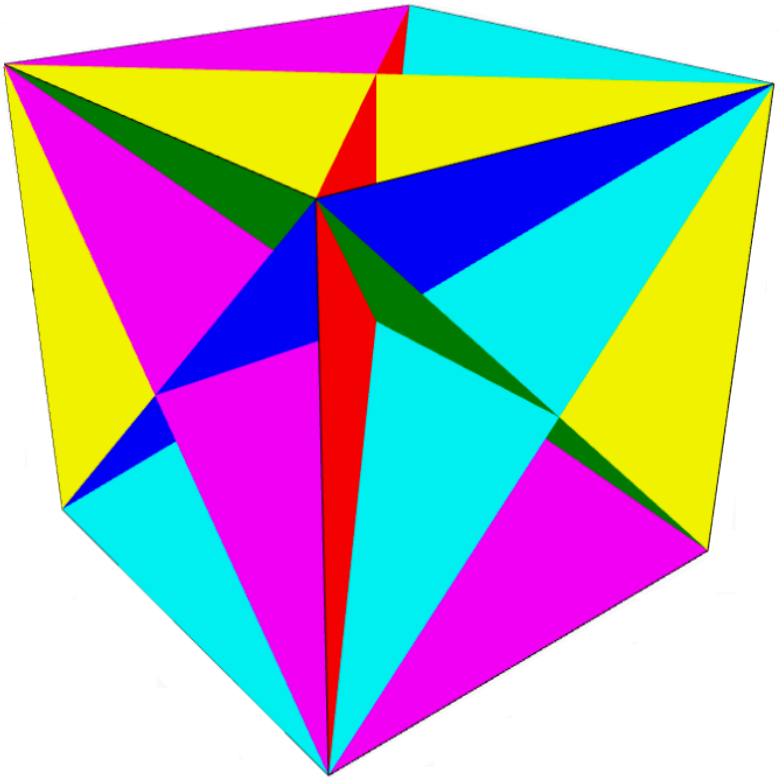
Lecture Notes on Algebraic Combinatorics



Jeremy L. Martin
University of Kansas
jlmartin@ku.edu

Contents

Foreword

These lecture notes began as my notes from Vic Reiner's Algebraic Combinatorics course at the University of Minnesota in Fall 2003. I currently use them for graduate courses at the University of Kansas. They will always be a work in progress. Please use them and share them freely for any research purpose. I have added and subtracted some material from Vic's course to suit my tastes, but any mistakes are my own; if you find one, please contact me at `jlmartin@ku.edu` so I can fix it. Thanks to those who have suggested additions and pointed out errors, including but not limited to: Kevin Adams, Nitin Aggarwal, Zaahir Ali, Trevor Arrigoni, Dylan Beck, Jonah Berggren, Lucas Chaffee, Matthew Chen, Geoffrey Critzer, Mark Denker, Souvik Dey, Joseph Doolittle, Ken Duna, Monalisa Dutta, Josh Fenton, Logan Godkin, Bennet Goeckner, Darij Grinberg (especially!), Brent Holmes, Arturo Jaramillo, Alex Lazar, Kevin Marshall, Dania Morales, George Nasr (especially!), Nick Packauskas, Abraham Pascoe, Smita Praharaj, John Portin, Enrique Salcido, Billy Sanders, Tony Se, Chris Uchizono, and Amanda Wilkens. Marge Bayer contributed the material on Ehrhart theory in §7.6.

Chapter 1

Posets and Lattices

1.1 Posets

Definition 1.1.1. A **partially ordered set** or **poset** is a set P equipped with a relation $\leq$ that is reflexive, antisymmetric, and transitive. That is, for all $x, y, z \in P$:

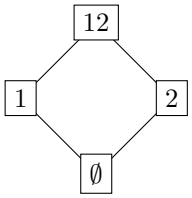
1. $x \leq x$ (reflexivity).
2. If $x \leq y$ and $y \leq x$, then $x = y$ (antisymmetry).
3. If $x \leq y$ and $y \leq z$, then $x \leq z$ (transitivity).

We say that x is **covered** by y, written $x \lessdot y$, if $x < y$ and there exists no z such that $x < z < y$. Two posets P, Q are **isomorphic** if there is a bijection $\phi : P \to Q$ that is order-preserving; that is, $x \leq y$ in P iff $\phi(x) \leq \phi(y)$ in Q. It is easy to check that isomorphism is an equivalence relation. A **subposet** of P is a subset $P' \subseteq P$ equipped with the order relation given by restriction from P.

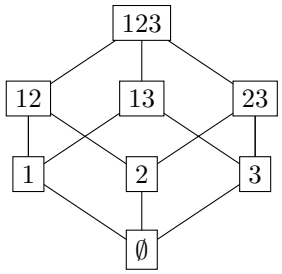
We will usually assume that P is finite. Sometimes a weaker assumption suffices, such that P is **chain-finite** (every chain is finite) or **locally finite** (every interval is finite). (We will say what "chains" and "intervals" are soon.)

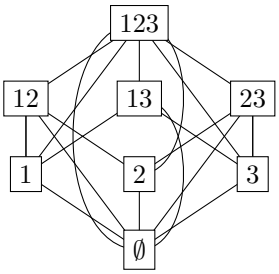
Remark 1.1.2. A **digraph** (short for "directed graph") consists of a set V of vertices and a set E of edges, which have the form $\overrightarrow{vw}$ for $v, w \in V$. A **directed acyclic graph** (or DAG) is a digraph with no *directed cycles*, i.e., edge sets of the form $\{\overrightarrow{v_1v_2}, \overrightarrow{v_2v_3}, \ldots, \overrightarrow{v_{n-1}v_n}, \overrightarrow{v_nv_1}\}$. If P is a poset, then the digraph with edges $\{\overrightarrow{xy} \mid x < y\}$ is a DAG. Conversely, for any DAG, the relation "$x \leq y$ if there exists a directed path $x \to \cdots \to y$ with zero or more edges" is a partial order. Thus posets and DAGs are essentially equivalent.

Example 1.1.3 (Boolean lattices). Let $[n] = \{1, 2, \ldots, n\}$ (a standard piece of notation in combinatorics) and let $2^{[n]}$ be the power set of $[n]$. We can partially order $2^{[n]}$ by writing $S \leq T$ if $S \subseteq T$. A poset isomorphic to $2^{[n]}$ is called a **Boolean lattice of rank n**. We may also use 2^S or Bool_S for the Boolean lattice of subsets of any finite set S; clearly $\mathsf{Bool}_S \cong \mathsf{Bool}_{|S|}$.



Bool_2





Bool_3

◀

The first two pictures are **Hasse diagrams**: graphs whose vertices are the elements of the poset and whose edges represent the **covering relations**, which are enough to generate all the relations in the poset by transitivity. (As you can see on the right, including *all* the relations would make the diagram unnecessarily complicated.) By convention, bigger elements in P are at the top of the picture.

The Boolean lattice 2^S has a unique minimum element (namely $\emptyset$) and a unique maximum element (namely S). Not every poset has to have such elements, but if a poset does, we will call them $\hat{\mathbf{0}}$ and $\hat{\mathbf{1}}$ respectively (or if necessary $\hat{\mathbf{0}}_P$ and $\hat{\mathbf{1}}_P$).

Definition 1.1.4. A poset that has both a $\hat{\mathbf{0}}$ and a $\hat{\mathbf{1}}$ is called **bounded**.[1] An element that covers $\hat{\mathbf{0}}$ is called an **atom**, and an element that is covered by $\hat{\mathbf{1}}$ is called a **coatom**. For example, the atoms in 2^S are the singleton subsets of S, and the coatoms are the subsets of cardinality $|S| - 1$.

We can make a poset P bounded: define a new poset $\hat{P}$ by adjoining new elements $\hat{\mathbf{0}}, \hat{\mathbf{1}}$ such that $\hat{\mathbf{0}} < x < \hat{\mathbf{1}}$ for every $x \in P$. Meanwhile, sometimes we have a bounded poset and want to delete the bottom and top elements.

Definition 1.1.5. Let $x, y \in P$ with $x \leq y$. The **interval** from x to y is the set

$$[x, y] = \{z \in P \colon x \leq z \leq y\}.$$

This formula makes sense if $x \not\leq y$, when $[x, y] = \emptyset$, but typically we don't want to think of the empty set as a bona fide interval. Also, $[x, y]$ is a singleton set if and only if $x = y$.

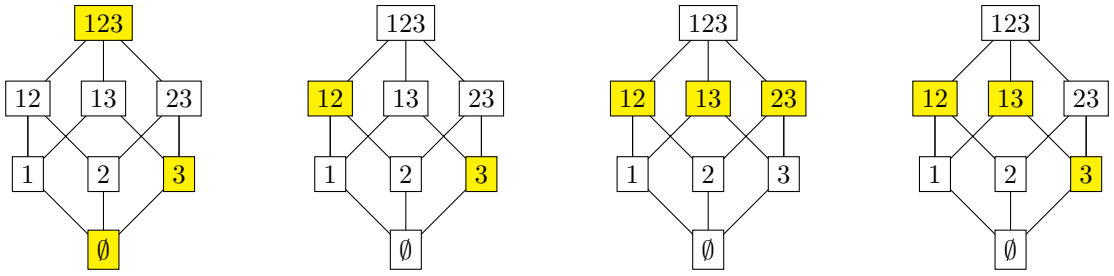
Definition 1.1.6. A subset $C \subseteq P$ (or P itself) is called a **chain** if its elements are pairwise comparable. Thus every chain is of the form $C = \{x_0, \ldots, x_n\}$, where $x_0 < \cdots < x_n$. The number n is called the **length** of the chain; notice that the length is one less than the cardinality of the chain. The chain C is called **saturated** if $x_0 \lessdot \cdots \lessdot x_n$; equivalently, C is maximal among all chains with bottom element x_0 and top element x_n. (Note that not all such chains necessarily have the same length — we will get back to that soon.) An **antichain** is a subset of P (or, again, P itself) in which no two of its elements are comparable.[2]

For example, in the Boolean lattice Bool_3, the subset[3] $\{\emptyset, 3, 123\}$ is a chain of length 2 (note that it is not saturated), while $\{12, 3\}$ and $\{12, 13, 23\}$ are antichains. The subset $\{12, 13, 3\}$ is neither a chain nor an antichain: 13 is comparable to 3 but not to 12.

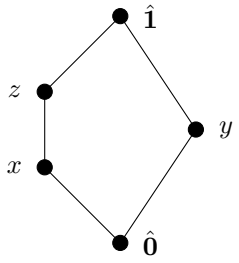
[1] This term has nothing to do with the more typical metric-space definition of "bounded".

[2] To set theorists, "antichain" means something stronger: a set of elements such that no two have a common lower bound. On the other hand, combinatorialists frequently want to talk about antichains in a bounded poset, where the more restrictive definition would be trivial.

[3] It is very common to drop the braces and commas when writing subsets of $[n]$: it is easier and cleaner to write $\{\emptyset, 3, 123\}$ rather than $\{\emptyset, \{3\}, \{1, 2, 3\}\}$.



One of the many nice properties of the Boolean lattice Bool_n is that its elements fall into horizontal slices (sorted by their cardinalities). Whenever $S \lessdot T$, it is the case that $|T| = |S| + 1$. A poset for which we can do this is called a **ranked** poset. However, it would be tautological to define a ranked poset to be a poset in which we can rank the elements! The actual definition of rankedness is a little more subtle, but makes perfect sense after a little thought, particularly after looking at an example of how a poset might fail to be ranked:



You can see what goes wrong — the chains $\hat{\mathbf{0}} \lessdot x \lessdot z \lessdot \hat{\mathbf{1}}$ and $\hat{\mathbf{0}} \lessdot y \lessdot \hat{\mathbf{1}}$ have the same bottom and top and are both saturated, but have different lengths. So the "rank" of $\hat{\mathbf{1}}$ is not well-defined; it could be either 2 or 3 more than the "rank" of $\hat{\mathbf{0}}$. Saturated chains are thus a key element in defining what "ranked" means.

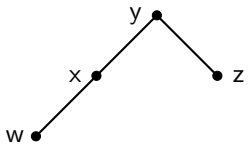
Definition 1.1.7. A poset P is **ranked** if for every $x, y \in P$, all saturated chains with bottom element x and top element y have the same length. A poset is **graded** if it is ranked and bounded.

In practice, most ranked posets we will consider are graded, or at least have a bottom element. To define a rank function $r : P \to \mathbb{Z}$, one can choose the rank of any single element arbitrarily, then assign the rest of the ranks by ensuring that

$$x \lessdot y \implies r(y) = r(x) + 1. \tag{1.1}$$

It is an exercise to prove that this definition results in no contradiction. It is standard to define $r(\hat{\mathbf{0}}) = 0$ so that all ranks are nonnegative; then $r(x)$ is the length of any saturated chain from $\hat{\mathbf{0}}$ to x. (Recall from Definition 1.1.6 that "length" means the number of *steps*, not the number of *elements* — i.e., edges rather than vertices in the Hasse diagram.)

Remark 1.1.8. The literature is not consistent on the usage of the term "ranked". In some sources "ranked" means what I am calling "graded"; it can also be used for the stronger condition that all maximal chains with the same top element have the same length. For example, the poset shown below, which satisfies Definition 1.1.7, is not ranked in this stronger sense, since $\{w, x, y\}$ and $\{z, y\}$ are both maximal elements of C_y:



There is no way to equip this poset with a rank function such that both minimal elements have rank 0 and (1.1) holds. On the other hand, many posets that arise in practice have a unique minimal element, so all these definitions are equivalent.

Definition 1.1.9. Let P be a ranked poset with rank function r. The **rank-generating function** of P is the formal power series

$$F_P(q) = \sum_{x \in P} q^{r(x)}.$$

Thus, for each k, the coefficient of q^k is the number of elements at rank k. (This definition, of course, makes sense only for posets with a finite number of elements at each rank.)

For example, the Boolean lattice is ranked by cardinality, with

$$F_{\mathsf{Bool}_n}(q) = \sum_{S \subseteq [n]} q^{|S|} = (1+q)^n.$$

The expansion of this polynomial is palindromic, because the coefficients are a row of Pascal's Triangle. That is, Bool_n is **rank-symmetric**.

More generally, if P and Q are ranked, then $P \times Q$ is ranked, with $r_{P\times Q}(x,y) = r_P(x) + r_Q(y)$, and $F_{P\times Q} = F_P F_Q$.

Definition 1.1.10. A **linear extension** of a poset P is a total order $\prec$ on the set P that refines $<_P$: that is, if $x <_P y$ then $x \prec y$. The set of all linear extensions is denoted $\mathscr{L}(P)$ (and sometimes called the *Jordan-Hölder set of* P).

If P is a chain then $\mathscr{L}(P) = \{P\}$, while if P is an antichain then $\mathscr{L}(P) = \mathfrak{S}_P$, the set of all permutations (= linear orders) of P. In general, the more relations P has, the fewer linear extensions. In this case $|\mathscr{L}(P + Q)| = \binom{|P|+|Q|}{|P|}$.

Definition 1.1.11. An **order ideal** (resp., an **order filter**) of P is a subposet $Q \subseteq P$ with the property that if $x, y \in P$, $x \in Q$, and $y \leq x$ (resp., $y \geq x$) then $y \in Q$.

Colloquially, an order ideal is a subset of P "closed under going down". Note that a subset of P is an order ideal if and only if its complement is an order filter. The order ideal **generated** by $Q \subseteq P$ is the smallest order ideal containing it, namely $\langle Q \rangle = \{x \in P \colon x \leq q \text{ for some } q \in Q\}$. Conversely, every order ideal has a unique minimal set of generators, namely its maximal elements (which form an antichain).

Example 1.1.12. Let $\{F_1, \ldots, F_k\}$ be a nonempty family of subsets of $[n]$. The order ideal they generate is

$$\Delta = \langle F_1, \ldots, F_k \rangle = \{G \subseteq [n] \colon G \subseteq F_i \text{ for some } i\}.$$

These order ideals are called **abstract simplicial complexes**, and are the standard combinatorial models for topological spaces (at least well-behaved ones). If each F_i is regarded as a simplex (i.e., the convex hull of a set of affinely independent points) then the order-ideal condition says that if Δ contains a simplex, then it contains all sub-simplices. For example, Δ cannot contain a triangle without also containing its edges and vertices. Simplicial complexes are the fundamental objects of topological combinatorics, and we will have much more to say about them in Chapter 6. ◀

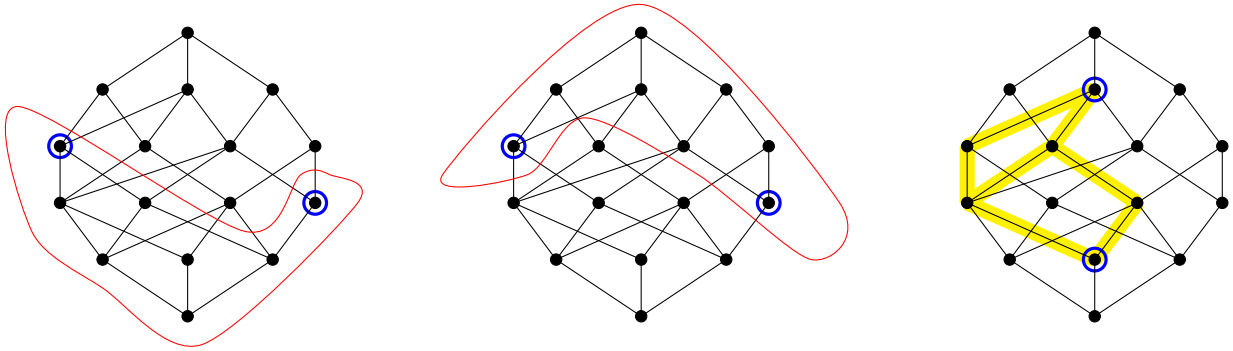


Figure 1.1: Order ideals, order filters, and intervals.

There are several ways to make new posets out of old ones. Here are some of the most basic.

Definition 1.1.13. Let P, Q be posets.

- The **dual** P^* of P is obtained by reversing all the order relations: $x \leq_{P^*} y$ iff $x \geq_P y$. The Hasse diagram of P^* is the same as that of P, turned upside down. A poset is **self-dual** if $P \cong P^*$; the map realizing the self-duality is called an **anti-automorphism**. For example, chains and antichains are self-dual, as is Bool_n (via the anti-automorphism $S \mapsto [n] \setminus S$). Any self-dual ranked poset is clearly rank-symmetric.
- The **disjoint union** $P + Q$ is the poset on $P \cup Q$ that inherits the relations from P and Q but no others, so that elements of P are incomparable with elements of Q. The Hasse diagram of $P + Q$ can be obtained by drawing the Hasse diagrams of P and Q side by side.
- The **Cartesian product** $P \times Q$ has a poset structure as follows: $(p, q) \leq (p', q')$ if $p \leq_P p'$ and $q \leq_Q q'$. This is a very natural and useful operation. For example, it is not hard to check that $\mathsf{Bool}_k \times \mathsf{Bool}_\ell \cong \mathsf{Bool}_{k+\ell}$.
- Assume that P has a $\hat{\mathbf{1}}$ and Q has a $\hat{\mathbf{0}}$. Then the **ordinal sum** $P \oplus Q$ is defined by identifying $\hat{\mathbf{1}}_P = \hat{\mathbf{0}}_Q$ and setting $p \leq q$ for all $p \in P$ and $q \in Q$. Note that this operation is not in general commutative (although it is associative).

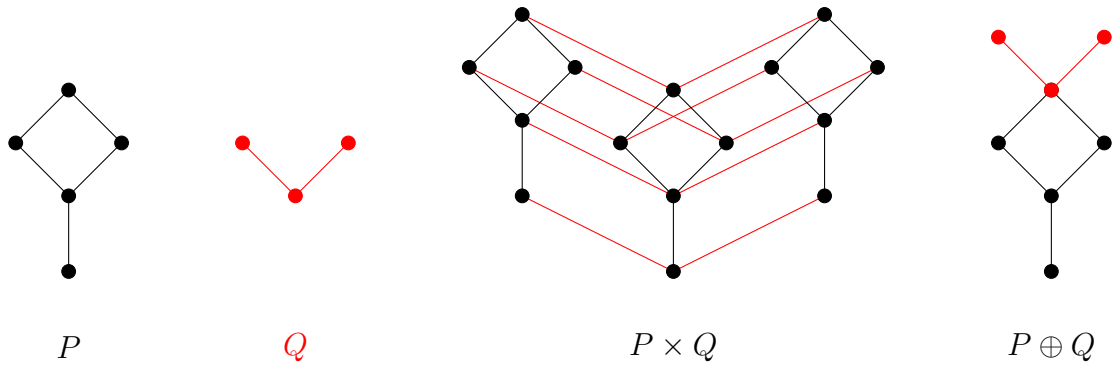


Figure 1.2: Direct product $\times$ and ordinal sum $\oplus$.

1.2 Lattices

Definition 1.2.1. A poset L is a **lattice** if every pair $x, y \in L$ has (i) a unique largest common lower bound, called their **meet** and written $x \wedge y$; (ii) a unique smallest common upper bound, called their **join** and written $x \vee y$. That is, for all $z \in L$,

$$z \leq x \text{ and } z \leq y \;\Rightarrow\; z \leq x \wedge y,$$
$$z \geq x \text{ and } z \geq y \;\Rightarrow\; z \geq x \vee y,$$

Note that, e.g., $x \wedge y = x$ if and only if $x \leq y$. Meet and join are easily seen to be commutative and associative, so for any finite $M \subseteq L$, the meet $\wedge M$ and join $\vee M$ are well-defined elements of L. In particular, every finite lattice is bounded, with $\hat{\mathbf{0}} = \wedge L$ and $\hat{\mathbf{1}} = \vee L$. (In an infinite lattice, the join or meet of an infinite set of elements may not be well-defined.[4]) For convenience, we set $\wedge\emptyset = \hat{\mathbf{1}}$ and $\vee\emptyset = \hat{\mathbf{0}}$.

It is easy to see that any poset isomorphic to a lattice is a lattice, and meet and join are equivariant under isomorphism (i.e., if f is an isomorphism, then $f(x \wedge y) = f(x) \wedge f(y)$ and $f(x \vee y) = f(x) \vee f(y)$). (Therefore, in order to show that two lattices are isomorphic, it is necessary only to show that they are isomorphic as posets.)

The canonical example of a lattice is the Boolean lattice $2^{[n]}$. Its meet and join are intersection and union, respectively. (In fact, the symbols $\wedge$ and $\vee$ were probably chosen to resemble $\cap$ and $\cup$.)

Example 1.2.2 (The partition lattice). An **[unordered] set partition** of S is a set of pairwise-disjoint, non-empty sets ("blocks") whose union is S. It is the same data as an equivalence relation on S, whose equivalence classes are the blocks. It is important to keep in mind that neither the blocks, nor the elements of each block, are ordered.

Let Π_n be the poset of all set partitions of $[n]$. For example, two elements of Π_5 are

$$\pi = \big\{\{1,3,4\},\ \{2,5\}\big\} \qquad \text{(abbr.: } 134|25\text{)}$$
$$\sigma = \big\{\{1,3\},\ \{4\},\ \{2,5\}\big\} \qquad \text{(abbr.: } 13|4|25\text{)}$$

We can impose a partial order on Π_n as follows: $\sigma \leq \pi$ if every block of σ is contained in a block of π; for short, σ **refines** π (as here). To put it another way, σ can be formed by further splitting up π, or equivalently every block of σ is a subset of some block of π. The lattices Π_3 and Π_4 are shown in Figure 1.3.

Observe that Π_n is bounded, with $\hat{\mathbf{0}} = 1|2|\cdots|n$ and $\hat{\mathbf{1}} = 12\cdots n$. For each set partition σ, the partitions that cover σ in Π_n are those obtained from σ by merging two of its blocks into a single block. Therefore, Π_n is graded, with rank function $r(\pi) = n - |\pi|$. The coefficients of the rank-generating function of Π_n are by definition the Stirling numbers of the second kind. Recall that $S(n,k)$ is the number of partitions of $[n]$ into k blocks, so

$$F_{\Pi_n}(q) = \sum_{k=1}^{n} S(n,k) q^{n-k}.$$

Furthermore, Π_n is a lattice: any two set partitions π, σ have a unique coarsest common refinement

$$\pi \wedge \sigma = \{A \cap B :\ A \in \pi,\ B \in \sigma,\ A \cap B \neq \emptyset\}.$$

Meanwhile, $\pi \vee \sigma$ is defined as the transitive closure of the union of the equivalence relations corresponding to π and σ.

[4] A lattice in which every set has a well-defined meet and join is called a *complete lattice*, although the concept does not arise often in a combinatorial setting.

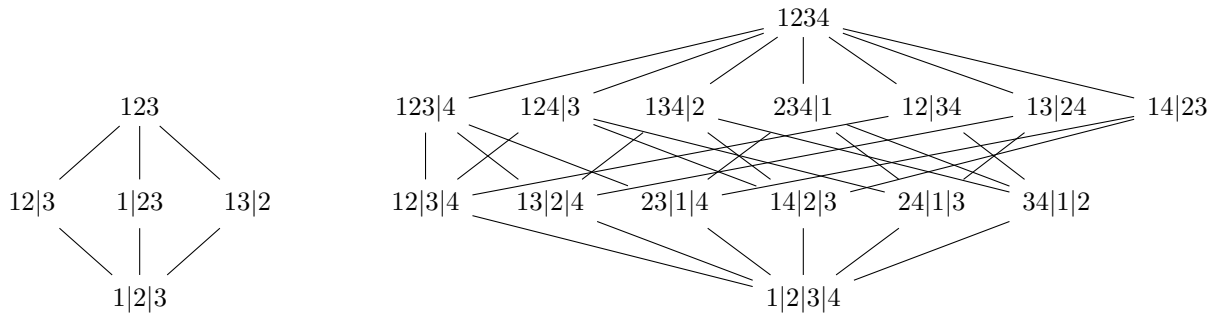


Figure 1.3: The partition lattices Π_3 and Π_4.

Finally, for any finite set, we can define Π_X to be the poset of set partitions of X, ordered by reverse refinement; evidently $\Pi_X \cong \Pi_{|X|}$. ◀

Example 1.2.3 (The connectivity lattice of a graph). Let $G = (V, E)$ be a graph. Recall that for $X \subseteq V$, the *induced subgraph* $G|_X$ is the graph on vertex set X, with two edges adjacent in $G|_X$ if and only if they are adjacent in G. The **connectivity lattice** of G is the subposet of Π_V defined by

$$K(G) = \{\pi \in \Pi_V : G|_X \text{ is connected for every block } X \in \pi\}.$$

For an example, see Figure 1.4. It is not hard to see that $K(G) = \Pi_V$ if and only if G is the complete graph K_V, and $K(G)$ is Boolean if and only if G is acyclic. Also, if H is a subgraph of G then $K(H)$ is a subposet of $K(G)$. The proof that $K(G)$ is in fact a lattice (justifying the terminology) is left as an exercise.

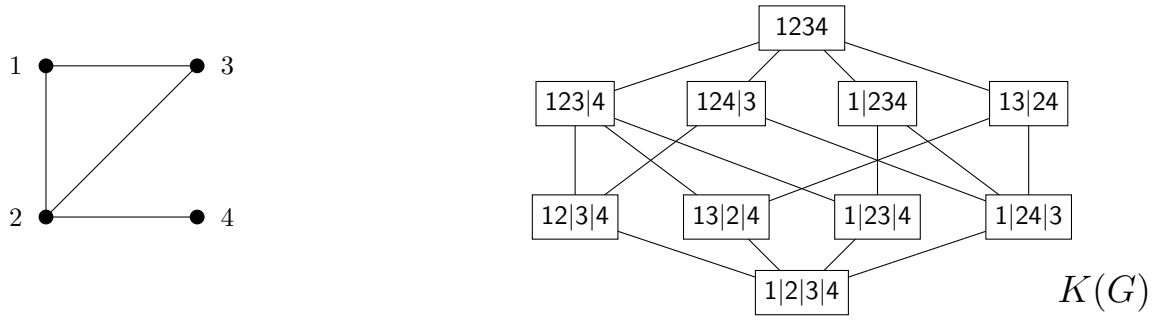


Figure 1.4: A graph and its connectivity lattice.

◀

Example 1.2.4 (Partitions, tableaux, and Young's lattice). An **(integer) partition** is a sequence $\lambda = (\lambda_1, \ldots, \lambda_\ell)$ of weakly decreasing positive integers: i.e., $\lambda_1 \geq \cdots \geq \lambda_\ell > 0$. If $n = \lambda_1 + \cdots + \lambda_\ell$, we write $\lambda \vdash n$ and/or $n = |\lambda|$. For convenience, we often set $\lambda_i = 0$ for all $i > \ell$.

Partitions are fundamental objects that will come up in many contexts. Let Y be the set of all partitions, partially ordered by $\lambda \geq \mu$ if $\lambda_i \geq \mu_i$ for all $i = 1, 2, \ldots$. Then Y is a ranked lattice, with rank function $r(\lambda) = |\lambda|$. Join and meet are given by component-wise max and min — we will shortly see another description of the lattice operations. ◀

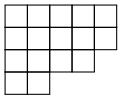
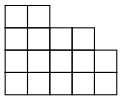
This is an infinite poset, but the number of partitions at any given rank is finite. In particular Y is **locally finite**, i.e., every interval is finite.[5] Moreover, the rank-generating function

$$\sum_{\lambda} q^{|\lambda|} = \sum_{n \geq 0} \sum_{\lambda \vdash n} q^n$$

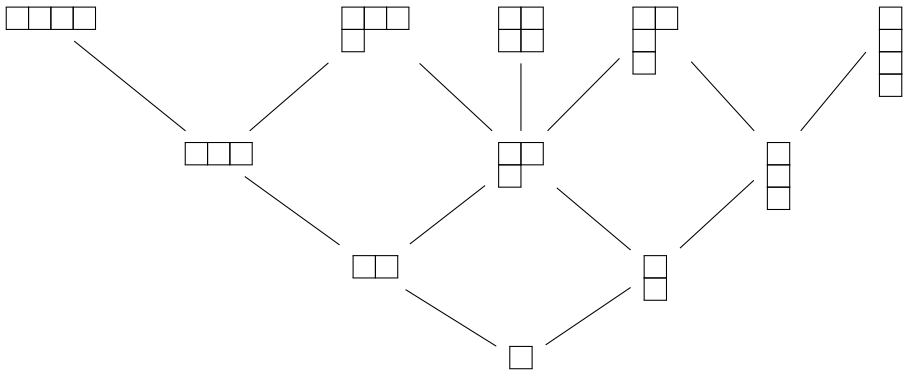
is a well-defined formal power series, and it is given by the justly celebrated formula

$$\prod_{k=1}^{\infty} \frac{1}{1-q^k}.$$

There is a nice pictorial way to look at Young's lattice. Instead of thinking about partitions as sequence of numbers, view them as their corresponding **Ferrers diagrams** (or **Young diagrams**): northwest-justified piles of boxes whose ith row contains λ_i boxes. The northwest-justification convention is called "English notation", and I will use that throughout, but a significant minority of combinatorialists prefer "French notation", in which the vertical axis is reversed. For example, the partition $(5, 5, 4, 2)$ is represented by the Ferrers diagram

 (English) or  (French).

Now the order relation in Young's lattice is as follows: $\lambda \geq \mu$ if and only if the Ferrers diagram of λ contains that of μ. The bottom part of the Hasse diagram of Y looks like this:

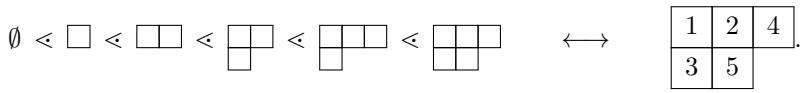


In terms of Ferrers diagrams, join and meet are simply union and intersection respectively.

Young's lattice Y has a nontrivial automorphism $\lambda \mapsto \tilde{\lambda}$ called **conjugation**. This is most easily described in terms of Ferrers diagrams: reflect across the line $x + y = 0$ so as to swap rows and columns. It is easy to check that if $\lambda \geq \mu$, then $\tilde{\lambda} \geq \tilde{\mu}$.

A maximal chain from $\emptyset$ to λ in Young's lattice can be represented by a **standard tableau**: a filling of λ with the numbers $1, 2, \ldots, |\lambda|$, using each number once, with every row increasing to the right and every column increasing downward. The kth element in the chain is the Ferrers diagram containing the numbers $1, \ldots, k$. For example:

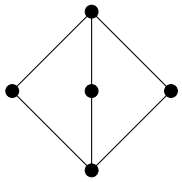
[5] In general, if X is any adjective, then "poset P is locally X" means "every interval in P is X".



Example 1.2.5 (Subspace lattices). Let q be a prime power, let $\mathbb{F}_q$ be the field of order q, and let $V = \mathbb{F}_q^n$ (a vector space of dimension n over $\mathbb{F}_q$). The **subspace lattice** $L_V(q) = L_n(q)$ is the set of all vector subspaces of V, ordered by inclusion. (We could replace $\mathbb{F}_q$ with an infinite field. The resulting poset is infinite, although chain-finite.)

The meet and join operations on $L_n(q)$ are given by $W \wedge W' = W \cap W'$ and $W \vee W' = W + W'$. We could construct analogous posets by ordering the (normal) subgroups of a group, or the prime ideals of a ring, or the submodules of a module, by inclusion. (However, these posets are not necessarily ranked, while $L_n(q)$ is ranked, by dimension.)

The simplest example is when $q = 2$ and $n = 2$, so that $V = \{(0,0), (0,1), (1,0), (1,1)\}$. Of course V has one subspace of dimension 2 (itself) and one of dimension 0 (the zero space). Meanwhile, it has three subspaces of dimension 1; each consists of the zero vector and one nonzero vector. Therefore, $L_2(2) \cong M_5$.



M_5

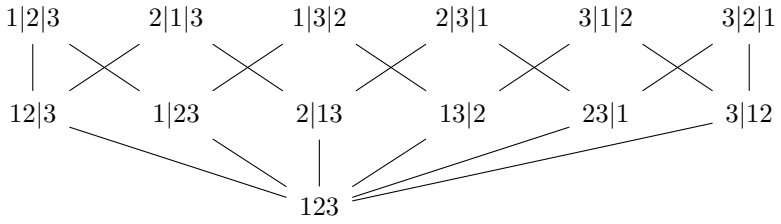
Note that $L_n(q)$ is self-dual, under the anti-automorphism $W \to W^\perp$ (the orthogonal complement with respect to any non-degenerate bilinear form).

The number of elements at rank k in $L_n(q)$, i.e., the number of k-dimensional subspaces of $\mathbb{F}_q^n$, is the ***q*-binomial coefficient**

$$\begin{bmatrix} n \\ k \end{bmatrix}_q = \frac{(q^n-1)(q^n-q)\cdots(q^n-q^{k-1})}{(q^k-1)(q^k-q)\cdots(q^k-q^{k-1})},$$

The proof is left as an exercise (Problem 1.14(b)). For more on q-binomial coefficients (including a proof that they are actually polynomials in q, not merely rational functions), see Problem 2.8. ◀

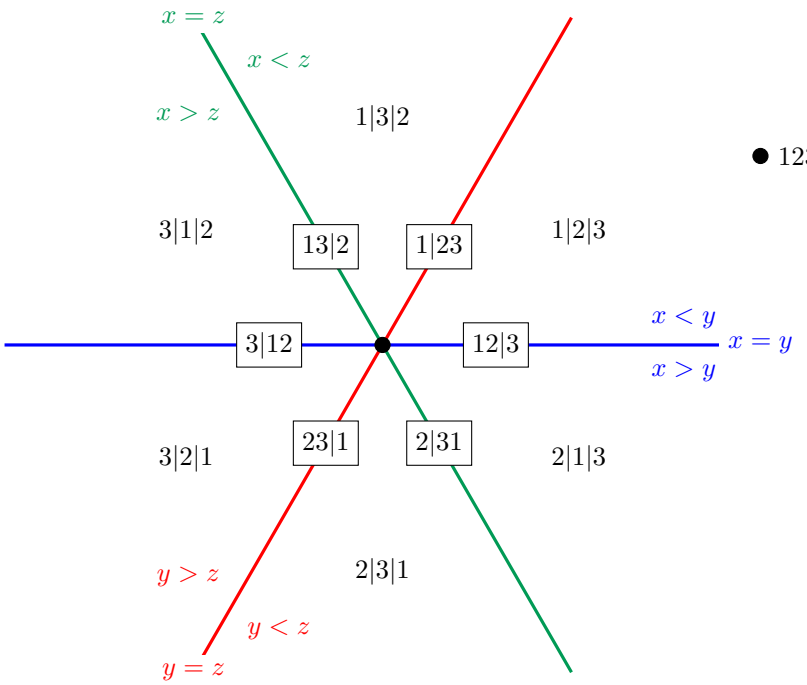
Example 1.2.6 (The lattice of ordered set partitions)**.** An **ordered set partition** (OSP) of S is an ordered list of pairwise-disjoint, non-empty sets ("blocks") whose union is S. Note the difference from unordered set partitions (Example 1.2.2). We use the same notation for OSPs as for their unordered cousins, but now, for example, $14|235$ and $235|14$ represent different OSPs. The set $\mathcal{O}_n$ of OSPs of $[n]$ is a poset under refinement: σ refines π if π can be obtained from σ by removing zero or more separator bars. For example, $16|247|389|5 \leq 16|2|4|7|38|9|5$, but $1|23|45$ and $12|345$ are incomparable. The Hasse diagram for $\mathcal{O}_3$ is as follows.



This poset is ranked, with rank function $r(\pi) = |\pi| - 1$ (i.e., the number of bars, or one less than the number of blocks, just like Π_n). Technically $\mathcal{O}_n$ is not a lattice but only a meet-semilattice, since join is not always well-defined. However, we can make it into a true lattice by appending an artificial $\hat{\mathbf{1}}$ at rank n.

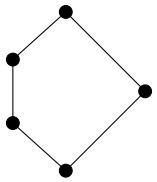
Interestingly, $\mathcal{O}_n$ is locally Boolean, i.e., every interval $[\pi, \sigma] \subseteq \mathcal{O}_n$ is a Boolean lattice, whose atoms correspond to the bars that appear in σ but not in π.

There is a nice geometric way to picture $\mathcal{O}_n$. Every point $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ gives rise to an OSP $\phi(\mathbf{x})$ that describes which coordinates are less than, equal to, or greater than others. For example, if $\mathbf{x} = (6, 6, 0, 4, 7) \in \mathbb{R}^5$, then $\phi(\mathbf{x}) = 3|4|12|5$, since $x_3 < x_4 < x_1 = x_2 < x_5$. Let $C_\pi = \phi^{-1}(\mathbf{x}) \subset \mathbb{R}^n$; that is, C_π is the set of points whose relative order of coordinates is given by π. Each set C_π is a **cone** (i.e., it is closed under addition and multiplication by positive scalars) and evidently the C_π decompose $\mathbb{R}^n$, so they give a good picture of $\mathcal{O}_n$. For example, the picture for $n = 3$ looks like this. (The picture is actually the cross-section in the plane $x_1 + x_2 + x_3 = 0$, but this cross-section is enough to see the full combinatorial structure.)



The topology matches the combinatorics: for example, each C_π is a $|\pi|$-dimensional space, and $\pi \leq \sigma$ in $\mathcal{O}_n$ if and only if $C_\pi \subseteq \overline{C_\sigma}$ (where the bar means closure). We will come back to this in more detail when we study hyperplane arrangements in Chapter 5; see especially Example5.7.1. ◀

Example 1.2.7. Lattices don't have to be ranked. For example, the poset N_5 shown below is a perfectly good lattice.



N_5

◀

Proposition 1.2.8 (Absorption laws). *Let L be a lattice and $x, y \in L$. Then $x \vee (x \wedge y) = x$ and $x \wedge (x \vee y) = x$. (Proof left to the reader.)*

The following result is a very common way of proving that a poset is a lattice.

Proposition 1.2.9. *Let P be a bounded poset that is a meet-semilattice (i.e., every nonempty $B \subseteq P$ has a well-defined meet $\wedge B$). Then every nonempty subset of P has a well-defined join, and consequently P is a lattice. Similarly, every bounded join-semilattice is a lattice.*

Proof. Let P be a bounded meet-semilattice. Let $A \subseteq P$, and let $B = \{b \in P : b \geq a \text{ for all } a \in A\}$. Note that $B \neq \emptyset$ because $\hat{\mathbf{1}} \in B$. Then $\wedge B$ is the unique least upper bound for A, for the following reasons. First, $\wedge B \geq a$ for all $a \in A$ by definition of B and of meet. Second, if $x \geq a$ for all $a \in A$, then $x \in B$ and so $x \geq \wedge B$. So every bounded meet-semilattice is a lattice, and the dual argument shows that every bounded join-semilattice is a lattice, □

This statement can be weakened slightly: any poset that has a unique top element and a well-defined meet operation is a lattice (the bottom element comes free as the meet of the entire set), as is any poset with a unique bottom element and a well-defined join.

Definition 1.2.10. Let L be a lattice. A **sublattice** of L is a subposet $L' \subseteq L$ that (a) is a lattice and (b) inherits its meet and join operations from L. That is,

$$x \wedge_{L'} y = x \wedge_L y \quad \text{and} \quad x \vee_{L'} y = x \vee_L y \qquad \forall x, y \in L'.$$

Equivalently, a sublattice of L is a subset that is closed under meet and join. We can speak of the sublattice of L *generated* by any subset $S \subseteq L$; it is just the smallest sublattice containing S.

Note that the maximum and minimum elements of a sublattice of L need not be the same as those of L. As an important example, every interval $L' = [x, z] \subseteq L$ (i.e., $L' = \{y \in L : x \leq y \leq z\}$) is a sublattice with minimum element x and maximum element z. (We might write $\hat{\mathbf{0}}_{L'} = x$ and $\hat{\mathbf{1}}_{L'} = z$.)

Example 1.2.11. Young's lattice Y is an infinite lattice. Meets of arbitrary sets are well-defined, as are finite joins. There is a $\hat{\mathbf{0}}$ element (the empty Ferrers diagram), but no $\hat{\mathbf{1}}$. On the other hand, Y is **locally finite** — every interval $[\lambda, \mu] \subseteq Y$ is finite. Similarly, the set of natural numbers, partially ordered by divisibility, is an infinite, locally finite lattice with a $\hat{\mathbf{0}}$. ◀

Example 1.2.12. Consider the set $M = \{A \subseteq [4] : A \text{ has even size}\}$. This is a lattice, but it is *not* a sublattice of Bool_4, because for example $12 \wedge_M 13 = \emptyset$ while $12 \wedge_{\mathsf{Bool}_4} 13 = 1$. ◀

Example 1.2.13. [Weak Bruhat order] Let $\mathfrak{S}_n$ be the set of permutations of $[n]$ (i.e., the symmetric group).[6] Write elements $w \in \mathfrak{S}_n$ as strings $w_1 w_2 \cdots w_n$ of distinct digits, e.g., $47182635 \in \mathfrak{S}_8$. (This is called *one-line*

[6] That's a Fraktur S, obtainable in LaTeX as `\mathfrak{S}`. The letter S has many other standard uses in combinatorics: Stirling numbers, symmetric functions, etc. The symmetric group is important enough to merit an ornate symbol!

notation.) The **weak Bruhat order** $\leq_W$ on $\mathfrak{S}_n$ is defined as follows: $w \lessdot_W v$ if v can be obtained by swapping w_i with w_{i+1}, where $w_i < w_{i+1}$. For example,

$$4716\underline{\mathbf{28}}35 \lessdot_W 4716\underline{\mathbf{82}}35 \quad \text{and} \quad 4\underline{\mathbf{71}}62835 \gtrdot_W 4\underline{\mathbf{17}}62835.$$

In other words, $v = ws_i$, where s_i is the transposition that swaps i with $i+1$. The weak order actually is a lattice, though this is not so easy to prove.

Another characterization of weak order is in terms of inversions. A **inversion** of $w \in \mathfrak{S}_n$ is an ordered pair (i, j) such that $i < j$ and $w_i > w_j$. Let $I(w)$ be the set of inversions of w and $\mathrm{inv}(w) = |I(w)|$ the number of inversions, which we label for future reference:

$$\mathrm{inv}(w) = \#\{\{i, j\} : i < j \text{ and } w_i > w_j\}. \tag{1.2}$$

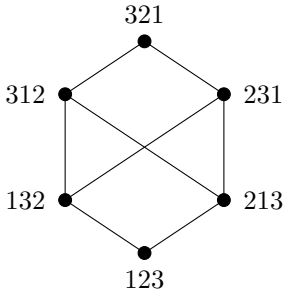
Weak order is ranked by inversion number, and in fact $v \leq_W w$ if and only if $I(v) \subseteq I(w)$.

Example 1.2.14. [**Bruhat order**] The **Bruhat order** $\leq_B$ on permutations is a related partial order with more relations (i.e., "stronger") than the weak order. It is defined as follows: $w \lessdot_B v$ if $\mathrm{inv}(v) > \mathrm{inv}(w)$ and $v = wt$ for some transposition t. For example,

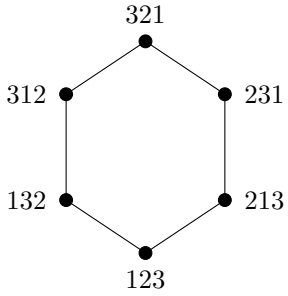
$$471\underline{\mathbf{6}}2\underline{\mathbf{8}}35 \lessdot_B 471\underline{\mathbf{8}}2\underline{\mathbf{6}}35$$

in Bruhat order (because this transposition has introduced exactly one more inversion), but not in weak order (since the positions transposed, namely 4 and 6, are not adjacent). On the other hand, $47\underline{\mathbf{1}}62\underline{\mathbf{8}}35$ is not covered by $47\underline{\mathbf{8}}62\underline{\mathbf{1}}35$ because this transposition increases the inversion number by 5, not by 1. (So this is a relation, but not a cover, in Bruhat order.) ◀

The Bruhat and weak orders on $\mathfrak{S}_3$ are shown below. You should be able to see from the picture that Bruhat order is not a lattice.



Bruhat order



Weak Bruhat order

A *Coxeter group* is a finite group generated by elements $s_1, \ldots, s_n$, called *simple reflections*, satisfying $s_i^2 = 1$ and $(s_i s_j)^{m_{ij}} = 1$ for all $i \neq j$ and some integers $m_{ij} \geq 2$. For example, setting $m_{ij} = 3$ if $|i - j| = 1$ and $m_{ij} = 2$ if $|i - j| > 1$, we obtain the symmetric group $\mathfrak{S}_{n+1}$. Coxeter groups are fantastically important in geometric combinatorics and we could spend at least a semester on them. The standard resources are the books by Brenti and Björner [BB05], which has a more combinatorial approach, and Humphreys [Hum90], which has a more geometric flavor. For now, it's enough to mention that every Coxeter group has associated Bruhat and weak orders, whose definitions generalize those for the symmetric group.

The Bruhat and weak order give graded, self-dual poset structures on $\mathfrak{S}_n$, both ranked by number of **inversions**:

$$r(w) = \Big|\big\{\{i, j\} : i < j \text{ and } w_i > w_j\big\}\Big|.$$

(For a general Coxeter group, the rank of an element w is the minimum number r such that w is the product of r simple reflections.) The rank-generating function of $\mathfrak{S}_n$ is a very nice polynomial called the **q-factorial** (or "the q-analogue of n factorial", "n factorial base q", etc.):

$$F_{\mathfrak{S}_n}(q) = 1(1+q)(1+q+q^2)\cdots(1+q+\cdots+q^{n-1}) = \prod_{i=1}^{n} \frac{1-q^i}{1-q}. \qquad \blacktriangleleft$$

1.3 Distributive lattices

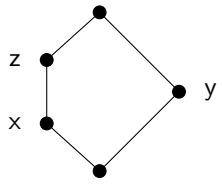
Definition 1.3.1. A lattice L is **distributive** if the following two equivalent conditions hold:

$$x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z) \qquad \forall x, y, z \in L, \tag{1.3a}$$
$$x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z) \qquad \forall x, y, z \in L. \tag{1.3b}$$

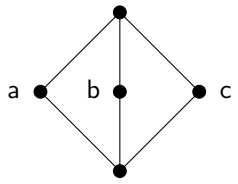
Proving that the two conditions (1.3a) and (1.3b) are equivalent is not too hard, but is not trivial (Problem 1.8). Note that replacing the equalities with $\geq$ and $\leq$ respectively gives statements that are true for all lattices.

The condition of distributivity seems natural, but in fact distributive lattices are quite special.

1. The Boolean lattice $2^{[n]}$ is a distributive lattice, because the set-theoretic operations of union and intersection are distributive over each other.
2. Every sublattice of a distributive lattice is distributive. In particular, Young's lattice Y is distributive because it is a sublattice of a Boolean lattice (recall that meet and join in Y are given by intersection and union on Ferrers diagrams).
3. The lattices M_5 and N_5 are not distributive:



$$(x \vee y) \wedge z = \hat{\mathbf{1}} \wedge z = z$$
$$(x \wedge z) \vee (y \wedge z) = x \vee \hat{\mathbf{0}} = x$$



$$(a \vee b) \wedge c = c$$
$$(a \wedge c) \vee (b \wedge c) = \hat{\mathbf{0}}.$$

4. The partition lattice Π_n is not distributive for $n \geq 3$, because $\Pi_3 \cong M_5$, and for $n \geq 4$ every Π_n contains a sublattice isomorphic to Π_3 (see Problem 1.1). Likewise, if $n \geq 2$ then the subspace lattice $L_n(q)$ contains a copy of M_5 (take any plane together with three distinct lines in it), hence is not distributive.
5. The set D_n of all positive integer divisors of a fixed integer n, ordered by divisibility, is a distributive lattice (Problem 1.3).

Every poset P gives rise to a distributive lattice in the following way. The set $J(P)$ of order ideals of P (see Definition 1.1.11) is itself a bounded poset, ordered by containment. In fact $J(P)$ is a distributive lattice: the union or intersection of order ideals is an order ideal (this is easy to check) which means that $J(P)$ is a sublattice of the distributive lattice Bool_P. (See Figure 1.5 for an example.)

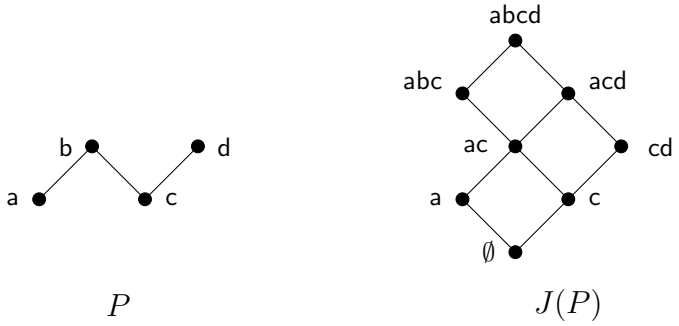


Figure 1.5: A poset P and the corresponding distributive lattice $J(P)$.

For example, if P is an antichain, then every subset is an order ideal, so $J(P) = \mathsf{Bool}_P$, while if P is a chain with n elements, then $J(P)$ is a chain with $n+1$ elements. As an infinite example, if $P = \mathbb{N}^2$ with the product ordering (i.e., $(x, y) \leq (x', y')$ if $x \leq x'$ and $y \leq y'$), then $J(P)$ is Young's lattice Y.

Remark 1.3.2. There is a natural bijection between $J(P)$ and the set of antichains of P, since the maximal elements of any order ideal form an antichain that generates it. (Recall that an antichain is a set of elements that are pairwise incomparable.) Moreover, for each order ideal I, the order ideals covered by I in $J(P)$ are precisely those of the form $I' = I \setminus \{x\}$, where x is a maximal element of I. In particular $|I'| = |I| - 1$ for all such I', and it follows by induction that $J(P)$ is ranked by cardinality.

We will shortly prove Birkhoff's theorem (Theorem 1.3.7), a.k.a. the Fundamental Theorem of Finite Distributive Lattices: the finite distributive lattices are *exactly* the lattices of the form $J(P)$, where P is a finite poset.

Definition 1.3.3. Let L be a lattice. An element $x \in L$ is **join-irreducible** if it cannot be written as the join of two other elements. That is, if $x = y \vee z$ then either $x = y$ or $x = z$. The subposet (not sublattice!) of L consisting of all join-irreducible elements is denoted $\mathrm{Irr}(L)$. Here is an example.

If L is finite, then an element of L is join-irreducible if it covers exactly one other element. (This is not true in a lattice such as $\mathbb{R}$ under the natural order, in which there are no covering relations!) The condition of finiteness can be relaxed; see Problem 1.10.

Definition 1.3.4. A **factorization** of $x \in L$ is an equation of the form

$$x = p_1 \vee \cdots \vee p_n$$

where $p_1, \ldots, p_n \in \mathrm{Irr}(L)$. The factorization is **irredundant** if the p_i form an antichain.

In analogy with ring theory, call a lattice **Artinian** if it has no infinite descending chains. (For example, L is Artinian if it is finite, or chain-finite, or locally finite and has a $\hat{\mathbf{0}}$.) If L is Artinian, then every element $x \in L$ has a factorization — if x itself is not join-irreducible, express it as a join of two smaller elements, then repeat. Moreover, every factorization can be reduced to an irredundant factorization by deleting each factor strictly less than another (which does not change the join of the factors). **Throughout the rest of the section, we will assume that L is Artinian.**

For general lattices, irredundant factorizations need not be unique. For example, the $\hat{\mathbf{1}}$ element of M_5 can be factored irredundantly as the join of any two atoms. On the other hand, *distributive* lattices do exhibit unique factorization, as we will soon prove (Proposition 1.3.6).

Proposition 1.3.5. *Let L be a distributive lattice and let $p \in \operatorname{Irr}(L)$. Suppose that $p \leq q_1 \vee \cdots \vee q_n$. Then $p \leq q_i$ for some i.*

Proof. By distributivity,

$$p = p \wedge (q_1 \vee \cdots \vee q_n) = (p \wedge q_1) \vee \cdots \vee (p \wedge q_n)$$

and since p is join-irreducible, it must equal $p \wedge q_i$ for some i, whence $p \leq q_i$. □

Proposition 1.3.5 is a lattice-theoretic analogue of the statement that if a prime p divides a product of positive numbers, then it divides at least one of them. (This is in fact exactly what the result says when applied to the divisor lattice D_n.)

Proposition 1.3.6 (Unique factorization for distributive lattices)**.** *Let L be a distributive lattice. Then every $x \in L$ can be written uniquely as an irredundant join of join-irreducible elements.*

Proof. Suppose that we have two irredundant factorizations

$$x = p_1 \vee \cdots \vee p_n = q_1 \vee \cdots \vee q_m \tag{1.4}$$

with $p_i, q_j \in \operatorname{Irr}(L)$ for all i, j. Then $p_1 \leq x = q_1 \vee \cdots \vee q_m$, so by Proposition 1.3.5, $p_1 \leq q_j$ for some j. Again by Proposition 1.3.5, $q_j \leq p_i$ for some i. If $i \neq 1$, then $p_1 \lneq p_i$, which contradicts the fact that the p_i form an antichain. Therefore $p_1 = q_j$. This argument implies that each p_i is one of the q_j's, and vice versa. Therefore, the two factorizations in (1.4) must be identical. □

Theorem 1.3.7 (Birkhoff 1933)**.** *Up to isomorphism, the finite distributive lattices are exactly the lattices $J(P)$, where P is a finite poset. Moreover, $L \cong J(\operatorname{Irr}(L))$ for every lattice L and $P \cong \operatorname{Irr}(J(P))$ for every poset P.*

Sketch of proof. The lattice isomorphism $L \to J(\operatorname{Irr}(L))$ is given by

$$\phi(x) = \{p \in \operatorname{Irr}(L) \colon p \leq x\}.$$

Meanwhile, the join-irreducible order ideals in P are just the principal order ideals, i.e., those generated by a single element. So the poset isomorphism $P \to \operatorname{Irr}(J(P))$ is given by

$$\psi(y) = \langle y \rangle.$$

These facts need to be checked; the details are left to the reader (Problem 1.12).

Corollary 1.3.8. *Every finite distributive lattice L is graded.*

Proof. The FTFDL says that $L \cong J(P)$ for some finite poset P. Then L is ranked by Remark 1.3.2, and it is bounded with $\hat{\mathbf{0}} = \emptyset$ and $\hat{\mathbf{1}} = P$. □

Corollary 1.3.9. *Let L be a finite distributive lattice. The following are equivalent:*

1. *L is a Boolean lattice.*
2. *$\mathrm{Irr}(L)$ is an antichain.*
3. *L is atomic (i.e., every element in L is the join of atoms). Equivalently, every join-irreducible element is an atom.*
4. *L is **complemented**. That is, for each $x \in L$, there exists a unique element $\bar{x} \in L$ such that $x \vee \bar{x} = \hat{\mathbf{1}}$ and $x \wedge \bar{x} = \hat{\mathbf{0}}$.*
5. *L is **relatively complemented**. That is, for every interval $[y, z] \subseteq L$ and every $x \in [y, z]$, there exists a unique element $u \in [y, z]$ such that $x \vee u = z$ and $x \wedge u = y$.*

Proof. (5) $\implies$ (4): Take $[x, y] = [\hat{\mathbf{0}}, \hat{\mathbf{1}}]$.

(4) $\implies$ (3): Suppose that L is complemented, and suppose that $y \in \mathrm{Irr}(L)$ is not an atom. Let x be an atom in $[\hat{\mathbf{0}}, y]$. Then

$$(x \vee \bar{x}) \wedge y = \hat{\mathbf{1}} \wedge y = y$$
$$(x \vee \bar{x}) \wedge y = (x \wedge y) \vee (\bar{x} \wedge y) = x \vee (\bar{x} \wedge y)$$

by distributivity. So $y = x \vee (\bar{x} \wedge y)$, which is a factorization of y, but y is join-irreducible, which implies $\bar{x} \wedge y = y$, i.e., $\bar{x} \geq y$. But then $\bar{x} \geq x$ and $\bar{x} \wedge x = x \neq \hat{\mathbf{0}}$, a contradiction.

(3) $\implies$ (2): This follows from the observation that no two atoms are comparable.

(2) $\implies$ (1): By the FTFDL, since $L = J(\mathrm{Irr}(L))$.

(1) $\implies$ (5): If $X \subseteq Y \subseteq Z$ are sets, then let $U = X \cup (Y \setminus Z)$. Then $Y \cap U = X$ and $Y \cup U = Z$. □

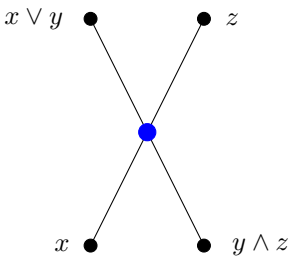
Join and meet could have been interchanged throughout this section. For example, the dual of Proposition 1.3.6 says that every element in a distributive lattice L has a unique "cofactorization" as an irredundant meet of meet-irreducible elements, and L is Boolean iff every element is the meet of coatoms. (In this case we would require L to be Noetherian instead of Artinian — i.e., to contain no infinite increasing chains. For example, Young's lattice is Artinian but not Noetherian.)

1.4 Modular lattices

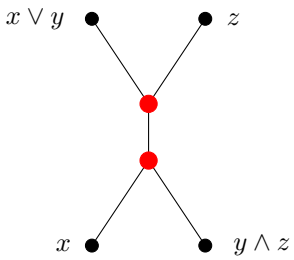
Definition 1.4.1. A lattice L is **modular** if every $x, y, z \in L$ with $x \leq z$ satisfy the *modular equation*:

$$x \vee (y \wedge z) = (x \vee y) \wedge z. \tag{1.5}$$

Note that for all lattices, if $x \leq z$, then $x \vee (y \wedge z) \leq (x \vee y) \wedge z$. Modularity says that, in fact, equality holds.



Modular

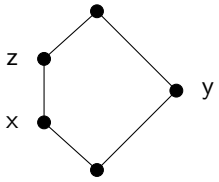


Non-modular

The term "modularity" arises in algebra: a canonical example of a modular lattice is the poset of modules over any ring, ordered by inclusion (Corollary 1.4.3).

Some basic facts and examples:

1. Every sublattice of a modular lattice is modular.
2. Distributive lattices are modular: if L is distributive and $x \leq z \in L$, then
$$x \vee (y \wedge z) \;=\; (x \vee y) \wedge (x \vee z) \;=\; (x \vee y) \wedge z.$$
3. The lattice L is modular if and only if its dual L^* is modular. Unlike the corresponding statement for distributivity, this is immediate, because the modular equation is invariant under dualization.
4. The nonranked lattice N_5 is not modular.



Here $x \leq z$, but
$$x \vee (y \wedge z) = x \vee \hat{\mathbf{0}} = x,$$
$$(x \vee y) \wedge z = \hat{\mathbf{1}} \wedge z = z.$$
In fact, N_5 is the unique obstruction to modularity, as we will soon see (Thm. 1.4.5).
5. The nondistributive lattice $M_5 \cong \Pi_3$ is modular. However, Π_4 is not modular (exercise).

Theorem 1.4.2. *[Characterizations of modularity] Let L be a lattice. Then the following are equivalent:*

1. *L is modular.*
2. *For all $x, y, z \in L$, if $x \in [y \wedge z, z]$, then $x = (x \vee y) \wedge z$.*
3. *For all $x, y, z \in L$, if $x \in [y, y \vee z]$, then $x = (x \wedge z) \vee y$.*
4. *For all $y, z \in L$, the lattices $L' = [y \wedge z, z]$ and $L'' = [y, y \vee z]$ are isomorphic, via the maps*
$$\begin{aligned} \alpha : L' &\to L'' & \qquad \beta : L'' &\to L' \\ q &\mapsto q \vee y, & p &\mapsto p \wedge z. \end{aligned}$$

Proof. (1) $\Longrightarrow$ (2): If $y \wedge z \leq x \leq z$, then the modular equation $x \vee (y \wedge z) = (x \vee y) \wedge z$ reduces to $x = (x \vee y) \wedge z$.

(2) $\Longrightarrow$ (1)): Suppose that (2) holds. Let $a, b, c \in L$ with $a \leq c$. Then
$$b \wedge c \;\leq\; a \vee (b \wedge c) \;\leq\; c \vee c \;=\; c$$

so applying (2) with $y = b$, $z = c$, $x = a \vee (b \wedge c)$ gives

$$a \vee (b \wedge c) \;=\; \big((a \vee (b \wedge c)) \vee b\big) \wedge c \;=\; (a \vee b) \wedge c$$

which is the modular equation for a, b, c.

(2) $\iff$ (3): These two conditions are duals of each other (i.e., L satisfies (2) iff L^* satisfies (3)), and modularity is a self-dual condition.

(2)+(3) $\iff$ (4): The functions α and β are always order-preserving functions with the stated domains and ranges. Conditions (2) and (3) say respectively that $\beta \circ \alpha$ and $\alpha \circ \beta$ are the identities on L' and L''; together, these conditions are equivalent to condition (4). □

Corollary 1.4.3. *Let R be a (not necessarily commutative) ring and M a (left) R-submodule. Then the (possibly infinite) poset $L(M)$ of (left) R-submodules of M, ordered by inclusion, is a modular lattice with operations $Y \vee Z = Y + Z$ and $Y \wedge Z = Y \cap Z$.*

Proof. The Second Isomorphism Theorem says that $Z/(Y \cap Z) \cong (Y + Z)/Y$ for all $Y, Z \in L(M)$. Therefore

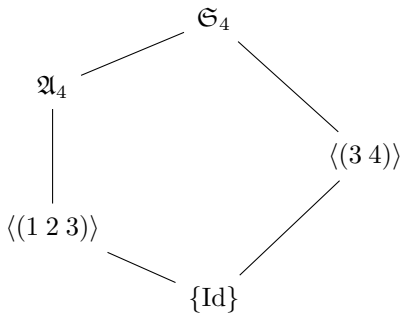
$$[Y \cap Z, Z] \;\cong\; L(Z/(Y \cap Z)) \;\cong\; L((Y + Z)/Y) \;\cong\; [Y, Y + Z]$$

so $L(M)$ satisfies condition 4 of Theorem 1.4.2. □

In particular, the subspace lattices $L_n(q)$ are modular (see Example 1.2.5).

Example 1.4.4. For a (finite) group G, let $L(G)$ denote the lattice of subgroups of G, with operations $H \wedge K = H \cap K$ and $H \vee K = HK$ (i.e., the group generated by $H \cup K$). If G is abelian then $L(G)$ is always modular, but if G is non-abelian then modularity can fail.

For example, let $G = \mathfrak{S}_4$, let X and Y be the cyclic subgroups generated by the cycles $(1\ 2\ 3)$ and $(3\ 4)$ respectively, and let $Z = \mathfrak{A}_4$ (the alternating group). Then $(XY) \cap Z = Z$ but $X(Y \cap Z) = Z$. Indeed, these groups generate a sublattice of $L(\mathfrak{S}_4)$ isomorphic to N_5:



◀

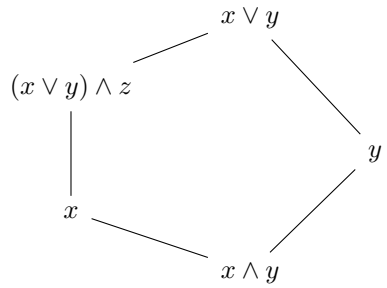
In fact, an occurrence of N_5 is the only obstruction to modularity:

Theorem 1.4.5. *Let L be a lattice.*

1. *L is modular if and only if it contains no sublattice isomorphic to N_5.*
2. *L is distributive if and only if it contains no sublattice isomorphic to N_5 or M_5.*

Proof. Both $\implies$ directions are easy, because distributivity and modularity are conditions inherited by sublattices, and N_5 is not modular and M_5 is not distributive.

Suppose that x, y, z is a triple for which modularity fails. One can check that



is a sublattice (details left to the reader), and it is isomorphic to N_5.

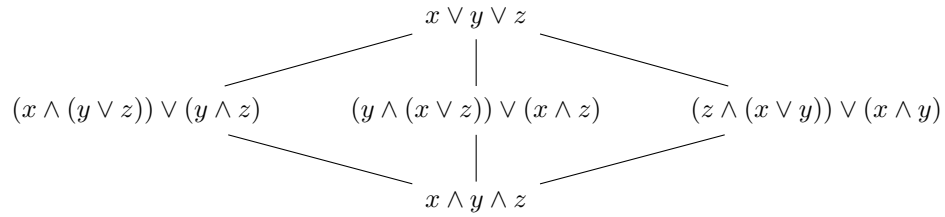
Suppose that L is not distributive. If it isn't modular then it contains an N_5, so there is nothing to prove. If it is modular, then choose x, y, z such that

$$x \wedge (y \vee z) > (x \wedge y) \vee (x \wedge z).$$

You can then show that

1. this inequality is invariant under permuting x, y, z;
2. $(x \wedge (y \vee z)) \vee (y \wedge z)$ and the two other lattice elements obtained by permuting x, y, z form an antichain;
3. $x \vee y = x \vee z = y \vee z$, and likewise for meets.

Hence we have constructed a sublattice of L isomorphic to M_5.



□

A corollary is that every modular lattice is graded, because a non-graded lattice must contain a sublattice isomorphic to N_5. The details are left to the reader; we will eventually prove the stronger statement that every *semimodular* lattice is graded.

1.5 Semimodular lattices

Recall that the notation $x \lessdot y$ means that x is covered by y, i.e., $x < y$ and there exists no z strictly between x, y (i.e., such that $x < z < y$).

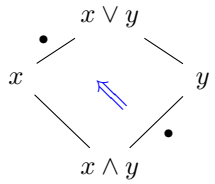
Definition 1.5.1. A lattice L is **(upper) semimodular** if for all incomparable $x, y \in L$,

$$x \wedge y \lessdot y \implies x \lessdot x \vee y. \tag{1.6}$$

Conversely, L is **lower semimodular** if the converse holds.

Note that both upper and lower semimodularity are inherited by sublattices, and that L is upper semimodular if and only if its dual L^* is lower semimodular. Also, the implication (1.6) is trivially true if x and y are comparable. If they are incomparable (as we will often assume), then there are several useful colloquial rephrasings of semimodularity:

- "If meeting with x merely nudges y down, then joining with y merely nudges x up."
- In the interval $[x \wedge y, x \vee y] \subseteq L$ pictured below, if the southeast relation is a cover, then so is the northwest relation.



(1.7)

- This condition is often used symmetrically: if x, y are incomparable and they both cover $x \wedge y$, then they are both covered by $x \vee y$.
- Contrapositively, "If there is other stuff between x and $x \vee y$, then there is also other stuff between $x \wedge y$ and y."

Example 1.5.2. The partition lattice Π_n is an important example of an upper semimodular lattice. To see that it is USM, let π and σ be incomparable set partitions of $[n]$, and suppose that $\sigma \gtrdot \sigma \wedge \pi$. Recall that this means that $\sigma \wedge \pi$ can be obtained from σ by splitting some block $B \in \sigma$ into two sub-blocks B', B''. More specifically, we can write $\sigma = A_1|\cdots|A_k|B$ and $\sigma \wedge \pi = A_1|\cdots|A_k|B'|B''$, where B is the disjoint union of B' and B''. Since $\sigma \wedge \pi$ refines π but σ does not, we know that $A_1, \ldots, A_k, B', B''$ are all subsets of blocks of π but B is not; in particular B' and B'' are subsets of different blocks of π, say C' and C'' respectively. But then merging C' and C'' produces a partition τ that covers π and is refined by σ, so it must be the case that $\tau = \sigma \vee \pi$, and we have proved that Π_n is USM. ◀

Lemma 1.5.3. *If a lattice L is modular, then it is both upper and lower semimodular.*

Proof. If $x \wedge y \lessdot y$, then the sublattice $[x \wedge y, y]$ has only two elements. If L is modular, then condition (4) of the characterization of modularity (Theorem 1.4.2) implies that $[x \wedge y, y] \cong [x, x \vee y]$, so $x \lessdot x \vee y$. Hence L is upper semimodular. The dual argument proves that L is lower semimodular. □

In fact, upper and lower semimodularity together imply modularity. We will show that any of these three conditions on a lattice L implies that it is graded, and that its rank function r satisfies

$$\begin{aligned} r(x \vee y) + r(x \wedge y) &\leq r(x) + r(y) && \text{iff } L \text{ is USM,} \\ r(x \vee y) + r(x \wedge y) &\geq r(x) + r(y) && \text{iff } L \text{ is LSM,} \\ r(x \vee y) + r(x \wedge y) &= r(x) + r(y) && \text{iff } L \text{ is modular.} \end{aligned}$$

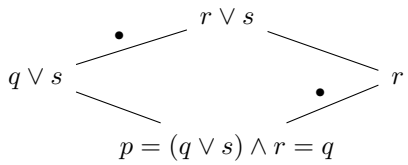
Lemma 1.5.4. *Suppose L is USM and let $q, r, s \in L$. If $q \lessdot r$, then either $q \vee s = r \vee s$ or $q \vee s \lessdot r \vee s$.*

In other words, if it only takes one step to walk up from q to r, then it takes *at most* one step to walk from $q \vee s$ to $r \vee s$.

Proof. Let $p = (q \vee s) \wedge r$, so that $q \leq p \leq r$. Since q is covered by r, it follows that either $p = q$ or $p = r$.

- If $p = r$, then $q \vee s \geq r$. So $q \vee s = r \vee (q \vee s) = (r \vee q) \vee s = r \vee s$.

- If $p = q$, then $p = (q \vee s) \wedge r = q \lessdot r$. Applying semimodularity to the diamond figure below, we obtain $(q \vee s) \lessdot (q \vee s) \vee r = r \vee s$.



□

Theorem 1.5.5. *Let L be a finite lattice. Then L is USM if and only if it is ranked, with rank function r satisfying the **submodular inequality** or **semimodular inequality***

$$r(x \vee y) + r(x \wedge y) \;\leq\; r(x) + r(y) \qquad \forall x, y \in L. \tag{1.8}$$

Proof. ($\Longleftarrow$) Suppose that L is a ranked lattice with rank function r satisfying (1.8). Suppose that x, y are incomparable and $x \wedge y \lessdot y$ so that $r(y) = r(x \wedge y) + 1$. Incomparability implies $x \vee y > x$, so $r(x \vee y) - r(x) > 0$. On the other hand, rearranging (1.8) gives

$$0 \;<\; r(x \vee y) - r(x) \;\leq\; r(y) - r(x \wedge y) \;=\; 1$$

so $r(x \vee y) - r(x) = 1$, i.e., $x \vee y \gtrdot x$.

($\Longrightarrow$) For later use, observe that if L is semimodular, then

$$x \wedge y \;\lessdot\; x, y \;\Longrightarrow\; x, y \;\lessdot\; x \vee y. \tag{1.9}$$

Denote by $c(L)$ the maximum length[7] of a chain in L. We will induct on $c(L)$. For the base cases, if $c(L) = 0$ then L has one element, while if $c(L) = 1$ then L has two elements. If $c(L) = 2$ then $L = \{\hat{\mathbf{0}}, \hat{\mathbf{1}}, x_1, \ldots, x_n\}$, where $n \geq 1$ and $\hat{\mathbf{0}} \lessdot x_i \lessdot \hat{\mathbf{1}}$ for all i. It is easy to see that these lattices are ranked, USM and satisfy (1.8) (in fact equality holds and these lattices are modular). Therefore, suppose $c(L) = n \geq 3$. Assume inductively that if $\tilde{L}$ is USM and $c(\tilde{L}) < c(L)$, then $\tilde{L}$ is ranked and its rank function satisfies (1.8).

First, we show that L is ranked.

Let $X = \{\hat{\mathbf{0}} = x_0 \lessdot x_1 \lessdot \cdots \lessdot x_{n-1} \lessdot x_n = \hat{\mathbf{1}}\}$ be a chain of maximum length. Let $Y = \{\hat{\mathbf{0}} = y_0 \lessdot y_1 \lessdot \cdots \lessdot y_{m-1} \lessdot y_m = \hat{\mathbf{1}}\}$ be any maximal chain in L. We wish to show that $m = n$.

Let $L' = [x_1, \hat{\mathbf{1}}]$ and $L'' = [y_1, \hat{\mathbf{1}}]$. (See Figure 1.6.) By induction, these sublattices are both ranked. Moreover, $c(L') = n - 1$. If $x_1 = y_1$ then Y and X are both saturated chains in the ranked lattice L' and we are done, so suppose that $x_1 \neq y_1$. Let $z_2 = x_1 \vee y_1$. By (1.9), z_2 covers both x_1 and y_1. Let $z_2, \ldots, \hat{\mathbf{1}}$ be a saturated chain in L (thus, in $L' \cap L''$).

Since L' is ranked and $z \gtrdot x_1$, the chain $z_1, \ldots, \hat{\mathbf{1}}$ has length $n-2$. So the chain $y_1, z_1, \ldots, \hat{\mathbf{1}}$ has length $n-1$.

On the other hand, L'' is ranked and $y_1, y_2, \ldots, \hat{\mathbf{1}}$ is a saturated chain, so it also has length $n-1$. Therefore the chain $\hat{\mathbf{0}}, y_1, \ldots, \hat{\mathbf{1}}$ has length n as desired.

Second, we show that the rank function r of L satisfies (1.8). Let $x, y \in L$ and take a saturated chain

$$x \wedge y = c_0 \lessdot c_1 \lessdot \cdots \lessdot c_{n-1} \lessdot c_n = x.$$

[7] Recall that the *length* of a saturated chain is the number of minimal relations in it, which is one less than its cardinality as a subset of L. For example, $c(2^{[n]}) = n$, not $n + 1$.

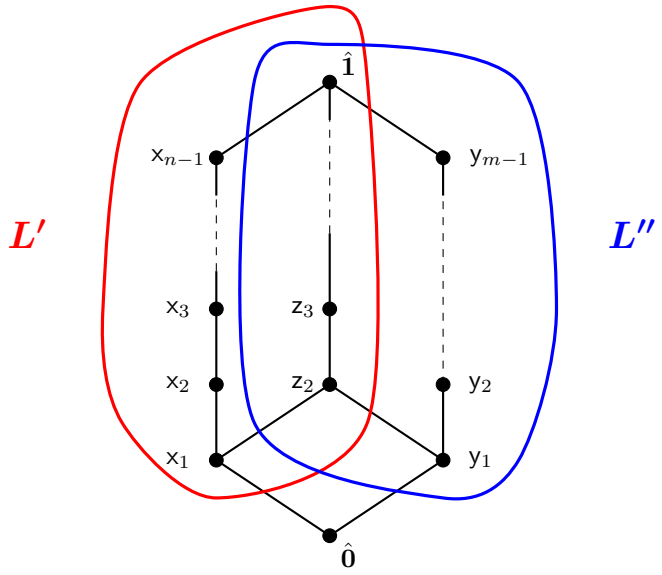


Figure 1.6: A semimodular lattice.

Note that $n = r(x) - r(x \wedge y)$. Then there is a chain

$$y = c_0 \vee y \leq c_1 \vee y \leq \cdots \leq c_n \vee y = x \vee y.$$

By Lemma 1.5.4, each $\leq$ in this chain is either an equality or a covering relation. Therefore, the *distinct* elements $c_i \vee y$ form a saturated chain from y to $x \vee y$, whose length must be $\leq n$. Hence

$$r(x \vee y) - r(y) \leq n = r(x) - r(x \wedge y)$$

which implies the submodular inequality (1.8). □

The same argument shows that L is lower semimodular if and only if it is ranked, with a rank function satisfying the reverse inequality of (1.8).

Theorem 1.5.6. *L is modular if and only if it is ranked, with rank function r satisfying the* ***modular equality***

$$r(x \vee y) + r(x \wedge y) = r(x) + r(y) \qquad \forall x, y \in L. \tag{1.10}$$

Proof. If L is modular, then it is both upper and lower semimodular, so the conclusion follows by Theorem 1.5.5. On the other hand, suppose that L is a lattice whose rank function r satisfies (1.10). Let $x \leq z \in L$. We already know that $x \vee (y \wedge z) \leq (x \vee y) \wedge z$, so it suffices to show that these two elements have the same rank. Indeed,

$$\begin{aligned} r(x \vee (y \wedge z)) &= r(x) + r(y \wedge z) - r(x \wedge y \wedge z) \\ &= r(x) + r(y \wedge z) - r(x \wedge y) \\ &= r(x) + r(y) + r(z) - r(y \vee z) - r(x \wedge y) \end{aligned}$$

and

$$\begin{aligned} r((x \vee y) \wedge z) &= r(x \vee y) + r(z) - r(x \vee y \vee z) \\ &= r(x \vee y) + r(z) - r(y \vee z) \\ &= r(x) + r(y) - r(x \wedge y) + r(z) - r(y \vee z). \end{aligned}$$

□

1.6 Geometric lattices

The following construction gives the prototype of a geometric lattice. Let $\Bbbk$ be a field, let V be a vector space over $\Bbbk$, and let E be a finite subset of V (with repeated elements allowed). We may as well assume that E spans V, so in particular $\dim V < \infty$. Say that a **flat** is a subset of E of the form $W \cap E$, where $W \subseteq E$ is a vector subspace. Define the **vector lattice of E** as

$$L(E) = \{W \cap E \colon W \subseteq V \text{ is a vector subspace}\}. \tag{1.11}$$

Then $L(E)$ is a subposet of Bool_E. Moreover,

$$L(E) \cong \{\Bbbk A \colon A \subseteq E\}. \tag{1.12}$$

(where $\Bbbk A$ denotes the vector subspace of V generated by A), via the map sending $A \mapsto \Bbbk A$. (Note that different subspaces of W can have the same intersection with E, and different subsets of E can span the same vector space.) The poset $L(E)$ is easily checked to be a lattice under the operations

$$(W \cap E) \wedge (X \cap E) = (W \cap X) \cap E, \qquad (W \cap E) \vee (X \cap E) = (W + X) \cap E.$$

The elements of $L(E)$ are called **flats**. Certainly $E = V \cap E$ is a flat, hence the top element of $L(E)$. The bottom element is $O \cap E$, where $O \subseteq V$ is the zero subspace; thus $O \cap E$ consists of the copies of the zero vector in E.

The tricky thing about the isomorphism (1.12) is that it is not so obvious which elements of E are flats. For every $A \subseteq E$, there is a unique minimal flat containing A, namely $\bar{A} := \Bbbk A \cap E$ — that is, the set of elements of E in the linear span of A. On the other hand, if $v, w, x \in E$ with $v + w = x$, then $\{v, w\}$ is not a flat, because any vector subspace that contains both v and w must also contain x. So, an equivalent definition of "flat" is that $A \subseteq E$ is a flat if no vector in $E \setminus A$ is in the linear span of the vectors in A.

The lattice $L(E)$ is ranked, with rank function $r(A) = \dim \Bbbk A$. It is upper semimodular (Problem 1.17) but is not in general modular (see Example 1.6.3 below). On the other hand, $L(E)$ is always an *atomic* lattice: every element is the join of atoms. This is a consequence of the simple fact that $\Bbbk\langle v_1, \ldots, v_k\rangle = \Bbbk v_1 + \cdots + \Bbbk v_k$. This motivates the following definition:

Definition 1.6.1. A lattice L is **geometric** if it is (upper) semimodular and atomic. If $L \cong L(E)$ for some set of vectors E, we say that E is a **(linear) representation** of L.

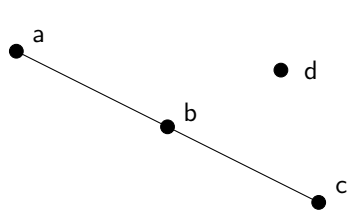
For example, the set $E = \{(0,1), (1,0), (1,1)\} \subseteq \mathbb{F}_2^2$ is a linear representation of the geometric lattice M_5. (For that matter, so is any set of three nonzero vectors in a two-dimensional space over any field, provided none is a scalar multiple of another.)

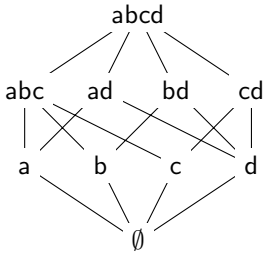
A closely related construction is the **affine lattice of E**, defined by

$$L^{\text{aff}}(E) = \big\{W \cap E \colon W \subseteq V \text{ is an affine subspace}\big\}.$$

(An affine subspace of V is a translate of a vector subspace; for example, a line or plane not necessarily containing the origin.) In fact, any lattice of the form $L^{\text{aff}}(E)$ can be expressed in the form $L(\hat{E})$, where $\hat{E}$ is a certain point set constructed from E (homework problem). However, the dimension of the affine span of a set $A \subseteq E$ is one less than its rank — which means that we can draw geometric lattices of rank 3 conveniently as planar point configurations. If $L \cong L^{\text{aff}}(E)$, we could say that E is a **(affine) representation** of L.

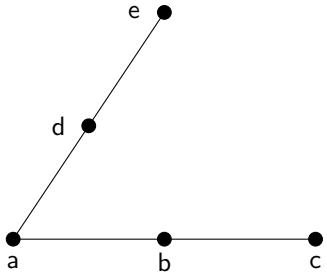
Example 1.6.2. Let $E = \{\mathsf{a}, \mathsf{b}, \mathsf{c}, \mathsf{d}\}$, where $\mathsf{a}, \mathsf{b}, \mathsf{c}$ are collinear but no other set of three points is. Then $L^{\text{aff}}(E)$ is the lattice shown below (which happens to be modular).

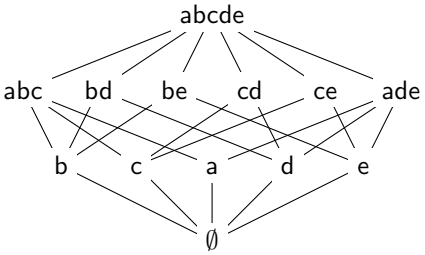




◀

Example 1.6.3. If E is the point configuration on the left with the only collinear triples $\{\mathsf{a},\mathsf{b},\mathsf{c}\}$ and $\{\mathsf{a},\mathsf{d},\mathsf{e}\}$, then $L^{\text{aff}}(E)$ is the lattice on the right.





This lattice is not modular: consider the two elements bd and ce. ◀

Example 1.6.4. Recall from Example 1.5.2 that the partition lattice Π_n is USM for all n. In fact it is geometric. To see that it is atomic, observe that the atoms are the set partitions with $n-1$ blocks, necessarily one doubleton block and $n-2$ singletons; let π_{ij} denote the atom whose doubleton block is $\{i,j\}$. Then every set partition σ is the join of the set $\{\pi_{ij} : i \sim_\sigma j\}$.

In fact, Π_n is a vector lattice. Let $\Bbbk$ be any field, llet $\{\mathbf{e}_1,\ldots,\mathbf{e}_n\}$ be the standard basis of $V = \Bbbk^n$, let $\mathbf{p}_{ij} = \mathbf{e}_i - \mathbf{e}_j$ for all $1 \le i < j \le n$, and let E be the set of all such vectors $\mathbf{p}_{ij}$.. Then in fact $\Pi_n \cong L(E)$. The atoms π_{ij} of Π_n correspond to the atoms $\Bbbk\langle\mathbf{p}_{ij}\rangle$ of $L(E)$; the rest of the isomorphism is left as Problem 1.18. Note that this construction works over any field $\Bbbk$.

More generally, if G is any simple graph on vertex set $[n]$ then the connectivity lattice $K(G)$ is isomorphic to $L(E_G)$, where $E_G = \{\mathbf{a}_{ij} : ij \text{ is an edge of } G\}$. ◀

1.7 Exercises

Posets

Problem 1.1. (a) Prove that every nonempty interval in a Boolean lattice is itself isomorphic to a Boolean lattice.
(b) Prove that every interval in the subspace lattice $L_n(q)$ is isomorphic to a subspace lattice.
(c) Prove that every interval in the partition lattice Π_n is isomorphic to a product of partition lattices. (The **product** of posets $P_1,\ldots,P_k$ is the Cartesian product $P_1 \times \cdots \times P_k$, equipped with the partial order $(x_1,\ldots,x_k) \le (y_1,\ldots,y_k)$ if $x_i \le_{P_i} y_i$ for all $i \in [k]$.)

Problem 1.2. Let P and Q be posets. Describe $\mathscr{L}(P+Q)$ in terms of $\mathscr{L}(P)$ and $\mathscr{L}(Q)$, and give a formula for $|\mathscr{L}(P+Q)|$.

Problem 1.3. Let n be a positive integer. Let D_n be the set of all positive-integer divisors of n (including n itself), partially ordered by divisibility.

(a) Prove that D_n is a ranked poset, and describe the rank function.
(b) For which values of n is D_n (i) a chain; (ii) a Boolean lattice? For which values of n, m is it the case that $D_n \cong D_m$?
(c) Prove that D_n is a distributive lattice. Describe its meet and join operations and its join-irreducible elements.
(d) Prove that D_n is *self-dual*, i.e., there is a bijection $f : D_n \to D_n$ such that $f(x) \leq f(y)$ if and only if $x \geq y$.

Problem 1.4. Let G be a graph on vertex set $V = [n]$. Recall from Example 1.2.3 that the *connectivity lattice* of a graph is the subposet $K(G)$ of Π_n consisting of set partitions in which every block induces a connected subgraph of G. Prove that $K(G)$ is a lattice. Is it a sublattice of Π_n?

Problem 1.5. Let $\mathcal{A}$ be a finite family of sets. For $\mathcal{A}' \subseteq \mathcal{A}$, define $\cup\mathcal{A}' = \bigcup_{A \in \mathcal{A}'} A$. Let $U(\mathcal{A}) = \{\cup\mathcal{A}' \colon \mathcal{A}' \subseteq \mathcal{A}\}$, considered as a poset ordered by inclusion.

(a) Prove that $U(\mathcal{A})$ is a lattice. (Hint: Don't try to specify the meet operation explicitly.)
(b) Construct a set family $\mathcal{A}$ such that $U(\mathcal{A})$ is isomorphic to weak Bruhat order on $\mathfrak{S}_3$ (see Example 2.11).
(c) Construct a set family $\mathcal{A}$ such that $U(\mathcal{A})$ is not ranked.
(d) Is every finite lattice of this form?

Problem 1.6. For $1 \leq i \leq n-1$, let s_i be the transposition in $\mathfrak{S}_n$ that swaps i with $i+1$. (The s_i are called *elementary transpositions*.) You probably know that $\{s_1, \ldots, s_{n-1}\}$ is a generating set for $\mathfrak{S}_n$ (and if you don't, you will shortly prove it). For $w \in \mathfrak{S}_n$, an expression $w = s_{i_1} \cdots s_{i_k}$ is called a **reduced word** if there is no way to express w as a product of fewer than k generators.

(a) Show that every reduced word for w has length equal to $\mathrm{inv}(w)$ (as defined in (1.2)).
(b) Define a partial order $\prec$ on $\mathfrak{S}_n$ as follows: $w \prec v$ if there exists a reduced word $s_{i_1} \cdots s_{i_k}$ for v such that w is the product of some proper subword $w = s_{i_{j_1}} \cdots s_{i_{j_\ell}}$. (Sorry about the triple subscripts; this just means that v is obtained by deleting some of the letters from the reduced word for w.) Prove that if $w \prec v$, then $w < v$ in Bruhat order. (The converse is true but requires significantly more work; see [BB05], in particular Theorems 1.4.3 and 2.2.2.)

Problem 1.7. Prove that the rank-generating functions of weak order and Bruhat order on $\mathfrak{S}_n$ are both

$$\prod_{i=1}^{n} \frac{1-q^i}{1-q}.$$

(Hint: Induct on n, and use one-line notation for permutations, not cycle notation.)

Distributive lattices

Problem 1.8. Prove that the two formulations (1.3a) and (1.3b) of distributivity of a lattice L are equivalent, i.e.,

$$x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z) \quad \forall x, y, z \in L \qquad \Longleftrightarrow \qquad x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z) \quad \forall x, y, z \in L.$$

Problem 1.9. In Problem 1.3 you proved that the divisor lattice D_n is distributive. Characterize all posets P such that $J(P) \cong D_n$ for some $n \in \mathbb{N}$. (In other words, prove a statement of the form "A distributive lattice $L = J(P)$ is isomorphic to a divisor lattice if and only if the poset $P = \mathrm{Irr}(L)$ is _____.")

Problem 1.10. Let L be a finite lattice and $x \in L$. Prove that x is join-irreducible if it covers exactly one other element. What weaker conditions than "finite" suffice?

Problem 1.11. Let Y be Young's lattice (which we know is distributive).

(a) Describe the join-irreducible elements of Young's lattice Y.
(b) Let $\lambda \in Y$. If $\lambda = \mu_1 \vee \cdots \vee \mu_k$ is an irredundant factorization, then what quantity does k correspond to in the Ferrers diagram of λ?
(c) Let λ be a $2 \times n$ rectangle. Show that the number of maximal chains in the interval $[\emptyset, \lambda] \subseteq Y$ is the Catalan number C_n.
(d) Count the maximal chains in the interval $[\emptyset, \lambda] \subseteq Y$ if λ is a hook shape (i.e., $\lambda = (n+1, 1, 1, \ldots, 1)$, with a total of m copies of 1).

Problem 1.12. Fill in the details in the proof of the FTFDL (Theorem 1.3.7) by showing the following facts.

(a) For a finite distributive lattice L, show that the map $\phi : L \to J(\mathrm{Irr}(L))$ given by

$$\phi(x) \;=\; \langle p \colon p \in \mathrm{Irr}(L),\ p \leq x \rangle$$

is indeed a lattice isomorphism.
(b) For a finite poset P, show that an order ideal in P is join-irreducible in $J(P)$ if and only if it is principal (i.e., generated by a single element).

Problem 1.13. Let L be a sublattice of Bool_n that is *accessible*: if $S \in L \setminus \{\emptyset\}$ then there exists some $x \in S$ such that $S \setminus \{x\} \in L$. Construct a poset P on $[n]$ such that $J(P) = L$. (Notice that I wrote "$= L$", not "$\cong L$." It is not enough to invoke Birkhoff's theorem to say that such a P must exist! The point is to explicitly construct a poset P on $[n]$ whose order ideals *are* the sets in L.)

Modular lattices

Problem 1.14. Let $L_n(q)$ be the poset of subspaces of an n-dimensional vector space over the finite field $\mathbb{F}_q$ (so $L_n(q)$ is a modular lattice by Corollary 1.4.3).

(a) Prove directly from the definition of modularity that $L_n(q)$ is modular. (I.e., verify algebraically that the join and meet operations obey the modular equation (1.5).)
(b) Prove the assertion in Example 1.2.5 that the number of k-dimensional subspaces of $\mathbb{F}_q^n$ is $\begin{bmatrix} n \\ k \end{bmatrix}_q$. Hint: Every vector space of dimension k is determined by an ordered basis $v_1, \ldots, v_k$. How many ordered bases does each k-dimensional vector space $V \in L_n(q)$ have? How many sequences of vectors in $\mathbb{F}_q^n$ are ordered bases for some k-dimensional subspace?
(c) Count the maximal chains in $L_n(q)$.

Problem 1.15. Verify that the lattice Π_4 is not modular.

Semimodular and geometric lattices

Problem 1.16. Let L be a lattice with the following property: for all $x, y \in L$, if $x \wedge y$ is covered by both x and y, then $x \vee y$ covers x and y. Prove that L is upper semimodular. (Obviously upper-semimodular lattices have this property, so this exercise provides an alternative definition of upper semimodularity.)

Problem 1.17. Prove that the lattice $L(E)$ defined in (1.11) is upper semimodular.

Problem 1.18. Prove that the lattices Π_n and $L(E)$ are isomorphic, where E is the vector set described in Example 1.6.4. To do this, you need to characterize the vector spaces spanned by subsets of $A \subseteq E$ and show that they are in bijection with set partitions. (Hint: It may be useful to look at the orthogonal complements of those vector spaces under the standard inner product on $\Bbbk^n$.)

Problem 1.19. The purpose of this exercise is to show that the constructions L and L^{aff} produce the same class of lattices. Let $\Bbbk$ be a field and let $E = \{e_1, \ldots, e_n\} \subseteq \Bbbk^d$.

(a) The *augmentation* of a vector $e_i = (e_{i1}, \ldots, e_{id})$ is the vector $\tilde{e}_i = (1, e_{i1}, \ldots, e_{id}) \in \Bbbk^{d+1}$. Prove that $L^{\text{aff}}(E) = L(\tilde{E})$, where $\tilde{E} = \{\tilde{e}_1, \ldots, \tilde{e}_n\}$.

(b) Let $\mathbf{v}$ be a vector in $\Bbbk^d$ that is not a scalar multiple of any e_i, let H Let $H \subseteq \Bbbk^d$ be a generic affine hyperplane, let $\hat{e}_i$ be the projection of e_i onto H, and let $\hat{E} = \{\hat{e}_1, \ldots, \hat{e}_n\}$. Prove that $L(E) = L^{\text{aff}}(\hat{E})$. (The first part is figuring out what "generic" means. A generic hyperplane might not exist for all fields, but if $\Bbbk$ is infinite then almost all hyperplanes are generic.)

Problem 1.20. Recall from Corollary 1.3.9 that a lattice L is *relatively complemented* if, whenever $y \in [x, z] \subseteq L$, there exists $u \in [x, z]$ such that $y \wedge u = x$ and $y \vee u = z$. Prove that a finite semimodular lattice is atomic (hence geometric) if and only if it is relatively complemented.

(Here is the geometric interpretation of being relatively complemented. Suppose that V is a vector space, $L = L(E)$ for some point set $E \subseteq V$, and that $X \subseteq Y \subseteq Z \subseteq V$ are vector subspaces spanned by flats of $L(E)$. For starters, consider the case that $X = O$. Then we can choose a basis B of the space Y and extend it to a basis B' of Z, and the vector set $B' \setminus B$ spans a subspace of Z that is complementary to Y. More generally, if X is any subspace, we can choose a basis B for X, extend it to a basis B' of Y, and extend B' to a basis B'' of Z. Then $B \cup (B'' \setminus B')$ spans a subspace $U \subseteq Z$ that is relatively complementary to Y, i.e., $U \cap Y = X$ and $U + Y = Z$.)

Chapter 2

Poset Algebra

Throughout this chapter, every poset we consider will be assumed to be **locally finite**, i.e., every interval is finite.

2.1 The incidence algebra of a poset

Let P be a poset and let $\mathrm{Int}(P)$ denote the set of (nonempty) intervals of P. Recall that an interval is a subset of P of the form $[x, y] := \{z \in P : x \leq z \leq y\}$; if $x \not\leq y$ then $[x, y] = \emptyset$.

Definition 2.1.1. The **incidence algebra** $I(P)$ is the set of functions $\alpha : \mathrm{Int}(P) \to \mathbb{C}$ ("incidence functions")[1], made into a $\mathbb{C}$-vector space with pointwise addition, subtraction and scalar multiplication. It is equivalent to think of $I(P)$ as the set of functions on $P \times P$, with $\alpha(x, y) = 0$ if $x \not\leq y$ — this lets us write $\alpha(x, y)$ instead of the more awkward $\alpha([x, y])$. We make $I(P)$ into a ring with the **convolution product**:

$$(\alpha * \beta)(x, y) = \sum_{z \in [x,y]} \alpha(x, z)\beta(z, y).$$

Note that the assumption of local finiteness is both necessary and sufficient for convolution to be well-defined for all incidence functions.

Proposition 2.1.2. *Convolution is associative (although it is not in general commutative).*

[1] More generally, we could allow incidence functions to take values in any (commutative) ring.

Proof. The basic idea is to reverse the order of summation:

$$\begin{aligned}
[(\alpha * \beta) * \gamma](x,y) &= \sum_{z \in [x,y]} (\alpha * \beta)(x,z) \cdot \gamma(z,y) \\
&= \sum_{z \in [x,y]} \left(\sum_{w \in [x,z]} \alpha(x,w)\beta(w,z) \right) \gamma(z,y) \\
&= \sum_{w,z:\ x \le w \le z \le y} \alpha(x,w)\beta(w,z)\gamma(z,y) \\
&= \sum_{w \in [x,y]} \alpha(x,w) \left(\sum_{z \in [w,y]} \beta(w,z)\gamma(z,y) \right) \\
&= \sum_{w \in [x,y]} \alpha(x,w) \cdot (\beta * \gamma)(w,y) \\
&= [\alpha * (\beta * \gamma)](x,y).
\end{aligned}$$

□

The ring $I(P)$ has a multiplicative identity, namely the Kronecker delta function, regarded as an incidence function:

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

Therefore, we sometimes write 1 for δ.

Once you know a ring has a multiplicative identity, the next natural question is which elements are invertible. This question has a nice answer:

Proposition 2.1.3. *An incidence function $\alpha \in I(P)$ has a left/right/two-sided convolution inverse if and only if $\alpha(x,x) \neq 0$ for all x (the "nonzero condition"). In that case, the inverse is given by the recursive formula*

$$\alpha^{-1}(x,y) = \begin{cases} \alpha(x,x)^{-1} & \text{if } x = y, \\ -\alpha(y,y)^{-1} \sum\limits_{z:\ x \le z < y} \alpha^{-1}(x,z)\alpha(z,y) & \text{if } x < y. \end{cases} \tag{2.1}$$

This formula is well-defined by induction on the size of $[x,y]$, with the cases $x = y$ and $x \neq y$ serving as the base case and inductive step respectively.

Proof. Let β be a left convolution inverse of α. In particular, $\alpha(x,x) = \beta(x,x)^{-1}$ for all x (use the equation $(\alpha * \beta)(x,x) = \delta(x,x) = 1$), so the nonzero condition is necessary.

On the other hand, if $x < y$, then

$$(\beta * \alpha)(x,y) = \sum_{z \in [x,y]} \beta(x,z)\alpha(z,y) = \delta(x,y) = 0$$

and solving for $\beta(x,y)$ (by pulling the $z = y$ term out of the sum) gives the formula (2.1), which is well-defined provided that $\alpha(y,y) \neq 0$. So the nonzero condition is also sufficient.

A similar argument shows that the nonzero condition is necessary and sufficient for α to have a right convolution inverse. Moreover, the left and right inverses coincide: if $\beta * \alpha = \delta = \alpha * \gamma$ then $\beta = \beta * \delta = \beta * \alpha * \gamma = \gamma$ by associativity. □

Now we have a ring in which algebraic identities can encode facts about the poset P. We need some interesting incidence functions to play with. The **zeta function** and **eta function** of P are defined as

$$\zeta(x,y) = \begin{cases} 1 & \text{if } x \le y, \\ 0 & \text{if } x \not\le y, \end{cases} \qquad \eta(x,y) = \begin{cases} 1 & \text{if } x < y, \\ 0 & \text{if } x \not< y, \end{cases}$$

i.e., $\eta = \zeta - 1 = \zeta - \delta$. Note that ζ is invertible and η is not.

These trivial-looking incidence functions are useful because their convolution powers count important things, namely multichains and chains in P. In other words, enumerative questions about posets can be expressed algebraically. Specifically,

$$\begin{aligned} \zeta^2(x,y) &= \sum_{z\in[x,y]} \zeta(x,z)\zeta(z,y) = \sum_{z\in[x,y]} 1 \\ &= \#\{z:\ x \le z \le y\}, \\ \zeta^3(x,y) &= \sum_{z\in[x,y]} \sum_{w\in[z,y]} \zeta(x,z)\zeta(z,w)\zeta(w,y) = \sum_{x\le z\le w\le y} 1 \\ &= \#\{(z,w):\ x \le z \le w \le y\}, \\ \zeta^k(x,y) &= \#\{(x_1,\dots,x_{k-1}):\ x \le x_1 \le x_2 \le \cdots \le x_{k-1} \le y\}. \end{aligned}$$

That is, $\zeta^k(x,y)$ counts the number of **multichains** of length k between x and y (chains with possible repeats). If we replace ζ with η, then the calculations all work the same way, except that all the $\le$'s are replaced with $<$'s, so we get

$$\eta^k(x,y) = \#\{(x_1,\dots,x_{k-1}):\ x < x_1 < x_2 < \cdots < x_{k-1} < y\},$$

the number of *chains* of length k (not necessarily saturated) between x and y. In particular, if the chains of P are bounded in length (e.g., if P is finite), then $\eta^n = 0$ for $n \gg 0$.

Direct products of posets play nicely with the incidence algebra construction. Specifically, let P, Q be bounded finite posets. For $\alpha \in I(P)$ and $\phi \in I(Q)$, define $\alpha\phi \in I(P\times Q)$ by

$$\alpha\phi[(x,x'),(y,y')] = \alpha(x,y)\phi(x',y').$$

This defines a linear transformation $F : I(P)\otimes I(Q) \to I(P\times Q)$. [2] In other words, $(\alpha+\beta)\phi = \alpha\phi + \beta\phi$, and $\alpha(\phi+\psi) = \alpha\phi+\alpha\psi$, and $\alpha(c\phi) = (c\alpha)\phi = c(\alpha\phi)$ for all $c \in \mathbb{C}$. It is actually a vector space isomorphism, because there is a bijection $\mathrm{Int}(P)\times\mathrm{Int}(Q) \to \mathrm{Int}(P\times Q)$ given by $(I,J) \to I\times J$, and $F(\chi_I\otimes\chi_J) = \chi_{I\times J}$ (where χ_I is the characteristic function of I, i.e., the incidence function that is 1 on I and zero on other intervals). In fact, more is true:

Proposition 2.1.4. *The map F just defined is a ring isomorphism. That is, for all $\alpha,\beta \in I(P)$ and $\phi,\psi \in I(Q)$,*

$$\alpha\phi * \beta\psi = (\alpha*\beta)(\phi*\psi).$$

Furthermore, the incidence functions δ and ζ are multiplicative on direct products, i.e.,

$$\delta_{P\times Q} = \delta_P\delta_Q \quad \textit{and} \quad \zeta_{P\times Q} = \zeta_P\zeta_Q.$$

[2] See §8.5 for an extremely brief introduction to the tensor product operation $\otimes$.

Proof. Let (x, x') and (y, y') be elements of $P \times Q$. Then

$$\begin{aligned}(\alpha\phi * \beta\psi)[(x,x'),(y,y')] &= \sum_{(z,z')\in[(x,x'),(y,y')]} \alpha\phi[(x,x'),(z,z')] \cdot \beta\psi[(z,z'),(y,y')] \\ &= \sum_{z\in[x,y]} \sum_{z'\in[x',y']} \alpha(x,z)\phi(x',z')\beta(z,y)\psi(z',y') \\ &= \left[\sum_{z\in[x,y]} \alpha(x,z)\beta(z,y)\right]\left[\sum_{z'\in[x',y']} \phi(x',z')\psi(z',y')\right] \\ &= (\alpha * \beta(x,y)) \cdot (\phi * \psi(x',y')).\end{aligned}$$

Multiplicativity of δ and ζ is immediate from their definitions. □

2.2 The Möbius function

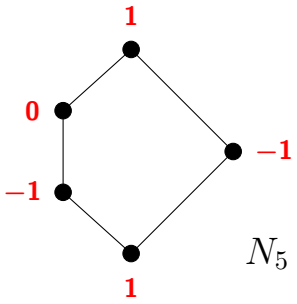
The **Möbius function** μ_P of a poset P is defined as the convolution inverse of its zeta function: $\mu_P = \zeta_P^{-1}$. This turns out to be one of the most important incidence functions on a poset. For a bounded poset, we abbreviate $\mu_P(x) = \mu_P(\hat{\mathbf{0}}, x)$ and $\mu(P) = \mu_P(\hat{\mathbf{0}}, \hat{\mathbf{1}})$. Proposition 2.1.3 provides a recursive formula for μ:

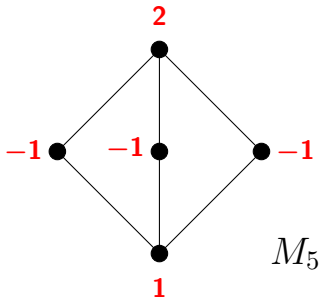
$$\mu(x,y) = \begin{cases} 0 & \text{if } y \not\geq x \text{ (i.e., if } [x,y] = \emptyset), \\ 1 & \text{if } y = x, \\ -\sum_{z:\ x\leq z<y} \mu(x,z) & \text{if } x < y. \end{cases} \tag{2.2}$$

This is equivalent to the familiar recursive formula: to find $\mu_P(x)$, add up the values of μ_P at all elements $< x$, then change the sign.

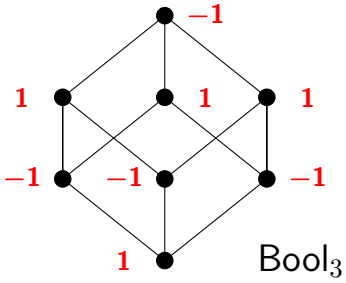
Example 2.2.1. If $P = \{0 < 1 < 2 < \cdots\}$ is a chain, then its Möbius function is given by $\mu(x,x) = 1$, $\mu(x, x+1) = -1$, and $\mu(x,y) = 0$ otherwise. ◀

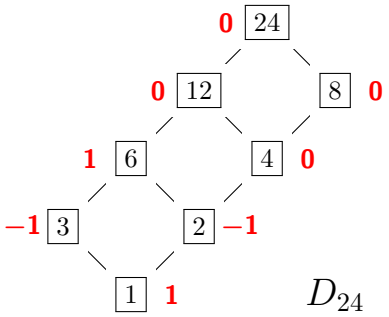
Example 2.2.2. Here are the Möbius functions $\mu_P(x) = \mu_P(\hat{\mathbf{0}}, x)$ for the lattices N_5 and M_5:





And here are the Boolean lattice Bool_3 and the divisor lattice D_{24}:





◀

Example 2.2.3 (Möbius functions of partition lattices)**.** What is $\mu(\Pi_n)$ in terms of n? Clearly $\mu(\Pi_1) = 1$ and $\mu(\Pi_2) = -1$, and $\mu(\Pi_3) = \mu(M_5) = 2$. For $n = 4$, we calculate $\mu(\Pi_4)$ from (2.2). The value of $\mu_{\Pi_4}(\hat{\mathbf{0}}, \pi)$ depends only on the block sizes of π, in fact, $[\hat{\mathbf{0}}, \pi] \cong \Pi_{\pi_1} \times \cdots \times \Pi_{\pi_k}$. We will use the fact that the Möbius function is multiplicative on direct products; we will prove this shortly (Prop. 2.2.5).

Block sizes	Number of π's	Isomorphism type of $[\hat{\mathbf{0}}, \pi]$	$\mu(\hat{\mathbf{0}}, \pi)$
1,1,1,1	1	Π_1	1
2,1,1	6	Π_2	-1
2,2	3	$\Pi_2 \times \Pi_2$	1
3,1	4	Π_3	2

Therefore, $\mu(\Pi_4) = -(1 \cdot 1 - 1 \cdot 6 + 1 \cdot 3 + 2 \cdot 4) = -6$. Let's try $n = 5$:

Block sizes	Number of π's	$\mu(\hat{\mathbf{0}}, \pi)$	Contribution to $-\mu(\Pi_5)$
1,1,1,1,1	1	$\mu(\Pi_1) = 1$	1
2,1,1,1	10	$\mu(\Pi_2) = -1$	-10
2,2,1	15	$\mu(\Pi_2 \times \Pi_2) = 1$	15
3,1,1	10	$\mu(\Pi_3) = 2$	20
3,2	10	$\mu(\Pi_3 \times \Pi_2) = -2$	-20
4,1	5	$\mu(\Pi_4) = -6$	-30

Adding up the last column and multiplying by -1 gives $\mu(\Pi_5) = 24$. At this point you might guess that $\mu(\Pi_n) = (-1)^{n-1}(n-1)!$, and you would be right. We will prove this soon. ◀

The Möbius function is useful in many ways. It can be used to formulate a more general version of inclusion-exclusion called *Möbius inversion*. It behaves nicely under poset operations such as product, and has geometric and topological applications. Even just the single number $\mu(P) = \mu_P(\hat{\mathbf{0}}, \hat{\mathbf{1}})$ tells you a lot about a bounded poset P. Confusingly, this number itself is sometimes called the "Möbius function" of P (I prefer "Möbius number" to avoid ambiguity). Here is the reason.

Definition 2.2.4. A family $\mathscr{F}$ of posets is **hereditary** if, for each $P \in \mathscr{F}$, every interval in P is isomorphic to some [other] poset in $\mathscr{F}$. It is **semi-hereditary** if every interval in a member of $\mathscr{F}$ is isomorphic to a product of members of $\mathscr{F}$.

For example, the families of Boolean lattices, divisor lattices, and subspace lattices are all hereditary, and the family of partition lattices is semi-hereditary (Problem 1.1). Knowing the Möbius number for every

poset in a hereditary family is equivalent to knowing their full Möbius functions. The same is true for semi-hereditary families, for the following reason.

Proposition 2.2.5. *The Möbius function is multiplicative on direct products, i.e., $\mu_{P\times Q} = \mu_P\mu_Q$ (in the notation of Proposition 2.1.4).*

Proof.

$$\zeta_{P\times Q} * \mu_P\mu_Q = \zeta_P\zeta_Q * \mu_P\mu_Q = (\zeta_P * \mu_P)(\zeta_Q * \mu_Q) = \delta_P\delta_Q = \delta_{P\times Q}$$

which says that $\mu_P\mu_Q = \zeta_{P\times Q}^{-1} = \mu_{P\times Q}$. Here the second equality is the definition of multiplication in a tensor product of rings. (It is also possible to prove that $\mu_P\mu_Q = \mu_{P\times Q}$ directly from the definition; this is Problem 2.2.) □

For example, the interval $[1|2|3|4|5|678|9,\ 123|45|6789] \subseteq \Pi_9$ is isomorphic to $\Pi_3 \times \Pi_2 \times \Pi_2$, so its Möbius number is $\mu(\Pi_3)\mu(\Pi_2)^2$.

Since $\mu(\mathsf{Bool}_1) = -1$ and Bool_n is a product of n copies of Bool_1, an immediate consequence of Proposition 2.2.5 is the formula

$$\mu(\mathsf{Bool}_n) = (-1)^n.$$

This can also be proved by induction on n (with the cases $n = 0$ and $n = 1$ easy). If $n > 0$, then

$$\begin{aligned}\mu(\mathsf{Bool}_n) = -\sum_{A\subsetneq[n]} (-1)^{|A|} &= -\sum_{k=0}^{n-1}(-1)^k\binom{n}{k} \quad \text{(by induction)}\\ &= (-1)^n - \sum_{k=0}^{n}(-1)^k\binom{n}{k}\\ &= (-1)^n - (1-1)^n \;=\; (-1)^n.\end{aligned}$$

In particular, the full Möbius function of the Boolean lattice Bool_S is given by $\mu(A, B) = \mu(\mathsf{Bool}_{|B\setminus A|}) = (-1)^{|B\setminus A|}$ for all $A \subseteq B \subseteq S$.

Example 2.2.6. Let P be a product of k chains of lengths $a_1, \ldots, a_k$. Equivalently,

$$P = \{\mathbf{x} = (x_1, \ldots, x_k) \colon 0 \le x_i \le a_i \text{ for all } i \in [k]\},$$

ordered by $\mathbf{x} \le \mathbf{y}$ iff $x_i \le y_i$ for all i. (Recall that the length of a chain is the number of covering relations, which is one less than the number of elements; see Definition 1.1.6.) Then Prop. 2.2.5 together with the formula for the Möbius function of a chain (above) gives

$$\mu(\hat{\mathbf{0}}, \mathbf{x}) = \begin{cases} 0 & \text{if } x_i \ge 2 \text{ for at least one } i;\\ (-1)^s & \text{if } \mathbf{x} \text{ consists of } s \text{ 1's and } k-s \text{ 0's.}\end{cases}$$

(The Boolean lattice is the special case that $a_i = 1$ for every i.) This conforms to the definition of Möbius function that you may have seen in enumerative combinatorics or number theory, since products of chains are precisely divisor lattices. As mentioned above, the family of divisor lattices is hereditary: $[a, b] \cong D_{b/a}$ for all $a, b \in D_n$ with $a|b$. ◀

Here are a couple of enumerative applications of the Möbius function. The first, known as Philip Hall's Theorem,[3] makes the connection between the Möbius function and topology more explicit.

[3]Not to be confused with the unrelated Hall's Marriage Theorem.

Theorem 2.2.7 (Philip Hall's Theorem). *[Sta12, Prop. 3.8.5] Let P be a finite bounded poset with at least two elements. For $k \geq 1$, let*

$$c_k = c_k(P) = \left|\{(x_0, \ldots, x_k) : \hat{\mathbf{0}} = x_0 < x_1 < \cdots < x_k = \hat{\mathbf{1}}\}\right|,$$

the number of chains of length k between $\hat{\mathbf{0}}$ and $\hat{\mathbf{1}}$. Then

$$\mu_P(\hat{\mathbf{0}}, \hat{\mathbf{1}}) = \sum_k (-1)^k c_k.$$

Proof. Recall that $c_k = \eta^k(\hat{\mathbf{0}}, \hat{\mathbf{1}}) = (\zeta - \delta)^k(\hat{\mathbf{0}}, \hat{\mathbf{1}})$. The trick is to use the geometric series expansion $1/(1+h) = 1 - h + h^2 - h^3 + h^4 - \cdots$. Clearing both denominators and replacing h with η and 1 with δ, we get

$$(\delta + \eta)^{-1} = \left(\sum_{k=0}^{\infty} (-1)^k \eta^k\right).$$

The RHS looks like an infinite power series, but it is actually a polynomial, because $\eta^k = 0$ for k sufficiently large. (here is where we need the assumption that P is finite). So we have a valid equation in $I(P)$ (which you can verify by multiplying $\delta + \eta$ by the RHS). Switching the two sides and evaluating on $[\hat{\mathbf{0}}, \hat{\mathbf{1}}]$ gives

$$\sum_{k=0}^{\infty} (-1)^k c_k \;=\; \sum_{k=0}^{\infty} (-1)^k \eta^k(\hat{\mathbf{0}}, \hat{\mathbf{1}}) \;=\; (\delta + \eta)^{-1}(\hat{\mathbf{0}}, \hat{\mathbf{1}}) \;=\; \zeta^{-1}(\hat{\mathbf{0}}, \hat{\mathbf{1}}) \;=\; \mu(\hat{\mathbf{0}}, \hat{\mathbf{1}}).$$ □

This alternating sum looks like an Euler characteristic (see (6.2) below). In fact it is.

Corollary 2.2.8. *Let P be a finite bounded poset with at least two elements, and let $\Delta(P)$ be its* order complex, *i.e., the simplicial complex (see Example 1.1.12) whose vertices are the elements of $P \setminus \{\hat{\mathbf{0}}, \hat{\mathbf{1}}\}$ and whose simplices are chains. Each chain $x_0 = \hat{\mathbf{0}} < x_1 < \cdots < x_k = \hat{\mathbf{1}}$ gives rise to a simplex $\{x_1, \ldots, x_{k-1}\}$ of $\Delta(P)$ of dimension $k-2$. Hence $f_{k-2}(\Delta(P)) = c_k(P)$ for all $k \geq 1$, and the reduced Euler characteristic of $\Delta(P)$ is*

$$\tilde{\chi}(\Delta(P)) \stackrel{\text{def}}{=} \sum_{k \geq -1} (-1)^k f_k(\Delta(P)) = \sum_{k \geq 1} (-1)^{k-2} c_k(P) = \mu_P(\hat{\mathbf{0}}, \hat{\mathbf{1}}).$$ □

Example 2.2.9. For the Boolean lattice $P = \mathrm{Bool}_3$ (see Example 2.2.2), we have $c_0 = 0$, $c_1 = 1$, $c_2 = 6$, $c_3 = 6$, and $c_k = 0$ for $k > 3$. Indeed, $c_0 - c_1 + c_2 - c_3 = -1 = \mu_P(\hat{\mathbf{0}}, \hat{\mathbf{1}})$. ◀

Corollary 2.2.10. *If P is any finite poset, then $\mu(P) = \mu(P^*)$.*

Proof. This is immediate from Philip Hall's Theorem, since $c_k(P) = c_k(P^*)$ for all k. (One can also prove this fact by comparing the algebras $I(P)$ and $I(P^*)$; see Problem 2.3.) □

2.3 Möbius inversion and the characteristic polynomial

The following result is one of the most frequent applications of the Möbius function.

Theorem 2.3.1 (Möbius inversion formula). *Let P be a locally finite*[4] *poset, let V be any $\mathbb{C}$-vector space (usually, but not always, $\mathbb{C}$ itself) and let $f, g : P \to V$. Then*

$$g(x) = \sum_{y:\, y \le x} f(y) \quad \forall x \in P \qquad \Longleftrightarrow \qquad f(x) = \sum_{y:\, y \le x} \mu(y,x) g(y) \quad \forall x \in P, \tag{2.3a}$$

$$g(x) = \sum_{y:\, y \ge x} f(y) \quad \forall x \in P \qquad \Longleftrightarrow \qquad f(x) = \sum_{y:\, y \ge x} \mu(x,y) g(y) \quad \forall x \in P. \tag{2.3b}$$

Proof. Stanley calls the proof "A trivial observation in linear algebra". Let V be the vector space of functions $f : P \to \mathbb{C}$. Consider the right action $\bullet$ and the left action $\bullet$ of $I(P)$ on V by

$$(f \bullet \alpha)(x) = \sum_{y:\, y \le x} \alpha(y,x) f(y),$$
$$(\alpha \bullet f)(x) = \sum_{y:\, y \ge x} \alpha(x,y) f(y).$$

In terms of these actions, formulas (2.3a) and (2.3b) are respectively just the "trivial" observations

$$g = f \bullet \zeta \qquad \Longleftrightarrow \qquad f = g \bullet \mu, \tag{2.4a}$$
$$g = \zeta \bullet f \qquad \Longleftrightarrow \qquad f = \mu \bullet g. \tag{2.4b}$$

We just have to prove that these colored dots indeed define actions, i.e.,

$$f \bullet (\alpha * \beta) = (f \bullet \alpha) \bullet \beta \quad \text{and} \quad (\alpha * \beta) \bullet f = \alpha \bullet (\beta \bullet f).$$

We prove the first identity:

$$\begin{aligned}
(f \bullet (\alpha * \beta))(y) &= \sum_{x:\, x \le y} (\alpha * \beta)(x,y) f(x) \\
&= \sum_{x:\, x \le y} \sum_{z:\, z \in [x,y]} \alpha(x,z) \beta(z,y) f(x) \\
&= \sum_{z:\, z \le y} \left(\sum_{x:\, x \le z} \alpha(x,z) f(x) \right) \beta(z,y) \\
&= \sum_{z:\, z \le y} (f \bullet \alpha)(z) \beta(z,y) \quad = \quad ((f \bullet \alpha) \bullet \beta)(y).
\end{aligned}$$

and the other verification is a mirror image of this one. □

In the case $P = 2^{[n]}$, Möbius inversion says that

$$g(x) = \sum_{B \subseteq A} f(B) \quad \forall A \subseteq [n] \qquad \Longleftrightarrow \qquad f(x) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} g(B) \qquad \forall A \subseteq [n]$$

which is nothing more or less than the inclusion-exclusion formula. So Möbius inversion can be thought of as a generalized form of inclusion-exclusion in which the Boolean lattice is replaced by an arbitrary locally finite poset P. If we know the Möbius function of P, then knowing a combinatorial formula for either f or g allows us to write down a formula for the other one. This is frequently useful when we can express an enumerative problem in terms of a function on a poset.

[4] In fact (2.3a) requires only that every *principal* order ideal is finite (for (2.3b), every principal order filter).

Remark 2.3.2. The proof of Möbius inversion goes through more generally for functions $f, g : P \to X$, where X is any $\mathbb{C}$-vector space (for example, polynomials over $\mathbb{C}$).

Example 2.3.3. Here's an oldie-but-goodie. A **derangement** is a permutation $\sigma \in \mathfrak{S}_n$ with no fixed points. If $\mathcal{D}_n$ is the set of derangements in $\mathfrak{S}_n$, then

$$\begin{aligned}
&|\mathcal{D}_1| = 0,\\
&|\mathcal{D}_2| = 1 = |\{21\}|,\\
&|\mathcal{D}_3| = 2 = |\{231, 312\}|,\\
&|\mathcal{D}_4| = 9 = |\{2341, 2314, 2413, 3142, 3412, 3421, 4123, 4312, 4321\}|,\\
&\quad \dots
\end{aligned}$$

The problem is to determine $|\mathcal{D}_n|$ in general. For $S \subseteq [n]$, let

$$\begin{aligned}
f(S) &= |\{\sigma \in \mathfrak{S}_n \colon \sigma(i) = i \text{ iff } i \in S\}|,\\
g(S) &= |\{\sigma \in \mathfrak{S}_n \colon \sigma(i) = i \text{ if } i \in S\}|.
\end{aligned}$$

Thus $\mathcal{D}_n = f(\emptyset)$.

It is easy to calculate $g(S)$ directly. If $s = |S|$, then a permutation fixing the elements of S is equivalent to a permutation on $[n] \setminus S$, so $g(S) = (n-s)!$.

It is hard to calculate $f(S)$ directly. However,

$$g(S) = \sum_{R \supseteq S} f(R).$$

Rewritten in the incidence algebra $I(2^{[n]})$, this is just $g = \zeta \bullet f$. Thus $f = \mu \bullet g$, or in terms of the Möbius inversion formula (2.3b),

$$f(S) \;=\; \sum_{R \supseteq S} \mu(S,R) g(R) \;=\; \sum_{R \supseteq S} (-1)^{|R|-|S|}(n-|R|)! \;=\; \sum_{r=s}^{n} \binom{n-s}{r-s} (-1)^{r-s}(n-r)!\,.$$

The number of derangements is then $f(\emptyset)$, which is given by the well-known formula

$$\sum_{r=0}^{n} \binom{n}{r} (-1)^r (n-r)!$$

◀

Example 2.3.4. As a number-theoretic application, we will use Möbius inversion to compute the closed formula for Euler's totient function

$$\phi(n) = \#\{a \in [n] :\ \gcd(a,n) = 1\}.$$

Let $n = p_1^{a_1} \cdots p_s^{a_s}$ be the prime factorization of n, and let $P = \{p_1, \dots, p_s\}$. We work in the lattice $D_n \cong C_{a_1} \times \cdots \times C_{a_s}$. **Warning:** To avoid confusion with the cardinality symbol, we will use the symbol $\leq$ to mean the order relation in D_n: i.e., $x \leq y$ means that x divides y. For $x \in D_n$, define

$$\begin{aligned}
f(x) &= \#\{a \in [n] :\ x = \gcd(a,n)\},\\
g(x) &= \#\{a \in [n] :\ x \leq \gcd(a,n)\} \;=\; \sum_{y \geq x} f(y).
\end{aligned}$$

Applying formulation (2.3b) of Möbius inversion gives

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

On the other hand $g(x) = n/x$, since $x \leq \gcd(a,n)$ iff a is a multiple of x. Moreover, $\phi(n) = f(1)$, and

$$\mu(1,y) = \begin{cases} (-1)^q & \text{if } y \text{ is a product of distinct elements of } P, \\ 0 & \text{otherwise (i.e., if } p_i^2 \leq y \text{ for some } i). \end{cases}$$

Therefore,

$$\begin{aligned} \phi(n) = f(1) &= \sum_{y \in D_n} \mu(1,y)(n/y) \\ &= n \sum_{Q \subseteq P} \frac{(-1)^{|Q|}}{\prod_{p_i \in Q} p_i} \\ &= \frac{n}{p_1 \cdots p_r} \sum_{Q \subseteq P} (-1)^{|Q|} \frac{p_1 \cdots p_r}{\prod_{p_i \in Q} p_i} \\ &= \frac{n}{p_1 \cdots p_r} \sum_{S = P \setminus Q \subseteq P} (-1)^{r-|S|} \prod_{p_i \in S} p_i \\ &= \frac{n}{p_1 \cdots p_r} (-1)^r \prod_{i=1}^{r} (1 - p_i) \\ &= p_1^{a_1 - 1} \cdots p_r^{a_r - 1} (p_1 - 1) \cdots (p_r - 1) \end{aligned}$$

as is well known. ◀

Example 2.3.5. Let $G = (V, E)$ be a finite graph with $V = [n]$. We may as well assume that G is simple (no loops or parallel edges) and connected. A **coloring of G with t colors**, or for short a **t-coloring**, is just a function $\kappa : V(G) \to [t]$. An edge xy is **monochromatic** with respect to κ if $\kappa(x) = \kappa(y)$, and a coloring is **proper** if it has no monochromatic edges. What can we say about the number $p_G(t)$ of proper t-colorings?

This question can be expressed in terms of the connectivity lattice $K(G)$ (see Example 1.2.3 and Problem 1.4). For each t-coloring κ, let G_κ be the subgraph of G induced by the monochromatic edges, and let $P(\kappa)$ be the set partition of $V(G)$ whose blocks are the components of G_κ; then $P(\kappa)$ is an element of $K(G)$. The coloring κ is proper if and only if $P(\kappa) = \hat{\mathbf{0}}_{K(G)}$, the partition of $V(G)$ into singleton blocks. Accordingly, if we define $f : K(G) \to \mathbb{N}_{\geq 0}$ by

$$f(\pi) = |P^{-1}(\pi)| = \#\{t\text{-colorings } \kappa : \ P(\kappa) = \pi\},$$

then the number of proper t-colorings is $f(\hat{\mathbf{0}})$. We can find another expression for this number by Möbius inversion. Let

$$g(\pi) = \#\{\kappa : \ P(\kappa) \geq \pi\} = \sum_{\sigma \geq \pi} f(\sigma).$$

The condition $P(\kappa) \geq \pi$ is equivalent to saying that the vertices in each block of π are colored the same. The number of such colorings is just $t^{|\pi|}$ (choosing a color for each block, not necessarily different). Therefore, Möbius inversion (version (2.3b)) says that

$$p_G(t) = f(\hat{\mathbf{0}}) = \sum_{\pi \in K(G)} \mu(\hat{\mathbf{0}}, \pi) g(\pi) = \sum_{\pi \in K(G)} \mu(\hat{\mathbf{0}}, \pi) t^{|\pi|}. \tag{2.5}$$

While this formula is not necessarily easy to calculate, it does show that $p_G(t)$ is a polynomial in t; it is called the **chromatic polynomial**. (There are other ways to show this fact.)

If $G = K_n$ is the complete graph, then the connectivity lattice $K(K_n)$ is just the full partition lattice Π_n. On the other hand, we can calculate the chromatic polynomial of K_n directly: it is $p_{K_n}(t) = t(t-1)(t-2)\cdots(t-n+1)$ (since a proper coloring must assign different colors to all vertices). Combining this observation with (2.5) gives

$$\sum_{\pi \in K(K_n)} \mu(\hat{\mathbf{0}}, \pi) t^{|\pi|} = t(t-1)(t-2)\cdots(t-n+1).$$

This is an identity of polynomials in t. Extracting the coefficients of the lowest degree (t^1) terms on each side gives

$$\mu(\hat{\mathbf{0}}, \hat{\mathbf{1}}) = (-1)^{n-1}(n-1)!$$

so we have calculated the Möbius number of the partition lattice! There are many other ways to obtain this result. ◀

Example 2.3.6. Here is another way to use Möbius inversion to compute the Möbius function itself. In this example, we will do this for the lattice $L_n(q)$.

For small n, it is possible to work out the Möbius function of $L_n(q)$ by hand. For instance, $\mu(L_1(q)) = \mu(\mathsf{Bool}_1) = -1$, and $L_2(q)$ is a poset of rank 2 with $q+1$ elements in the middle (since each line in $\mathbb{F}_q^2$ is defined by a nonzero vector up to scalar multiples, so there are $(q^2-1)/(q-1)$ lines), so $\mu(L_2(q)) = -(-(q+1)+1) = q$. With a moderate amount of effort, one can check that $\mu(L_3(q)) = -q^3$ and $\mu(L_4(q)) = q^6$. Here is a way to calculate $\mu(L_n(q))$ for general n, which will lead into the discussion of the characteristic polynomial of a ranked poset.

Let $V = \mathbb{F}_q^n$, let $L = L_n(q)$ (ranked by dimension) and let X be a $\mathbb{F}_q$-vector space of *cardinality* t (yes, cardinality, not dimension!) Let

$$\begin{aligned} g(W) &= \#\{\mathbb{F}_q\text{-linear maps } \phi : V \to X \mid \ker\phi \supseteq W\} \\ &= \#\{\mathbb{F}_q\text{-linear maps } \bar{\phi} : V/W \to X\} \\ &= t^{n-\dim W} \end{aligned}$$

since $\bar{\phi}$ is determined by its values on a basis of V/W. Now let

$$f(W) = \#\{\mathbb{F}_q\text{-linear maps } \phi : V \to X \mid \ker\phi = W\}$$

so that

$$g(W) = \sum_{U \supseteq W} f(U)$$

and by Möbius inversion

$$f(W) = \sum_{U:\, V \supseteq U \supseteq W} \mu_L(W, U) t^{n-\dim U}.$$

In particular, if we take W to be the zero subspace $0 = \hat{\mathbf{0}}_L$, we obtain

$$\begin{aligned} f(\hat{\mathbf{0}}) &= \sum_{U \in L} \mu_L(\hat{\mathbf{0}}, U) t^{n-\dim U} \\ &= \#\{\text{1-1 linear maps } V \to X\} \\ &= (t-1)(t-q)(t-q^2)\cdots(t-q^{n-1}). \end{aligned} \tag{2.6}$$

For this last count, choose an ordered basis $\{v_1, \ldots, v_n\}$ for V, and send each v_i to a vector in X not in the linear span of $\{\phi(v_1), \ldots, \phi(v_{i-1})\}$; there are $t - q^{i-1}$ such vectors. The identity (2.6) holds for infinitely

many integer values of t and is thus an identity of polynomials in the ring $\mathbb{Q}[t]$. Therefore, it remains true upon setting t to 0 (even though no vector space can have cardinality zero!), whereupon the second and fourth terms in the equality (2.6) become

$$\mu_{L_n(q)}(\hat{\mathbf{0}}, \hat{\mathbf{1}}) = (-1)^n q^{\binom{n}{2}}$$

which is consistent with the $n \leq 4$ cases given at the start of the example. ◀

The two previous examples suggest that in order to understand a finite graded poset P, one should study the following polynomial.

Definition 2.3.7. Let P be a graded poset with rank function r. Its **characteristic polynomial** is

$$\chi(P; t) = \sum_{x \in P} \mu(\hat{\mathbf{0}}, x) t^{r(\hat{\mathbf{1}}) - r(x)}.$$

In particular,

$$\chi(P, 0) = \mu(P). \tag{2.7}$$

So far, we have shown that

$$\begin{aligned} t \cdot \chi(\Pi_n; t) &= t(t-1) \cdot (t-n+1), \\ t \cdot \chi(K(G); t) &= p_G(t), \\ \chi(L_n(q); t) &= (t-1)(t-q)(t-q^2) \cdots (t-q^{n-1}). \end{aligned}$$

Moreover, the characteristic polynomial of the Boolean lattice Bool_n is

$$\chi(\mathsf{Bool}_n; t) = \sum_{j=0}^{n} (-1)^j \binom{n}{j} t^{n-j} = (t-1)^n.$$

In fact, since the Möbius function is multiplicative on direct products of posets (Proposition 2.2.5), so is the characteristic polynomial.

The characteristic polynomial generalizes the Möbius number of a poset and contains additional information as well. For example, let $\mathcal{A}$ be a *hyperplane arrangement* in $\mathbb{R}^n$: a finite collection of affine linear spaces of dimension $n-1$. The arrangement separates $\mathbb{R}^n$ into *regions*, the connected components of $X = \mathbb{R}^n \setminus \bigcup_{H \in \mathcal{A}} H$. Let P be the poset of intersections of hyperplanes in H, ordered by reverse refinement. A famous result of Zaslavsky, which we will prove in Chapter 5, is that $|\chi_P(-1)|$ and $|\chi_P(1)|$ count the number of regions and bounded regions of X, respectively.

2.4 Möbius functions of lattices

There are additional techniques we can use for computing Möbius functions and characteristic polynomials of lattices, particularly lattices with good structural properties (e.g., semimodular).

Definition 2.4.1. Let L be a lattice. The **Möbius algebra** $\mathrm{Möb}(L)$ is the vector space of formal $\mathbb{C}$-linear combinations of elements of L, with multiplication given by the meet operation and extended linearly. (In particular, $\hat{\mathbf{1}}$ is the multiplicative unit of $\mathrm{Möb}(L)$.)

The elements of L form a vector space basis of $\mathrm{Möb}(L)$ consisting of idempotents (elements that are their own squares), since $x \wedge x = x$ for all $x \in L$. For example, if $L = 2^{[n]}$ then $\mathrm{Möb}(L) \cong \mathbb{C}[x_1, \ldots, x_n]/(x_1^2 - x_1, \ldots, x_n^2 - x_n)$, with a natural vector space basis given by squarefree monomials.

It seems as though $\mathrm{Möb}(L)$ could have a complicated ring structure, but actually it is quite simple.

Proposition 2.4.2. *Let L be a finite lattice with n elements. For $x \in L$, define*

$$\varepsilon_x = \sum_{y \leq x} \mu(y, x) y \in \mathrm{Möb}(L).$$

Then the set $B = \{\varepsilon_x : x \in L\}$ is a $\mathbb{C}$-vector space basis for $\mathrm{Möb}(L)$, with $\varepsilon_x \varepsilon_y = \delta_{xy} \varepsilon_x$. In particular, $\mathrm{Möb}(L) \cong \mathbb{C}^n$ as rings.

Proof. Applying Möbius inversion with $f(x) = \varepsilon_x$ and $g(x) = x$ yields[5]

$$x = \sum_{y \leq x} \varepsilon_y. \tag{2.8}$$

In particular, B is a vector space basis for $\mathrm{Möb}(L)$ as claimed. Let $\mathbb{C}_x$ be a copy of $\mathbb{C}$ with unit 1_x, so that we can identify $\mathbb{C}^{|L|}$ with $\prod_{x \in L} \mathbb{C}_x$. This is the direct product of rings, with multiplication $1_x 1_y = \delta_{xy} 1_x$. We claim that the $\mathbb{C}$-linear map $\phi : \mathrm{Möb}(L) \to \mathbb{C}^n$ given by $\phi(\varepsilon_x) = 1_x$ is a ring isomorphism. It is certainly a vector space isomorphism, and (2.8) implies that

$$\phi(x)\phi(y) = \phi\left(\sum_{w \leq x} \varepsilon_w\right) \phi\left(\sum_{z \leq y} \varepsilon_z\right) = \left(\sum_{w \leq x} 1_w\right)\left(\sum_{z \leq y} 1_z\right) = \sum_{v \leq x \wedge y} 1_v = \phi(x \wedge y). \qquad \square$$

This proof looks like sleight-of-hand, since we never calculated $\varepsilon_x \varepsilon_y$ explicitly!

Remark 2.4.3. Darij Grinberg observes that Prop. 2.4.2 goes through if L is assumed merely to be a (finite) meet-semilattice, rather than a lattice. Interestingly, since L need not have a top element, it is not immediate from the definition of $\mathrm{Möb}(L)$ that it must have a unit, but the existence of a unit is implied by the isomorphism $\mathrm{Möb}(L) \cong \mathbb{C}^n$ (in fact the unit is $\sum_{x \in L} \varepsilon_x$).

The Möbius algebra leads to useful identities that rely on translating between the "combinatorial" basis L and the "algebraic" basis B. Some of these identities permit computation of $\mu(x, y)$ by summing over a cleverly chosen *subset* of $[x, y]$, rather than the entire interval. Of course we know that $\mu(P) = -\sum_{x \neq \hat{\mathbf{1}}} \mu(\hat{\mathbf{0}}, x)$ for any poset P, but calculating $\mu(P)$ explicitly using this formula requires a recursive computation that can be quite inefficient. The special structure of a lattice L leads to much more streamlined expressions for $\mu(L)$. The first of these, *Weisner's theorem* (Prop. 2.4.4), reduces the number of summands substantially; it is easy to prove and has useful consequences, but is still recursive. The second, *Rota's crosscut theorem* (Thm. 2.4.9), requires more setup but is non-recursive, which makes it a more versatile tool.

Proposition 2.4.4 (Weisner's theorem). *Let L be a finite lattice with $|L| \geq 2$, and let $a \in L \setminus \{\hat{\mathbf{1}}\}$. Then*

$$\sum_{\substack{x \in L: \\ x \wedge a = \hat{\mathbf{0}}}} \mu(x, \hat{\mathbf{1}}) = 0. \tag{2.9}$$

In particular, pulling off the $x = \hat{\mathbf{0}}$ summand gives

$$\mu(L) \;=\; \mu_L(\hat{\mathbf{0}}, \hat{\mathbf{1}}) \;=\; -\sum_{\substack{x \in L \setminus \{\hat{\mathbf{0}}\}: \\ x \wedge a = \hat{\mathbf{0}}}} \mu(x, \hat{\mathbf{1}}). \tag{2.10}$$

[5]Here the vector space V of Theorem 2.3.1 is not $\mathbb{C}$ itself, but rather another vector space, namely $A(L)$.

Proof. We work in $\mathrm{Möb}(L)$ and calculate $a\varepsilon_{\hat{\mathbf{1}}}$ in two ways. On the one hand

$$a\varepsilon_{\hat{\mathbf{1}}} \;=\; \left(\sum_{b\le a}\varepsilon_b\right)\varepsilon_{\hat{\mathbf{1}}} \;=\; 0.$$

On the other hand

$$a\varepsilon_{\hat{\mathbf{1}}} \;=\; a\sum_{x\in L}\mu(x,\hat{\mathbf{1}})x \;=\; \sum_{x\in L}\mu(x,\hat{\mathbf{1}})x\wedge a.$$

Now taking the coefficient of $\hat{\mathbf{0}}$ on both sides gives (2.9), and (2.10) follows immediately. □

Example 2.4.5 (The Möbius function of the partition lattice Π_n). Let $a = 1|23\cdots n \in \Pi_n$. Then the partitions x that show up in the sum of (2.10) are just the atoms whose non-singleton block is $\{1,i\}$ for some $i>1$. For each such x, the interval $[x,\hat{\mathbf{1}}]\subseteq\Pi_n$ is isomorphic to Π_{n-1}, so (2.10) gives

$$\mu(\Pi_n) \;=\; -(n-1)\mu(\Pi_{n-1})$$

from which it follows by induction that

$$\mu(\Pi_n) \;=\; (-1)^{n-1}(n-1)!.$$

(Wasn't that easy?) ◀

Example 2.4.6 (The Möbius function of the subspace lattice $L_n(q)$). Let $L=L_n(q)$, and let $A=\{(v_1,\ldots,v_n)\in\mathbb{F}_q^n\colon v_n=0\}$. Then $\dim A = n-1$, i.e., A is a coatom in L. If X is a nonzero subspace such that $X\cap A=0$, then X must be a line spanned by some vector $(u_1,\ldots,u_n)$ with $u_n\neq 0$. We may as well assume $u_n=1$ and choose $u_1,\ldots,u_{n-1}$ arbitrarily, so there are q^{n-1} such lines. Moreover, the interval $[X,\hat{\mathbf{1}}]\subseteq L$ is isomorphic to $L_{n-1}(q)$. Therefore (2.10) gives

$$\mu(L_n(q)) \;=\; -q^{n-1}\mu(L_{n-1}(q))$$

and by induction

$$\mu(L_n(q)) \;=\; (-1)^n q^{\binom{n}{2}}.$$

◀

Here is an important consequence of Weisner's theorem.

Theorem 2.4.7. *The Möbius function of any upper semimodular lattice L weakly alternates in sign. That is, $(-1)^{r(x)}\mu(\hat{\mathbf{0}},x)\ge 0$ for all $x\in L$.*

Proof. It is sufficient to prove that $(-1)^{r(L)}\mu(L)\ge 0$, since every interval in a USM lattice is USM.

Let $a\in L\setminus\{\hat{\mathbf{0}}\}$. Applying Weisner's theorem to L^* and using the fact that $\mu(P)=\mu(P^*)$ (Corollary 2.2.10), we see that

$$\sum_{x\in L:\ x\vee a=\hat{\mathbf{1}}}\mu(\hat{\mathbf{0}},x)=0. \tag{2.11}$$

Now, suppose L is USM of rank n. The theorem is certainly true if $n\le 1$, so we proceed by induction. Take a to be an atom. If $x\vee a=\hat{\mathbf{1}}$, then

$$\begin{aligned} r(x) \;&\ge\; r(x\vee a)+r(x\wedge a)-r(a)\\ &=\; n+r(x\wedge a)-1\\ &\ge\; n-1 \end{aligned}$$

so either $x = \hat{\mathbf{1}}$, or else x is a coatom whose meet with a is $\hat{\mathbf{0}}$. Therefore, we can solve for $\mu(\hat{\mathbf{0}}, \hat{\mathbf{1}})$ in (2.11) to get

$$\mu(\hat{\mathbf{0}}, \hat{\mathbf{1}}) \;=\; - \sum_{\text{coatoms } x:\; x \wedge a = \hat{\mathbf{1}}} \mu(\hat{\mathbf{0}}, x).$$

But each interval $[\hat{\mathbf{0}}, x]$ is itself a USM lattice of rank $n-1$, so by induction each summand has sign $(-1)^{n-1}$, which completes the proof. □

A drawback of Weisner's theorem is that it is still recursive; the right-hand side of (2.10) involves other values of the Möbius function. This is not a problem for integer-indexed families of lattices $\{L_n\}$ such that every rank-k element $x \in L_n$ has $[\hat{\mathbf{0}}, x] \cong L_k$ (as we have just seen), but this is too much to hope for in general. The next result, Rota's *crosscut theorem*, gives a non-recursive way of computing the Möbius function.

Definition 2.4.8. Let L be a lattice. An **upper crosscut** of L is a set $X \subseteq L \setminus \{\hat{\mathbf{1}}\}$ such that if $y \in L \setminus X \setminus \{\hat{\mathbf{1}}\}$, then $y < x$ for some $x \in X$. A **lower crosscut** of L is a set $X \subseteq L \setminus \{\hat{\mathbf{0}}\}$ such that if $y \in L \setminus X \setminus \{\hat{\mathbf{0}}\}$, then $y > x$ for some $x \in X$.

It would be simpler to define an upper (resp., lower) crosscut as a set that contains all coatoms (resp., atoms), but in practice the formulation in the previous definition is typically a convenient way to show that a particular set is a crosscut.

Theorem 2.4.9 (Rota's crosscut theorem). *Let L be a finite lattice and $X \subset L$ an upper crosscut. Then*

$$\mu(L) = \sum_{A \subseteq X:\; \bigwedge A = \hat{\mathbf{0}}} (-1)^{|A|}. \tag{2.12a}$$

Dually, if X is a lower crosscut, then

$$\mu(L) = \sum_{A \subseteq X:\; \bigvee A = \hat{\mathbf{1}}} (-1)^{|A|}. \tag{2.12b}$$

Proof. We prove only (2.12a); the proof of (2.12b) is dual. Fix $x \in L$ and start with the following equation in $\mathrm{Möb}(L)$ (recalling (2.8)):

$$\hat{\mathbf{1}} - x \;=\; \sum_{y \in L} \varepsilon_y \;-\; \sum_{y \le x} \varepsilon_y \;=\; \sum_{y \not\le x} \varepsilon_y.$$

Therefore, for any $X \subseteq L$,

$$\prod_{x \in X} (\hat{\mathbf{1}} - x) \;=\; \prod_{x \in X} \sum_{y \not\le x} \varepsilon_y \;=\; \sum_{y \in Y} \varepsilon_y$$

where $Y = \{y \in L \colon y \not\le x \text{ for all } x \in X\}$. (Expand the sum and recall that $\varepsilon_y \varepsilon_{y'} = \delta_{yy'} \varepsilon_y$.) But if X is an upper crosscut, then $Y = \{\hat{\mathbf{1}}\}$, and this last equation becomes

$$\prod_{x \in X} (\hat{\mathbf{1}} - x) \;=\; \varepsilon_{\hat{\mathbf{1}}} \;=\; \sum_{y \in L} \mu(y, \hat{\mathbf{1}}) y. \tag{2.13}$$

On the other hand, a direct binomial expansion gives

$$\prod_{x \in X} (\hat{\mathbf{1}} - x) \;=\; \sum_{A \subseteq X} (-1)^{|A|} \bigwedge A. \tag{2.14}$$

Now equating the coefficients of $\hat{\mathbf{0}}$ on the right-hand sides of (2.13) and (2.14) yields (2.12a). □

Corollary 2.4.10 (Möbius numbers of some lattices are boring). *Let L be a lattice in which $\hat{\mathbf{1}}$ is not a join of atoms (for example, a distributive lattice that is not Boolean, such as almost any principal order ideal in Young's lattice). Then $\mu(L) = 0$.*

The crosscut theorem will be useful in studying hyperplane arrangements. Another topological application is the following result due to J. Folkman (1966), whose proof (omitted) uses the crosscut theorem.

Theorem 2.4.11. *Let L be a geometric lattice of rank r, and let $P = L \setminus \{\hat{\mathbf{0}}, \hat{\mathbf{1}}\}$. Then*

$$\tilde{H}_i(\Delta(P), \mathbb{Z}) \cong \begin{cases} \mathbb{Z}^{|\mu(L)|} & \text{if } i = r-2, \\ 0 & \text{otherwise} \end{cases}$$

where $\tilde{H}_i$ denotes reduced simplicial homology. That is, $\Delta(P)$ has the homology type of the wedge of $\mu(L)$ spheres of dimension $r-2$.

2.5 Exercises

Problem 2.1. Let P be a locally finite poset. Consider the incidence function $\kappa \in I(P)$ defined by

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

(a) Give a combinatorial interpretation of $\kappa^n(x, y)$ for all $x, y \in P$ and $n \in \mathbb{N}$.
(b) How can you tell from κ and its convolution powers whether P is ranked?
(c) Give combinatorial interpretations of $\kappa * \zeta(x, y)$ and $\zeta * \kappa(x, y)$.

Problem 2.2. Prove that the Möbius function is multiplicative on direct products (i.e., $\mu_{P\times Q} = \mu_P \mu_Q$ in the notation of Proposition 2.1.4) directly from the definition of μ.

Problem 2.3. Let P be a finite bounded poset and let P^* be its dual; recall that this means that $x \leq_P y$ if and only if $y \leq_{P^*} x$. Consider the vector space map $F : I(P) \to I(P^*)$ given by $F(\alpha)(y, x) = \alpha(x, y)$.

(a) Show that F is an *anti-isomorphism* of algebras, i.e., it is a vector space isomorphism and $F(\alpha * \beta) = F(\beta) * F(\alpha)$.
(b) Show that $F(\delta_P) = \delta_{P^*}$ and $F(\zeta_P) = \zeta_{P^*}$. Conclude that $F(\mu_P) = \mu_{P^*}$ and therefore that $\mu(P) = \mu(P^*)$.

Problem 2.4. Let L be an atomic lattice with atoms A. Prove that

$$\mu(x) = \sum_{B \subseteq A:\ \bigvee B = x} (-1)^{|B|}.$$

for every $x \in L$. (In a sense, this says that all Möbius functions of atomic lattices come from the M"obius function of a Boolean lattice.)

Problem 2.5. (Based on an observation by Mark Denker) Prove that

$$p_G(t) = \sum_k c_k(G) t(t-1)(t-2)\cdots(t-k+1)$$

where $c_k(G)$ is the number of ways of properly coloring G using exactly k colors (i.e., proper colorings that are *surjective* functions $\kappa : V(G) \to [k]$).

Problem 2.6. A set partition in Π_n is a *noncrossing partition* (NCP) if its associated equivalence relation $\sim$ satisfies the following condition: for all $i < j < k < \ell$, if $i \sim k$ and $j \sim \ell$ then $i \sim j \sim k \sim \ell$. The set of all NCPs of order n is denoted NC_n. Ordering by reverse refinement makes NC_n into a subposet of the partition lattice Π_n. Note that $\mathsf{NC}_n = \Pi_n$ for $n \leq 3$ (the smallest partition that is *not* noncrossing is $13|24 \in \Pi_4$). NCPs can be represented pictorially by *chord diagrams*. The chord diagram of $\xi = 1|2\ 5|3|4|6\ 8\ 12|7|9|10\ 11 \in \mathsf{NC}_{12}$ is shown in Figure 2.1(a).

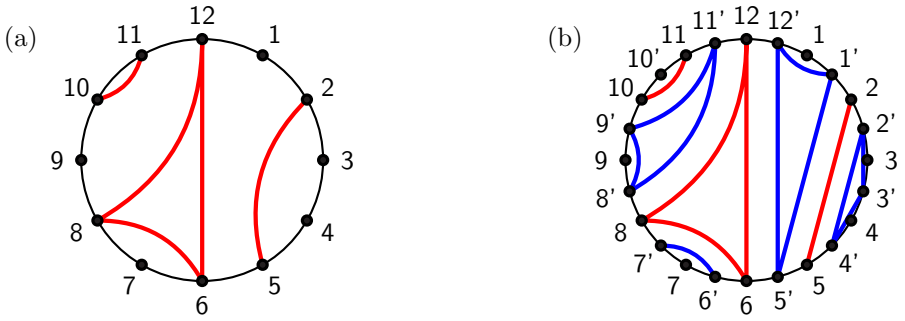


Figure 2.1: (a) A noncrossing partition $\xi \in \mathsf{NC}_{12}$. (b) Kreweras complementation.

(a) Prove that NC_n is a ranked lattice. Is it a sublattice of Π_n?

(b) Prove that the numbers $\mathsf{nc}_n = |\mathsf{NC}_n|$ satisfy the *Catalan recurrence*

$$\mathsf{nc}_n = \mathsf{nc}_{n-1} + \sum_{k=1}^{n-1} \mathsf{nc}_{k-1}\mathsf{nc}_{n-k};$$

therefore, nc_n is the nth Catalan number $C_n = \frac{1}{n+1}\binom{2n}{n}$. (You can formally define $\mathsf{nc}_0 = 1$, and establish the recurrence for $n \geq 1$.)

(c) Prove that the operation of *Kreweras complementation* is an anti-automorphism of NC_n. To define the Kreweras complement $K(\pi)$ of $\pi \in \mathsf{NC}_n$, start with the chord diagram of π and insert a point labeled i' between the points i and $i+1$ (mod n) for $i = 1, 2, \ldots, n$. Then a, b lie in the same block of $K(\pi)$ if it is possible to walk from a' to b' without crossing an arc of π. For instance, the Kreweras complement of the noncrossing partition $\xi \in \mathsf{NC}_{12}$ shown above is $K(\xi) = 1\ 5\ 12|2\ 3\ 4|6\ 7|8\ 9\ 11|10$ (see Figure 2.1(b)).

(d) Use Weisner's theorem to prove that $\mu(\mathsf{NC}_n) = (-1)^{n-1}C_{n-1}$ for all $n \geq 1$.

The characteristic polynomial of NC_n satisfies a version of the Catalan recurrence. For details see [LS00] (this might make a good end-of-semester project).

Problem 2.7. This problem is about how far Proposition 2.4.2 can be extended. Suppose that R is a commutative $\mathbb{C}$-algebra of finite dimension n as a $\mathbb{C}$-vector space, and that $x_1, \ldots, x_n \in R$ are linearly independent idempotents (i.e., $x_i^2 = x_i$ for all i). Prove that $R \cong \mathbb{C}^n$ as rings.

Problem 2.8. The ***q*-binomial coefficient** is the rational function

$$\begin{bmatrix} n \\ k \end{bmatrix}_q = \frac{(q^n-1)(q^n-q)\cdots(q^n-q^{k-1})}{(q^k-1)(q^k-q)\cdots(q^k-q^{k-1})}.$$

Here q is an indeterminate, and k, n are nonnegative integers with $k \leq n$. (For any other integers n, k, we set $\begin{bmatrix} n \\ k \end{bmatrix}_q = 0$.) We have already seen that $\begin{bmatrix} n \\ k \end{bmatrix}_q$ is the number of k-dimensional subspaces of $\mathbb{F}_q^n$ (Example 1.2.5, Problem 1.14).

(a) Check that setting $q = 1$ (after canceling out common terms), or equivalently applying $\lim_{q\to 1}$, recovers the ordinary binomial coefficient $\binom{n}{k}$.

(b) Prove the *q-Pascal identities*: for $n \geq 1$ and all k, we have

$$\begin{bmatrix} n \\ k \end{bmatrix}_q = q^k \begin{bmatrix} n-1 \\ k \end{bmatrix}_q + \begin{bmatrix} n-1 \\ k-1 \end{bmatrix}_q \qquad \text{and} \qquad \begin{bmatrix} n \\ k \end{bmatrix}_q = \begin{bmatrix} n-1 \\ k \end{bmatrix}_q + q^{n-k} \begin{bmatrix} n-1 \\ k-1 \end{bmatrix}_q.$$

Deduce that $\begin{bmatrix} n \\ k \end{bmatrix}_q$ is actually a polynomial in q (not merely a rational function).

(c) Prove that $\begin{bmatrix} n \\ k \end{bmatrix}_q$ is the generating function for partitions that fit into the rectangle $R_{k,n-k}$ with $n-k$ rows and k columns. That is,

$$\begin{bmatrix} n \\ k \end{bmatrix}_q = \sum_{\lambda \subseteq R(k,n-k)} q^{|\lambda|}.$$

(This result has a surprising application to the topology of algebraic varieties; see §11.4.)

(d) (Stanley, EC1, 2nd ed., 3.119) Prove the *q-binomial theorem*:

$$\prod_{k=0}^{n-1} (x - q^k) = \sum_{k=0}^{n} \begin{bmatrix} n \\ k \end{bmatrix}_q (-1)^k q^{\binom{k}{2}} x^{n-k}.$$

(Hint: Let $V = \mathbb{F}_q^n$ and let X be a vector space over $\mathbb{F}_q$ with x elements. Count the number of one-to-one linear transformations $V \to X$ in two ways.)

Problem 2.9. (Stanley, EC1, 3.129) Here is a cute application of combinatorics to elementary number theory. Let P be a finite poset, and let $\hat{P} = P \cup \{\hat{\mathbf{0}}, \hat{\mathbf{1}}\}$. Suppose that P has a fixed-point-free automorphism $\sigma : P \to P$ of prime order p; that is, $\sigma(x) \neq x$ and $\sigma^p(x) = x$ for all $x \in P$. Prove that $\mu_{\hat{P}}(\hat{\mathbf{0}}, \hat{\mathbf{1}}) \equiv -1 \pmod{p}$. What does this say in the case that $\hat{P} = \Pi_p$?

Chapter 3

Matroids

The motivating example of a geometric lattice is the lattice of flats of a finite set E of vectors. The underlying combinatorial data of this lattice can be expressed in terms of the rank function, which says the dimension of the space spanned by every subset of E. However, there are many other equivalent ways to describe the "combinatorial linear algebra" of a set of vectors: the family of linearly independent sets; the family of sets that form bases; which vectors lie in the span of which sets; etc. Each of these data sets defines the structure of a **matroid** on E. Matroids can also be regarded as generalizations of graphs, and are important in combinatorial optimization as well. A standard reference on matroid theory is [Oxl92], although I first learned the basics of the subject from an unusual (but very good) source, namely chapter 3 of [GSS93].

Conventions: Unless otherwise specified, E always denotes a finite set. We will be doing a lot of adding elements to and removing elements e from sets A, so for convenience we define $A + e = A \cup \{e\}$ and $A - e = A \setminus \{e\}$.

3.1 Closure operators

Definition 3.1.1. Let E be a finite set. A **closure operator on E** is a map $2^E \to 2^E$, written $A \mapsto \bar{A}$, such that

(i) $A \subseteq \bar{A}$;
(ii) $\bar{A} = \bar{\bar{A}}$; and
(iii) if $A \subseteq B$, then $\bar{A} \subseteq \bar{B}$

for all $A, B \subseteq E$. A set A is called **closed** or a **flat** if $\bar{A} = A$.

Proposition 3.1.2. *Every closure operator on E gives rise to a lattice with bottom element $\bar{\emptyset}$, top element E, and meet and join given by $F \wedge G = F \cap G$, $F \vee G = \overline{F \cup G}$.*

Proof. Any two subsets $A, B \subseteq E$ satisfy $\overline{A \cap B} \subseteq \overline{A}$ and $\overline{A \cap B} \subseteq \overline{B}$ (by (iii)), hence $\overline{A \cap B} \subseteq \overline{A} \cap \overline{B}$. In particular, if F and G are flats, then

$$\overline{F \cap G} \subseteq \overline{F} \cap \overline{G} = F \cap G \tag{3.1}$$

so equality holds. That is, the intersection of flats is a flat, so the flats form a meet-semilattice under intersection, hence a lattice. Meanwhile, if H is a flat containing both F and G, then $H = \bar{H} = \overline{F \cup G}$, so $\overline{F \cup G}$ must be the join of F and G. □

Definition 3.1.3. A **matroid closure operator** is a closure operator that satisfies the **exchange property:**

$$\text{if } e \notin \bar{A} \text{ but } e \in \overline{A + e'}, \text{ then } e' \in \overline{A + e} \qquad \forall A \subseteq E. \tag{3.2}$$

A **matroid** M is a set E (the "ground set") together with a matroid closure operator on E. A matroid is **simple** if the empty set and all singleton sets are closed.

Example 3.1.4. *Vector matroids.* Let V be a vector space over a field $\Bbbk$, and let $E \subseteq V$ be a finite set. Then

$$A \mapsto \bar{A} := \Bbbk A \cap E$$

is a matroid closure operator on E. It is easy to check the conditions for a closure operator. To check condition (3.2), if $e \in \overline{A + e'}$, then there is a linear equation

$$e = c_{e'} e' + \sum_{a \in A} c_a a$$

where $c_{e'}$ and all the c_a are scalars in $\Bbbk$. The condition $e \notin \bar{A}$ implies that $c_{e'} \neq 0$ in any equation of this form. Therefore, the equation can be rewritten to express e' as a linear combination of the vectors in $A + e$, obtaining (3.2). A matroid arising in this way (or, more generally, isomorphic to such a matroid) is called a **vector matroid**, **linear matroid** or **representable matroid**.[1] ◀

A vector matroid records information about linear dependence (i.e., which vectors belong to the linear spans of other sets of vectors) without having to worry about the actual coordinates of the vectors. More generally, a matroid can be thought of as a combinatorial, coordinate-free abstraction of linear dependence and independence. Note that a vector matroid is simple if none of the vectors is zero (so that $\bar{\emptyset} = \emptyset$) and if no vector is a scalar multiple of another (so that all singleton sets are closed).

3.2 Matroids and geometric lattices

The following theorem says that simple matroids and geometric lattices are essentially the same things. In Rota's language, they are "cryptomorphic": their definitions look very different, but they carry the same information. We will see many more ways to axiomatize the same information: rank functions, independence systems, basis systems, etc. Working with matroids requires a solid level of comfort with the cryptomorphisms between the various definitions of a matroid.

Theorem 3.2.1.
1. *Let M be a simple matroid with finite ground set E, and $L(M)$ its lattice of flats. Then $L(M)$ is a geometric lattice, under the operations $F \wedge G = F \cap G$, $F \vee G = \overline{F \cup G}$.*
2. *Let L be a geometric lattice and let E be its set of atoms. Then the function $A \mapsto \overline{A} = \{e \in E : e \leq \bigvee A\}$ is a matroid closure operator for a simple matroid on E.*
3. *These constructions are mutual inverses.*

Proof. (1) Recall that $L(M)$ is a lattice by Proposition 3.1.2. By definition of a simple matroid, the bottom element is $\emptyset$ and the atoms are the singleton subsets of E. Every flat is the join of the atoms corresponding to its elements, so the lattice $L(M)$ is atomic.

We now characterize the covering relations in $L(M)$.

Lemma 3.2.2. *If $F \in L(M)$ and $e \in E \setminus F$ (so that $F < F \vee e$), then in fact $F \lessdot F \vee e$.*

[1] Usually one of the first two terms is used for a matroid defined by a given set of vectors; "representable" just suggests that the matroid *could* be represented in that way.

Proof. Suppose that there is a flat G such that

$$F \subsetneq G \subseteq F \vee e = \overline{F + e}. \tag{3.3}$$

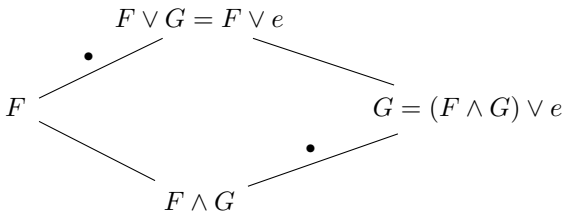
Let $e' \in G \setminus F$. Then $e' \in \overline{F + e}$, so the exchange axiom (3.2) implies $e \in \overline{F + e'}$, which in turn implies that $F \vee e \subseteq F \vee e' \subseteq G$. Hence the $\subseteq$ in (3.3) is actually an equality. We have shown that there are no flats strictly between F and $F \vee e$, proving the claim. □

Of course, if $F \lessdot G$ then $G = F \vee e$ for any $e \in G \setminus F$. So the covering relations in $L(M)$ are precisely of this form.

Suppose now that F and G are incomparable and that $G \gtrdot F \wedge G$. Then G is of the form $(F \wedge G) \vee e$, and we can take e to be any element of $G \setminus F$. In particular $F < F \vee e$, so by Lemma 3.2.2, $F \lessdot F \vee e$. Moreover,

$$F \vee G = F \vee ((F \wedge G) \vee e) = (F \vee e) \vee (F \wedge G) = F \vee e.$$

We have just proved that $L(M)$ is semimodular. Here is the diamond picture (cf. (1.7)):



In particular, $L(M)$ is ranked, with rank function

$$r(F) = \min\left\{|B| : \ B \subseteq E, \ F = \bigvee B\right\}.$$

Such a set B is called a **basis** of F.

(2) Let L be a geometric lattice with atoms E, and define $\overline{A} = \{e \in E \colon e \leq \bigvee A\}$ for $A \subseteq E$. It is easy to check that $A \mapsto \bar{A}$ is a closure operator, and that $\bar{A} = A$ whenever $|A| \leq 1$. So the only nontrivial part is to establish the exchange axiom (3.2).

Recall that if L is semimodular and $x, e \in L$ with e an atom and $x \not\geq e$ (so that $x < x \vee e$), then in fact $x \lessdot x \vee e$, because

$$r(x \vee e) - r(x) \leq r(e) - r(x \wedge e) = 1 - 0 = 1.$$

Accordingly, let $A \subseteq E$ and let $e, f \in E \setminus A$. Suppose that $e \in \overline{A + f}$; we must show that $f \in \overline{A + e}$. Let $x = \bigvee A \in L$. Then

$$x \lessdot x \vee f \quad \text{(by the previous remark)} \qquad \text{and} \qquad x < x \vee e \leq x \vee f,$$

which together imply that $x \vee f = x \vee e$. In particular $f \leq x \vee e$, i.e., $f \in \overline{A + e}$, proving that we have a matroid closure operator.

Part (3) is left as an exercise. □

Example 3.2.3. Let $E = \{a, b, c, d, e\}$. The subfamily of 2^E given by

$$L = \{\emptyset, \quad a,\ b,\ c,\ d,\ e, \quad ad,\ ae,\ bd,\ be,\ abc,\ cde, \quad abcde\},$$

regarded as a poset under inclusion, is a geometric lattice of rank 3 (feel free to check this yourself). In the displayed equation, the elements of L are grouped by rank. The associated matroid closure operator has, for example, $\overline{ad} = ad$, $\overline{ac} = \overline{abc} = abc$, $\overline{ace} = abcde$. ◀

In view of Theorem 3.2.1, we can describe a matroid on ground set E by the function $A \mapsto r(\bar{A})$, where r is the rank function of the associated geometric lattice. It is standard to abuse notation by calling this function r as well. Formally:

Definition 3.2.4. A *matroid rank function* on E is a function $r : 2^E \to \mathbb{N}$ satisfying the following conditions for all $A, B \subseteq E$:

(R1) $r(A) \leq |A|$.
(R2) If $A \subseteq B$ then $r(A) \leq r(B)$.
(R3) $r(A) + r(B) \geq r(A \cap B) + r(A \cup B)$ (the **submodular inequality**).

If r is a matroid rank function on E, then the corresponding matroid closure operator is given by

$$\overline{A} = \{e \in E :\ r(A + e) = r(A)\}.$$

Moreover, this matroid is simple if and only if $r(A) = |A|$ whenever $|A| \leq 2$.

Conversely, if $A \mapsto \bar{A}$ is a matroid closure operator on E, then the corresponding matroid rank function r is

$$r(A) = \min\{|B| :\ \bar{B} = \bar{A}\}.$$

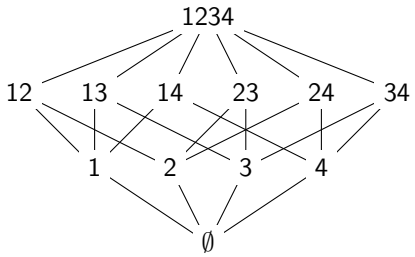
Example 3.2.5. Let $n = |E|$ and $0 \leq k \leq n$, and define

$$r(A) = \min(k, |A|).$$

It is easy to check that this satisfies the conditions of Definition 3.2.4. The corresponding matroid is called the **uniform matroid** $U_k(n)$. Its closure operator is

$$\overline{A} = \begin{cases} A & \text{if } |A| < k, \\ E & \text{if } |A| \geq k. \end{cases}$$

So the flats of M are the sets of cardinality $< k$, as well as E itself. Therefore, the lattice of flats looks like a Boolean lattice $2^{[n]}$ that has been truncated at the kth rank: that is, all elements of rank $\geq k$ have been deleted and replaced with a single $\hat{\mathbf{1}}$. For $n = 3$ and $k = 2$, this lattice is M_5. For $n = 4$ and $k = 3$, the Hasse diagram is as shown below.



If E is a set of n vectors in general position in $\Bbbk^k$, then the corresponding linear matroid is isomorphic to $U_k(n)$. This sentence is tautological, in the sense that it can be taken as a definition of "general position". If $\Bbbk$ is infinite and the points are chosen randomly (in some reasonable measure-theoretic sense), then $L(E)$ will be isomorphic to $U_k(n)$ with probability 1. On the other hand, $\Bbbk$ must be sufficiently large (in terms of n) in order for $\Bbbk^k$ to have n points in general position: for instance, $U_2(4)$ cannot be represented as a matroid over $\mathbb{F}_2$ simply because $\mathbb{F}_2^2$ contains only three nonzero vectors. ◀

At this point, let us formalize what isomorphism of matroids means.

Definition 3.2.6. Let M, M' be matroids on ground sets E, E' respectively. We say that M and M' are **isomorphic**, written $M \cong M'$, if there is a bijection $f : E \to E'$ satisfying any (hence all) of the following:

1. f induces a lattice isomorphism $\hat{f} : L(M) \cong L(M')$, i.e., $\hat{f}(x_1 \vee \cdots \vee x_k) = f(x_1) \vee \cdots \vee f(x_k)$.
2. $r(A) = r(f(A))$ for all $A \subseteq E$. (Here $f(A) = \{f(a) \colon a \in A\}$.)
3. $f(\bar{A}) = \overline{f(A)}$ for all $A \subseteq E$.

In general, every definition of "matroid" (and there are several more coming) will induce a corresponding equivalent for isomorphic.

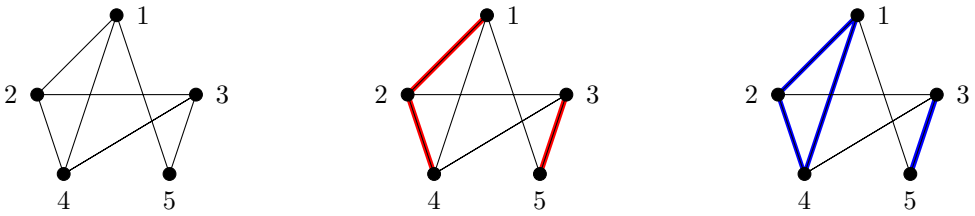
3.3 Graphic matroids

Let G be a finite graph with vertices V and edges E. For convenience, we will write $e = xy$ to mean "e is an edge with endpoints x, y". This notation does not exclude the possibility that e is a loop (i.e., $x = y$) or that some other edge might have the same pair of endpoints.

Definition 3.3.1. For each subset $A \subseteq E$, the corresponding **induced subgraph** of G is the graph $G|_A$ with vertices V and edges A. The **graphic matroid** or **complete connectivity matroid** $M(G)$ on E is defined by the closure operator

$$\bar{A} = \{e = xy \in E \colon x, y \text{ belong to the same component of } G|_A\}. \tag{3.4}$$

Equivalently, an edge $e = xy$ belongs to $\bar{A}$ if there is a path between x and y consisting of edges in A (for short, an *A-path*). For example, in the graph, $14 \in \bar{A}$ because $\{12, 24\} \subseteq A$.



$G = (V, E)$ $\quad A$ $\quad \bar{A}$

Proposition 3.3.2. *The operator $A \mapsto \bar{A}$ defined by (3.4) is a matroid closure operator.*

Proof. It is easy to check that $A \subseteq \bar{A}$ for all A, and that $A \subseteq B \implies \bar{A} \subseteq \bar{B}$. If $e = xy \in \bar{\bar{A}}$, then x, y can be joined by an $\bar{A}$-path P, and each edge in P can be replaced with an A-path, giving an A-path between x and y.

Finally, suppose $e = xy \notin \bar{A}$ but $e \in \overline{A+f}$. Let P be an $(A+f)$-path from x to y. Then $f \in P$ (because there is no A-path from x to y) and $P + e$ is a cycle. Deleting f produces an $(A+e)$-path between the endpoints of f. (See Figure 3.1.) □

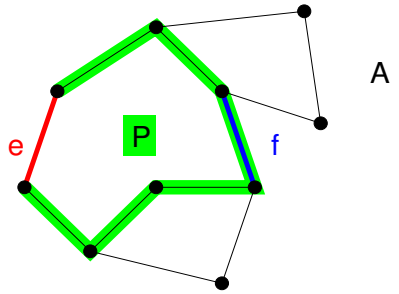


Figure 3.1: The closure axiom for a graphic matroid. Here A consists of all edges shown except e and f; neither e nor f belongs to $\bar{A}$, but $e \in \overline{A+f}$ and $f \in \overline{A+e}$.

The rank function of the graphic matroid is given by

$$r(A) = \min\{|B| : B \subseteq A, \overline{B} = \overline{A}\}.$$

Such a subset B is called a *spanning forest*[2] of A (or of $G|_A$). They are the bases of the graphic matroid $M(G)$. (I haven't yet said what a basis is — see the next section.)

Theorem 3.3.3. *Let $B \subseteq A$. Then any two of the following conditions imply the third (and characterize spanning forests of A):*

1. *$r(B) = r(A)$;*
2. *B is acyclic;*
3. *$|B| = |V| - c$, where c is the number of connected components of A.*

The flats of $M(G)$ correspond to the subgraphs of G in which every component is an induced subgraph of G. In other words, the geometric lattice corresponding to the graphic matroid $M(G)$ is precisely the connectivity lattice $K(G)$ introduced in Example 1.2.3.

Example 3.3.4. If G is a *forest* (a graph with no cycles), then no two vertices are joined by more than one path. Therefore, every edge set is a flat, and $M(G) \cong U_n(n)$. ◀

Example 3.3.5. If G is a cycle of length n, then every edge set of size $< n-1$ is a flat, but the closure of a set of size $n-1$ is the entire edge set. Therefore, $M(G) \cong U_{n-1}(n)$. ◀

Example 3.3.6. If $G = K_n$ (the complete graph on n vertices), then a flat of $M(G)$ is the same thing as an equivalence relation on $[n]$. Therefore, $M(K_n)$ is naturally isomorphic to the partition lattice Π_n. ◀

3.4 Matroid independence, basis and circuit systems

In addition to rank functions, lattices of flats, and closure operators, there are many other equivalent ways to define a matroid on a finite ground set E. In the fundamental example of a linear matroid M, some of these definitions correspond to linear-algebraic notions such as linear independence and bases.

[2]This terminology can cause confusion. By definition a subgraph H of G is spanning if $V(H) = V(G)$, but not every acyclic spanning subgraph is a spanning forest. A more accurate term would be "maximal forest".

Definition 3.4.1. A **(matroid) independence system** $\mathscr{I}$ is a family of subsets of E such that

(I1) $\emptyset \in \mathscr{I}$;
(I2) if $I \in \mathscr{I}$ and $I' \subseteq I$, then $I' \in \mathscr{I}$;
(I3) ("Donation") if $I, J \in \mathscr{I}$ and $|I| < |J|$, then there is some $x \in J \setminus I$ such that $I \cup x \in \mathscr{I}$.

Note that conditions **(I1)** and **(I2)** say that $\mathscr{I}$ is an abstract simplicial complex on E (see Example 1.1.12).

If E is a finite subset of a vector space, then the linearly independent subsets of E form a matroid independence system. Conditions **(I1)** and **(I2)** are clear. For **(I3)**, the span of J has greater dimension than that of I, so there must be some $x \in J$ outside the span of I, and then $I \cup x$ is linearly independent.

The next lemma generalizes the statement that any linearly independent set of vectors can be extended to a basis of any space containing it.

Lemma 3.4.2. *Let $\mathscr{I}$ be a matroid independence system on E. Suppose that $I \in \mathscr{I}$ and $I \subseteq X \subseteq E$. Then I can be extended to a maximum independent subset of X.*

Proof. If I already has maximum cardinality then we are done. Otherwise, let J be a maximum independent subset of X. Then $|J| > |I|$, so by **(I3)** there is some $x \in J \setminus I$ with $I \cup x$ independent. Replace I with $I \cup x$ and repeat. □

The argument shows also that for every $X \subseteq E$, all maximal independent subsets (or **bases**) of X have the same cardinality (so there is no irksome difference between "maximal" and "maximum"). In simplicial complex terms, every induced subcomplex of $\mathscr{I}$ is pure — an induced subcomplex is something of the form $\mathscr{I}|_X = \{I \in \mathscr{I} : I \subseteq X\}$, for $X \subseteq E$, and "pure" means that all maximal faces have the same cardinality. This condition actually characterizes matroid independence complexes; we will take this up again in §6.5.

A matroid independence system records the same combinatorial structure on E as a matroid rank function:

Proposition 3.4.3. *Let E be a finite set.*

1. *If r is a matroid rank function on E, then*
$$\mathscr{I} = \{A \subseteq E : \ r(A) = |A|\} \tag{3.5a}$$
is an independence system.
2. *If $\mathscr{I}$ is an independence system on E, then*
$$r(A) = \max\{|I| : \ I \subseteq A, \ I \in \mathscr{I}\} \tag{3.5b}$$
is a matroid rank function.
3. *These constructions are mutual inverses.*

Proof. Part 1: Let r be a matroid rank function on E and define $\mathscr{I}$ as in (3.5a). First, $r(I) \leq |I|$ for all $I \subseteq E$, so **(I1)** follows immediately. Second, suppose $I \in \mathscr{I}$ and $I' \subseteq I$; say $I' = \{x_1, \ldots, x_k\}$ and $I = \{x_1, \ldots, x_n\}$. Consider the "flag" (nested family of subsets)
$$\emptyset \subsetneq \{x_1\} \subsetneq \{x_1, x_2\} \subsetneq \cdots \subsetneq I' \subsetneq \cdots \subsetneq I.$$
The rank starts at 0 and increases at most 1 each time by submodularity. But since $r(I) = |I|$, it must increase by exactly 1 each time. In particular $r(I') = k = |I'|$ and so $I' \in \mathscr{I}$, establishing **(I2)**.

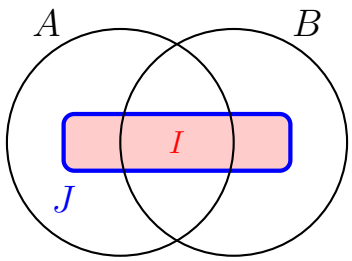
To show **(I3)**, let $I, J \in \mathscr{I}$ with $|I| < |J|$ and let $J \setminus I = \{x_1, \ldots, x_n\}$. If $n = 1$ then $J = I + x_1$ and there is nothing to show. Now suppose that $n \geq 2$ and $r(I + x_k) = r(I)$ for every $k \in [n]$. By submodularity,

$$\begin{aligned} r(I + x_1 + x_2) &\leq r(I + x_1) + r(I + x_2) - r(I) && = r(I), \\ r(I + x_1 + x_2 + x_3) &\leq r(I + x_1 + x_2) + r(I + x_3) - r(I) && = r(I), \\ &\cdots \\ r(I + x_1 + x_2 + \cdots + x_n) &\leq r(I + x_1 + \cdots + x_{n-1}) + r(I + x_n) - r(I) && = r(I), \end{aligned}$$

and equality must hold throughout. But then $r(I \cup J) = r(I) < r(J)$, which is a contradiction.

Part 2: Now suppose that $\mathscr{I}$ is an independence system on E, and define a function $r : 2^E \to \mathbb{Z}$ as in (3.5b). It is immediate from the definition that $r(A) \leq |A|$ and that $A \subseteq B$ implies $r(A) \leq r(B)$ for all $A, B \in \mathscr{I}$.

To prove submodularity, let $A, B \subseteq E$ and let I be a basis of $A \cap B$. By Lemma 3.4.2, we can extend I to a basis J of $A \cup B$. Note that no element of $J \setminus I$ can belong to both A and B, otherwise I would not be a maximal independent set in $A \cap B$. So we have the following Venn diagram:



Moreover, $J \cap A$ and $J \cap B$ are independent subsets of A and B respectively, but not necessarily maximal, so

$$r(A \cup B) + r(A \cap B) \;=\; |I| + |J| \;=\; |J \cap A| + |J \cap B| \;\leq\; r(A) + r(B). \qquad \square$$

If $M = M(G)$ is a graphic matroid, the associated independence system $\mathscr{I}$ is the family of *acyclic* edge sets in G. To see this, notice that if A is a set of edges and $e \in A$, then $r(A - e) < r(A)$ if and only if deleting e breaks a component of $G|_A$ into two smaller components (so that in fact $r(A - e) = r(A) - 1$). This is equivalent to the condition that e belongs to no cycle in A. Therefore, if A is acyclic, then deleting its edges one by one gets you down to $\emptyset$ and decrements the rank each time, so $r(A) = |A|$. On the other hand, if A contains a cycle, then deleting any of its edges won't change the rank, so $r(A) < |A|$.

Here's what the "donation" condition **(I3)** means in the graphic setting. Suppose that $|V| = n$, and let $c(H)$ denote the number of components of a graph H. If I, J are acyclic edge sets with $|I| < |J|$, then

$$c(G|_I) = n - |I| \;>\; c(G|_J) = n - |J|,$$

and there must be some edge $e \in J$ whose endpoints belong to different components of $G|_I$; that is, $I + e$ is acyclic.

The bases of M (the maximal independent sets) provide another way of defining a matroid.

Definition 3.4.4. A **(matroid) basis system on E** is a nonempty family $\mathscr{B} \subseteq 2^E$ such that for all $B, B' \in \mathscr{B}$,

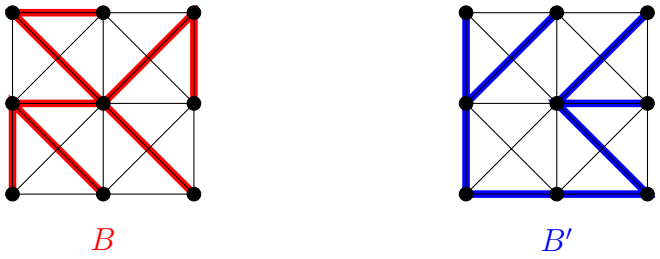
(B1) $|B| = |B'|$;
(B2) For all $e \in B \setminus B'$, there exists $e' \in B' \setminus B$ such that $(B - e) + e' \in \mathscr{B}$;

(B2′) For all $e \in B \setminus B'$, there exists $e' \in B' \setminus B$ such that $(B' + e) - e' \in \mathscr{B}$.

In fact, given **(B1)**, the conditions **(B2)** and **(B2′)** are equivalent, although this require some proof (Problem 3.2).

For example, if S is a finite set of vectors spanning a vector space V, then the subsets of S that are bases for V all have the same cardinality (namely $\dim V$) and satisfy the basis exchange condition **(B2)**.

If G is a graph, then the bases of $M(G)$ are its **spanning forests**, i.e., its maximal acyclic edge sets. If G is connected (which, as we will see, we may as well assume when studying graphic matroids) then the bases of $M(G)$ are its spanning trees.



Here is the graph-theoretic interpretation of **(B2)**. Let G be a connected graph, let B, B' be spanning trees, and let $e \in B \setminus B'$. Then $B - e$ has exactly two connected components. Since B' is connected, it must have some edge e' with one endpoint in each of those components, and then $B - e + e'$ is a spanning tree. See Figure 3.2.

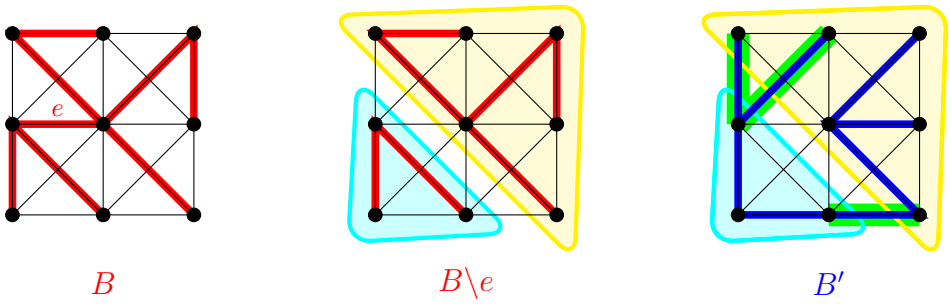


Figure 3.2: An example of basis axiom **(B2)** in a graphic matroid. The green edges are the possibilities for e' such that $B \backslash e + e'$ is a spanning tree.

As for **(B2′)**, if $e \in B \setminus B'$, then $B' + e$ must contain a unique cycle C (formed by e together with the unique path P in B' between the endpoints of e). Deleting any edge $e' \in P$ will produce a spanning tree, and there must be at least one such edge $e' \notin B$ (otherwise B contains the cycle C). See Figure 3.3.

If G is a graph with edge set E and $M = M(G)$ is its graphic matroid, then

$$\mathscr{I} = \{A \subseteq E \colon A \text{ is acyclic}\},$$
$$\mathscr{B} = \{A \subseteq E \colon A \text{ is a spanning forest of } G\}.$$

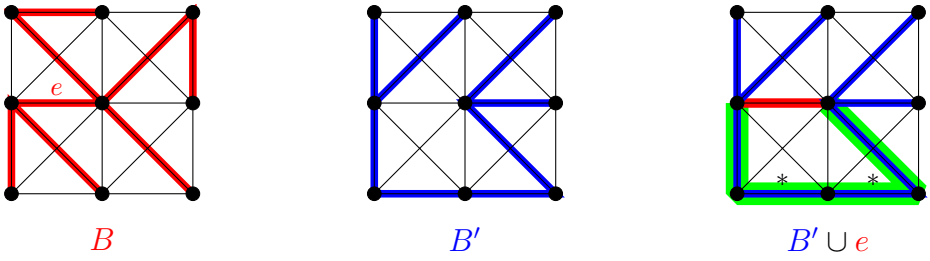


Figure 3.3: An example of basis axiom **(B2′)** in a graphic matroid. The path P is shown in green. The edges of $P \setminus B$, marked with stars, are valid choices for e'.

If S is a set of vectors and $M = M(S)$ is the corresponding linear matroid, then

$$\mathscr{I} = \{A \subseteq S \colon A \text{ is linearly independent}\},$$
$$\mathscr{B} = \{A \subseteq S \colon A \text{ is a basis for span}(S)\}.$$

Proposition 3.4.5. *Let E be a finite set.*

1. *If $\mathscr{I}$ is an independence system on E, then the family of maximal elements of $\mathscr{I}$ is a basis system.*
2. *If $\mathscr{B}$ is a basis system, then $\mathscr{I} = \bigcup_{B \in \mathscr{B}} 2^B$ is an independence system.*
3. *These constructions are mutual inverses.*

The proof is left as an exercise. We already have seen that an independence system on E is equivalent to a matroid rank function; Proposition 3.4.5 asserts that a basis system provides the same structure on E. Bases turn out to be especially convenient for describing fundamental operations on matroids such as duality, direct sum, and deletion/contraction (all of which are coming soon).

Instead of specifying the bases (maximal independent sets), a matroid can be defined by its minimal dependent sets, which are called **circuits**. These too can be axiomatized:

Definition 3.4.6. A **(matroid) circuit system on E** is a family $\mathscr{C} \subseteq 2^E$ such that, for all $C, C' \in \mathscr{C}$,

(C1) $\emptyset \notin \mathscr{C}$:
(C2) $C \not\subseteq C'$;
(C3) For all $e \in C \cap C'$, the set $(C \cup C') - e$ contains an element of $\mathscr{C}$.

In a linear matroid, the circuits are the minimal dependent sets of vectors. Indeed, if C, C' are such sets and $e \in C \cap C'$, then we can find two expressions for e as nontrivial linear combinations of vectors in C and in C', and equating these expressions and eliminating e shows that $(C \cup C') - e$ is dependent, hence contains a circuit.

In a graph, if two cycles C, C' meet in a (non-loop) edge $e = xy$, then $C - e$ and $C' - e$ are paths between x and y, so concatenating them forms a closed path. This path is not necessarily itself a cycle, but must contain some cycle.

Proposition 3.4.7. *Let E be a finite set.*

1. *If $\mathscr{I}$ is an independence system on E, then $\{C \notin \mathscr{I} \colon C' \in \mathscr{I}\ \forall C' \subsetneq C\}$ is a circuit system.*
2. *If $\mathscr{C}$ is a circuit system, then $\{I \subseteq E \colon C \not\subseteq I\ \forall C \in \mathscr{C}\}$ is an independence system.*
3. *These constructions are mutual inverses.*

In other words, the circuits are the minimal nonfaces of the independence complex (hence they correspond to the generators of the Stanley-Reisner ideal; see Defn. 6.3.1). The proof is left as an exercise.

The final definition of a matroid is different from what has come before, and gives a taste of the importance of matroids in combinatorial optimization.

Let E be a finite set and let Δ be an abstract simplicial complex on E (see Definition 3.4.1). Let $w : E \to \mathbb{R}_{\geq 0}$ be a function, which we regard as assigning weights to the elements of E, and for $A \subseteq E$, define $w(A) = \sum_{e \in A} w(e)$. Consider the problem of maximizing $w(A)$ over all subsets $A \in \Delta$; the maximum will certainly be achieved on a facet. A naive approach to find a maximal-weight A, which may or may not work for a given Δ and w, is the following "greedy" algorithm (known as **Kruskal's algorithm**):

1. Let $A = \emptyset$.
2. If A is a facet of Δ, stop.
   Otherwise, find $e \in E \setminus A$ of maximal weight such that $A + e \in \Delta$ (if there are several such e, pick one at random), and replace A with $A + e$.
3. Repeat step 2 until A is a facet of Δ.

Proposition 3.4.8. *Δ is a matroid independence system if and only if Kruskal's algorithm produces a facet of maximal weight for* every *weight function w.*

The proof is left as an exercise, as is the construction of a simplicial complex and a weight function for which the greedy algorithm does not produce a facet of maximal weight. This interpretation can be useful in algebraic combinatorics; see Example 9.19.2 below.

Summary of Matroid Axiomatizations

- *Geometric lattice*: lattice that is atomic and semimodular. Corresponds to a simple matroid.
- *Rank function*: monotonic function $r : 2^E \to \mathbb{N}$ such that $r(A) \leq |A|$ and $r(A) + r(B) \geq r(A \cup B) + r(A \cap B)$. Simple if $r(A) = |A|$ whenever $|A| \leq 2$.
- *Closure operator*: function $2^E \to 2^E$, $A \mapsto \bar{A}$ such that $A \subseteq \bar{A} = \bar{\bar{A}}$; $A \subseteq B \implies \bar{A} \subseteq \bar{B}$; and $x \notin \bar{A}$, $x \in \overline{A \cup y} \implies y \in \overline{A \cup x}$. Simple if $\bar{A} = A$ whenever $|A| \leq 1$.
- *Independence system*: set family $\mathscr{I} \subseteq 2^E$ such that $\emptyset \in \mathscr{I}$; $I \in \mathscr{I}$, $I' \subseteq I \implies I' \in \mathscr{I}$; and $I, J \in \mathscr{I}$, $|I| < |J| \implies \exists x \in J \setminus I : I \cup x \in \mathscr{I}$. Simple if $A \in \mathscr{I}$ whenever $|A| \leq 2$.
- *Basis system*: pure nonempty set family $\mathscr{B} \subseteq 2^E$ such that for all $B, B' \in \mathscr{B}$, $e \in B \setminus B'$: $\exists e' \in B' \setminus B$ with $B \setminus e \cup e' \in \mathscr{B}$. Simple if every element and every pair of elements belong to some basis.
- *Circuit system:* family $\mathscr{C} \subseteq 2^E$ of nonempty sets, none containing another; $C, C' \in \mathscr{C}$, $e \in C \cap C' \implies \exists C'' \in \mathscr{C} : C'' \subseteq (C \cup C') - e$. Simple if all elements have size at least 3.
- *Greedy algorithm:* simplicial complex Δ on E such that the greedy algorithm successfully constructs a maximum-weight facet for every weight function $w : E \to \mathbb{R}_{\geq 0}$.

3.5 Representability and regularity

The motivating example of a matroid is a finite collection of vectors in $\mathbb{R}^n$. What if we work over a different field? What if we turn this question on its head by specifying a matroid M purely combinatorially and then asking which fields give rise to vector sets whose matroid is M?

Definition 3.5.1. Let M be a matroid and V a vector space over a field $\Bbbk$. A set of vectors $S \subseteq V$ **represents** or **realizes M over** $\Bbbk$ if the linear matroid $M(S)$ associated with S is isomorphic to M.

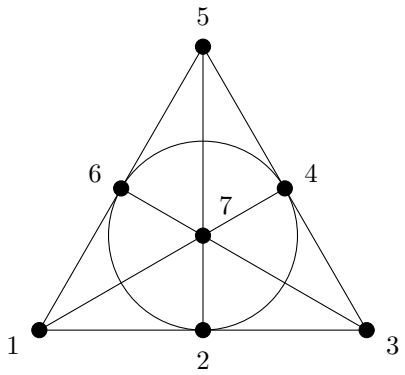
For example:

- The matroid $U_2(3)$ is representable over any field $\mathbb{F}$. Set $S = \{(1,0),\ (0,1),\ (1,1)\}$; any two of these vectors form a basis of $\mathbb{F}^2$.
- If $\Bbbk$ has at least three elements, then $U_2(4)$ is representable, by, e.g., $S = \{(1,0),\ (0,1),\ (1,1),\ (1,a)\}$. where $a \in \Bbbk \setminus \{0,1\}$. Again, any two of these vectors form a basis of $\Bbbk^2$.
- On the other hand, $U_2(4)$ is not representable over $\mathbb{F}_2$, because $\mathbb{F}_2^2$ doesn't contain four nonzero elements.

More generally, suppose that M is a simple matroid with n elements (i.e., the ground set E has $|E| = n$) and rank r (i.e., every basis of M has size r) that is representable over the finite field $\mathbb{F}_q$ of order q. Then each element of E must be represented by some nonzero vector in $\mathbb{F}_q^r$, and no two vectors can be scalar multiples of each other. Therefore,

$$n \leq \frac{q^r - 1}{q - 1}.$$

Example 3.5.2. *The Fano plane.* Consider the affine point configuration with 7 points and 7 lines (one of which looks like a circle), as shown:



This point configuration cannot be represented over $\mathbb{R}$. If you try to draw seven non-collinear points in $\mathbb{R}^2$ such that the six triples 123, 345, 156, 147, 257, 367 are each collinear, then 246 will not be collinear — try it. The same thing will happen over any field of characteristic $\neq 2$. On the other hand, over a field of characteristic 2, if the first six triples are collinear then 246 *must* be collinear. The configuration can be explicitly represented over $\mathbb{F}_2$ by the columns of the matrix

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix} \in (\mathbb{F}_2)^{3\times 7}$$

for which each of the seven triples of columns listed above is linearly dependent, and that each other triple is a column basis. (Note that over $\mathbb{R}$, the submatrix consisting of columns 2,4,6 has determinant 2.) The resulting matroid is called the **Fano plane** or **Fano matroid**. Note that each line in the Fano matroid corresponds to a 2-dimensional subspace of $\mathbb{F}_2^3$.

Viewed as a matroid, the Fano plane has rank 3. Its bases are the $\binom{7}{3} - 7 = 28$ noncollinear triples of points. Its circuits are the seven collinear triples and their complements (known as **ovals**). For instance, 4567 is an oval: it is too big to be independent, but on the other hand every three-element subset of it forms a basis (in particular, is independent), so it is a circuit.

The Fano plane is self-dual in the sense of discrete geometry[3]: the lines can be labeled $1, \ldots, 7$ so that point i lies on line j if and only if point j lies on line i. Here's how: recall that the points and lines of the Fano plane correspond respectively to 1- and 2-dimensional subspaces of $\mathbb{F}_2^3$, and assign the same label to orthogonally complementary spaces under the standard inner product. ◀

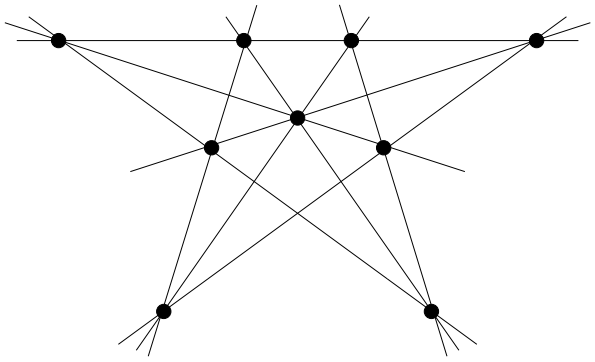
Example 3.5.3 (Finite projective planes)**.** Let $q \geq 2$ be a positive integer. A **projective plane of order q** consists of a collection P of points and a collection L of lines, each of which is a subset of P, such that:

- $|P| = |L| = q^2 + q + 1$;
- Each line contains $q + 1$ points, and each point lies in $q + 1$ lines;
- Any two points determine a unique line, and any two lines determine a unique point.

The Fano plane is thus a projective plane of order 2. More generally, if $\mathbb{F}_q$ is any finite field, then one can define a projective plane $\mathbb{P}_q^2$ whose points and lines are the 1- and 2-dimensional subspaces $\mathbb{F}_q^3$, respectively. Note that the number of lines is the number of nonzero vectors up to scalar multiplication, hence $(q^3 - 1)/(q - 1) = q^2 + q + 1$.

A notorious open question is whether any other finite projective planes exist. The best general result known is the *Bruck–Ryser–Chowla theorem* (1949), which states that if $q \equiv 1$ or $2 \pmod 4$, then q must be the sum of two squares. In particular, there exists no projective plane of order 6. Order 10 is also known to be impossible thanks to computer calculation, but the problem is open for other non-prime-power orders. It is also open whether there exists a projective plane of prime-power order that is not isomorphic to $\mathbb{P}_q^2$. One readily available survey of the subject is by Perrott [Per16]. ◀

Representability can be tricky. As we have seen, $U_2(4)$ can be represented over any field other than $\mathbb{F}_2$, while the Fano plane is representable only over fields of characteristic 2. The point configuration below is an affine representation of a rank-3 matroid over $\mathbb{R}$, but the matroid is not representable over $\mathbb{Q}$ [Grü03, pp. 93–94]. Put simply, it is impossible to construct a set of points with rational coordinates and exactly these collinearities.



A **regular** matroid is one that is representable over every field. (For instance, we will see that graphic matroids are regular.) For some matroids, the choice of field matters. For example, every uniform matroid is representable over every infinite field, but as we have seen before, $U_k(n)$ can be represented over $\mathbb{F}_q$ only if $n \leq (q^k - 1)/(q - 1)$. (For example, $U_2(4)$ is not representable over $\mathbb{F}_2$.) However, this inequality does not *suffice* for representability; as mentioned above, the Fano plane cannot be represented over, say, $\mathbb{F}_{101}$.

Recall that a *minor* of a matrix is the determinant of some square submatrix of M. A matrix is called **totally unimodular** if every minor is either 0, 1, or -1.

[3] But not self-dual as a matroid in the sense to be defined in §3.7.

Theorem 3.5.4. *A matroid M is regular if and only if it can be represented by the columns of a totally unimodular matrix.*

One direction is easy: if M has a unimodular representation then the coefficients can be interpreted as lying in any field, and the linear dependence of a set of columns does not depend on the choice of field (because $-1 \neq 0$ and $1 \neq 0$ in every field). The reverse direction is harder (see [Oxl92, chapter 6]), and the proof is omitted. In fact, something more is true: M is regular if and only if it is binary (representable over $\mathbb{F}_2$) and representable over at least one field of characteristic $\neq 2$.

Theorem 3.5.5. *Graphic matroids are regular.*

Proof. Let $G = (V, E)$ be a graph on vertex set $V = [n]$, and let $M = M(G)$ be the corresponding graphic matroid. We can represent M by the matrix X whose columns are the vectors $\mathbf{e}_i - \mathbf{e}_j$ for $ij \in E$. (Or $\mathbf{e}_j - \mathbf{e}_i$; it doesn't matter, since scaling a vector does not change the matroid.) Here $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is the standard basis for $\mathbb{R}^n$.

Consider any square submatrix X_{WB} of X with rows $W \subseteq V$ and columns $B \subseteq A$, where $|W| = |B| = k > 0$. If B contains a cycle $v_1, \ldots, v_k$ then the columns are linearly dependent, because

$$(\mathbf{e}_{v_1} - \mathbf{e}_{v_2}) + (\mathbf{e}_{v_2} - \mathbf{e}_{v_3}) + \cdots + (\mathbf{e}_{v_n} - \mathbf{e}_{v_1}) = 0,$$

so $\det X_{WB} = 0$. On the other hand, if B is acyclic, then I claim that $\det X_{WB} \in \{0, \pm 1\}$, which we will prove by induction on k. The base case $k = 1$ follows because all entries of X are 0 or ± 1. For $k > 1$, if there is some vertex of W with no incident edge in B, then the corresponding row of X_{WB} is zero and the determinant vanishes. Otherwise, by the handshaking theorem, there must be some vertex $w \in W$ incident to exactly one edge $b \in B$. The corresponding row of X_{WB} will have one entry ± 1 and the rest zero. Expanding on that row gives $\det X_{WB} = \pm \det_{W \setminus w, B \setminus b}$, and we are done by induction. The same argument shows that any set of columns corresponding to an acyclic edge set will in fact be linearly independent. □

Example 3.5.6. The matrix

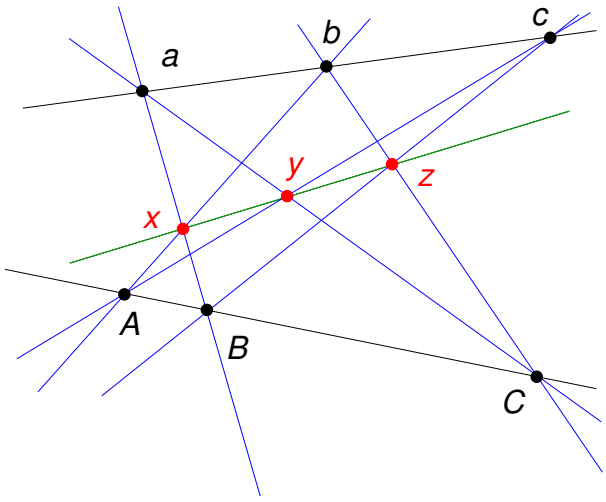
$$\begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & -1 \end{bmatrix}$$

represents $U_2(4)$ over any field of characteristic $\neq 2$, but the last two columns are dependent (in fact equal) in characteristic 2. ◀

Example 3.5.7. There exist matroids that are not representable over *any* field. The smallest ones have ground sets of size 8; one of these is the rank-4 *Vámos matroid* V_8 [Oxl92, p. 511]. The smallest rank-3 example is the **non-Pappus matroid**.

Pappus' Theorem from Euclidean geometry says that if a, b, c, A, B, C are distinct points in $\mathbb{R}^2$ such that a, b, c and A, B, C are collinear, then x, y, z are collinear, where

$$x = \overline{aB} \cap \overline{Ab}, \qquad y = \overline{aC} \cap \overline{Ac}, \qquad z = \overline{bC} \cap \overline{Bc}.$$



Accordingly, there is a rank-3 simple matroid on ground set $E = \{a, b, c, A, B, C, x, y, z\}$ whose flats are

$$\emptyset, \qquad a, b, c, a, b, c, x, y, z, \qquad abc,\ ABC,\ aBx,\ Abx,\ aCy,\ Acy,\ bCz,\ Bcz,\ xyz, \qquad E.$$

It turns out that deleting xyz from this list produces the family of closed sets of a matroid, called the *non-Pappus matroid* NP. Since Pappus' theorem can be proven using analytic geometry, and the equations that say that x, y, z are collinear are valid over any field (i.e., involve only ± 1 coefficients), it follows that NP is not representable over any field. ◀

3.6 Direct sum

There are several ways to construct new matroids from old ones. We will begin with a boring but useful one (direct sum) and then move on to the more exciting constructions of duality and deletion/contraction.

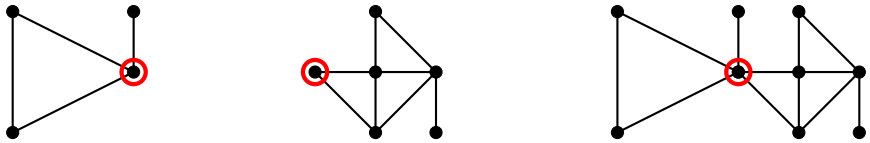
Definition 3.6.1. Let M_1, M_2 be matroids on disjoint sets E_1, E_2, with basis systems $\mathscr{B}_1, \mathscr{B}_2$. The **direct sum** $M_1 \oplus M_2$ is the matroid on $E_1 \cup E_2$ with basis system

$$\mathscr{B} = \{B_1 \cup B_2 \colon B_1 \in \mathscr{B}_1,\ B_2 \in \mathscr{B}_2\}.$$

I will omit the routine proof that $\mathscr{B}$ is a basis system.

If M_1, M_2 are linear matroids whose ground sets span vector spaces V_1, V_2 respectively, then $M_1 \oplus M_2$ is the matroid you get by regarding the vectors as living in $V_1 \oplus V_2$: the linear relations have to come either from V_1 or from V_2.

If G_1, G_2 are graphs, then $M(G_1) \oplus M(G_2) \cong M(G_1 + G_2)$, where $+$ denotes disjoint union. Actually, you can identify any vertex of G_1 with any vertex of G_2 and still get a graph whose associated graphic matroid is $M(G_1) \oplus M(G_2)$ (such as G in the following figure).



G_1 G_2 G

A useful corollary is that every graphic matroid arises from a *connected* graph. Actually, there may be many different connected graphs that give rise to the same matroid, since in the previous construction it did not matter which vertices of G_1 and G_2 were identified. This raises an interesting question: when does the isomorphism type of a graphic matroid $M(G)$ determine the graph G up to isomorphism?

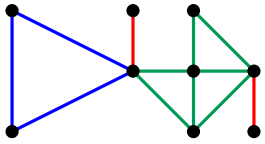
Definition 3.6.2. A matroid that cannot be written nontrivially as a direct sum of two smaller matroids is called **connected** or **indecomposable**.[4]

Proposition 3.6.3. *Let $G = (V, E)$ be a loopless graph. Then $M(G)$ is indecomposable if and only if G is* ***2-connected*** *— i.e., not only is it connected, but so is every subgraph obtained by deleting a single vertex.*

The "only if" direction is immediate: the discussion above implies that

$$M(G) = \bigoplus_H M(H)$$

where H ranges over all the **blocks** (maximal 2-connected subgraphs) of G.



We will prove the other direction later (maybe).

Remark 3.6.4. If $G \cong H$ as graphs, then clearly $M(G) \cong M(H)$. The converse is not true: if T is any tree (or even forest) on n vertices, then every set of edges is acyclic, so the independence complex is the Boolean lattice $2^{[n]}$ (and, for that matter, so is the lattice of flats).

In light of Proposition 3.6.3, it is natural to suspect that every 2-connected graph is determined up to isomorphism by its graphic matroid, but even this is not true; the two 2-connected graphs below are not isomorphic, but have isomorphic graphic matroids.

As you should expect from an operation called "direct sum," properties of $M_1 \oplus M_2$ should be easily deducible from those of its summands. In particular, direct sum is easy to describe in terms of the other matroid axiomatizations we have studied. It is additive on rank functions: if $A_1 \subseteq E_1$ and $A_2 \subseteq E_2$, then

$$r_{M_1 \oplus M_2}(A_1 \cup A_2) = r_{M_1}(A_1) + r_{M_2}(A_2).$$

[4]The first term is more common among matroid theorists, but I prefer "indecomposable" to avoid potential confusion with the graph-theoretic meaning of "connected".

Similarly, the closure operator is $\overline{A_1 \cup A_2} = \overline{A_1} \cup \overline{A_2}$. The circuit system of the direct sum is just the (necessarily disjoint) union of the circuit systems of the summands. Finally, the geometric lattice of a direct sum is just the poset product of the lattices of the summands, i.e.,

$$L(M_1 \oplus M_2) \cong L(M_1) \times L(M_2),$$

subject to the order relations $(F_1, F_2) \leq (F_1', F_2')$ iff $F_i \leq F_i'$ in $L(M_i)$ for each i.

3.7 Duality

Definition 3.7.1. Let M be a matroid on ground set $|E|$ with basis system $\mathscr{B}$. The **dual matroid** of M (also known as the **orthogonal matroid**) is the matroid M^* on E with basis system

$$\mathscr{B}^* = \{E \setminus B \colon B \in \mathscr{B}\}.$$

We often write e^* for elements of the ground set when talking about their behavior in the dual matroid.

Clearly the elements of $\mathscr{B}^*$ all have cardinality $|E| - r(M)$ (where r is the rank), and complementation swaps the basis exchange conditions **(B2)** and **(B2′)**, so if you believe that those conditions are logically equivalent (Problem 3.2) then you also believe that $\mathscr{B}^*$ is a matroid basis system.

It is immediate from the definition that $(M^*)^* = M$. In addition, the independent sets of M are the complements of the spanning sets of M^* (since $A \subseteq B$ for some $B \in \mathscr{B}$ if and only if $E \setminus A \supseteq E \setminus B$), and vice versa. The rank function r^* of the dual is given by

$$r^*(A) = r(E \setminus A) + |A| - r(E) \qquad \forall A \subseteq E. \tag{3.6}$$

The proof is left as Problem 3.8.

The dual of a vector matroid has an explicit description. Let $E = \{v_1, \ldots, v_n\} \subseteq \Bbbk^r$, and let $M = M(E)$. We may as well assume that E spans $\Bbbk^r$, so $r \leq n$, and the representing matrix $X = [v_1 | \cdots | v_n] \in \Bbbk^{r \times n}$ has full row rank r.

Let Y be any $(n-r) \times n$ matrix with $\mathrm{rowspace}(Y) = \mathrm{nullspace}(X)$. That is, the rows of Y span the orthogonal complement of $\mathrm{rowspace}(X)$ with respect to the standard inner product.

Theorem 3.7.2. *With this setup, the columns of Y are a representation for M^*.*

Before proving this theorem, we will do an example that will make things clearer.

Example 3.7.3. Let $E = \{v_1, \ldots, v_5\}$ be the set of column vectors of the following matrix (over $\mathbb{R}$, say):

$$X = \begin{bmatrix} 1 & 0 & 0 & 2 & 1 \\ 0 & 1 & 0 & 2 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

Notice that X has full row rank (it's in row-echelon form, after all), so it represents a matroid of rank 3 on 5 elements. We could take Y to be the matrix

$$Y = \begin{bmatrix} 0 & 0 & 0 & 1 & -2 \\ 1 & 1 & 0 & 0 & -1 \end{bmatrix}.$$

Then Y has rank 2. Call its columns $\{v_1^*, \ldots, v_5^*\}$; then the column bases are

$$\{v_1^*, v_4^*\},\ \{v_1^*, v_5^*\},\ \{v_2^*, v_4^*\},\ \{v_2^*, v_5^*\},\ \{v_4^*, v_5^*\},$$

whose (unstarred) complements (e.g., $\{v_2, v_3, v_5\}$, etc.) are precisely the column bases for X. In particular, every basis of M contains v_3 (so v_3 is a coloop), which corresponds to the fact that no basis of M^* contains v_3^* (so v_3^* is a loop). This makes sense linear-algebraically: v_3 is linearly independent of all the columns, so no vector with a nonzero entry in the 3rd position is orthogonal to any row of M, so v_3^* is the zero vector. ◀

Proof of Theorem 3.7.2. First, note that invertible row operations on a matrix $X \in \Bbbk^{r \times n}$ (i.e., multiplication on the left by an element of $GL_r(\Bbbk)$) do not change the matroid represented by its columns; they simply change the basis of $\Bbbk^r$.

Let B be a basis of M, and reindex so that $B = \{v_1, \ldots, v_r\}$. We can then perform invertible row-operations to put X into *reduced* row-echelon form, i.e.,

$$X \ = \ [I_r \mid A]$$

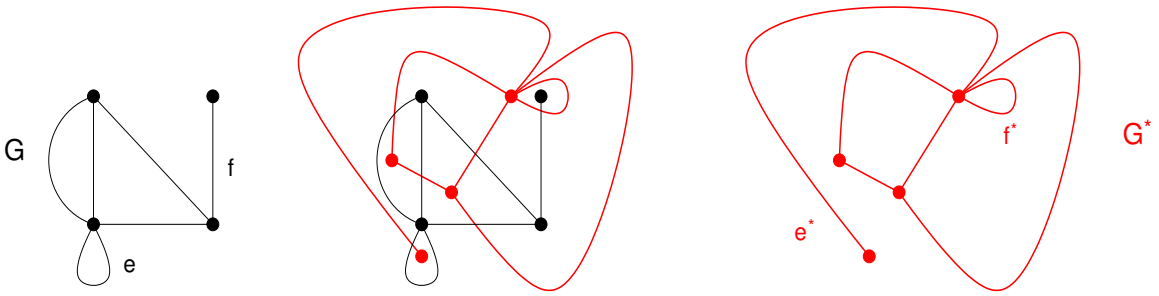
where I_r is the $r \times r$ identity matrix and A is arbitrary. It is easy to check that $\text{nullspace}\, X = (\text{rowspace}\, X^*)^T$, where

$$X^* \ = \ [-A^T \mid I_{n-r}],$$

(this is a standard recipe). But then the last $n-r$ elements of X^*, i.e., $E^* \setminus B^*$, are clearly a column basis. By the same logic, every basis of X is the complement of a column basis of Y, and the converse is true because X can be obtained from X^* in the same way that X^* is obtained from X. Therefore the columns of X and X^* represent dual matroids. Meanwhile, any matrix Y with the same rowspace as X^* can be obtained from it by invertible row operations, hence represents the same matroid. □

In particular, representability over a particular field is unchanged by dualization.

Duality and graphic matroids. Let G be a connected *planar* graph, i.e., one that can be drawn in the plane with no crossing edges. The **planar dual** is the graph G^* whose vertices are the regions into which G divides the plane, with two vertices of G^* joined by an edge e^* if the corresponding faces of G are separated by an edge e of G. (So e^* is drawn *across* e in the construction.)



Some facts to check about planar duality:

- $A \subseteq E$ is acyclic if and only if $E^* \setminus A^*$ is connected.
- $A \subseteq E$ is connected if and only if $E^* \setminus A^*$ is acyclic.
- G^{**} is naturally isomorphic to G.
- e is a loop (bridge) if and only if e^* is a bridge (loop).

If G is not planar then in fact $M(G)^*$ is not a graphic matroid (although it is certainly regular).

Definition 3.7.4. Let M be a matroid on E. A **loop** is an element of E that does not belongs to any basis of M. A **coloop** is an element of E that belongs to every basis of M. An element of E that is neither a loop nor a coloop is called **ordinary** (probably not standard terminology, but natural and useful).

In a linear matroid, a loop is a copy of the zero vector, while a coloop is a vector that is not in the span of all the other vectors.

A **cocircuit** of M is by definition a circuit of the dual matroid M^*. A matroid can be described by its cocircuit system, which satisfy the same axioms as those for circuits (Definition 3.4.6). Set-theoretically, a cocircuit is a minimal set not contained in any basis of M^*, so it is a minimal set that intersects every basis of M nontrivially. For a connected graph G, the cocircuits of the graphic matroid $M(G)$ are the **bonds** of G: the minimal edge sets K such that $G - K$ is not connected. Every bond C^* is of the following form: there is a partition $V(G) = X \cup Y$ such that C^* is the set of edges with one endpoint in each of X and Y, and both $G|_X$ and $G|_Y$ are connected.

3.8 Deletion and contraction

Definition 3.8.1. Let M be a matroid on E with independence system $\mathscr{I}$, and let $e \in E$.

1. The **deletion** of e is the matroid $M\backslash e$ on $E - e$ with independence system $\{I \in \mathscr{I} : e \notin I\}$.
2. The **contraction** of e is the matroid M/e on $E-e$ is defined as follows. If e is a loop then $\mathscr{I}(M/e) = \mathscr{I}$. Otherwise, the independence system of M/e is $\{I \subset E - e : I + e \in \mathscr{I}\}$.

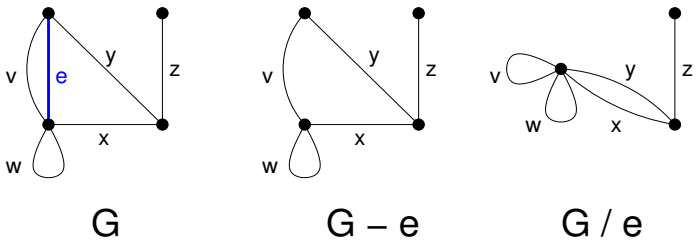
We can also describe deletion and contraction on the level of basis systems:

$$\mathscr{B}(M\backslash e) = \begin{cases} \{B \in \mathscr{B}(M) : e \notin B\} & \text{if } e \text{ is not a coloop,} \\ \{B - e : B \in \mathscr{B}(M)\} & \text{if } e \text{ is a coloop,} \end{cases}$$

$$\mathscr{B}(M/e) = \begin{cases} \{B - e : B \in \mathscr{B}(M),\ e \in B\} & \text{if } e \text{ is not a loop,} \\ \{B : B \in \mathscr{B}(M)\} & \text{if } e \text{ is a loop.} \end{cases}$$

Again, the terms come from graph theory. Deleting an edge e of a graph G means removing it from the graph, while contracting an edge means to shrink it down so that its two endpoints merge into one. The resulting graphs are called $G\backslash e$ and G/e, and these operations are consistent with the effect on graphic matroids, i.e.,

$$M(G\backslash e) = M(G)\backslash e, \qquad M(G/e) = M(G)/e. \tag{3.7}$$



Notice that contracting can cause some edges to become parallel, and can cause other edges (namely, those parallel to the edge being contracted) to become loops. In matroid language, deleting an element from a simple matroid always yields a simple matroid, but the same is not true for contraction.

We can define deletion and contraction of sets as well as single elements. To delete (resp., contract) a set, simply delete (resp., contract) each of its elements in some order.

Proposition 3.8.2. *Let M be a matroid on E.*

1. *For each $A \subseteq E$, the deletion $M\backslash A$ and contraction M/A are well-defined (i.e., do not depend on the order in which elements of A are deleted or contracted).*
2. *In particular*

$$\mathscr{I}(M\backslash A) = \{I \subseteq E \setminus A \colon I \in \mathscr{I}(M)\},$$
$$\mathscr{I}(M/A) = \{I \subseteq E \setminus A \colon I \cup B \in \mathscr{I}(M)\}$$

   where B is some (or, equivalently, any) basis of $M|_A$.
3. *Deletion and contraction commute in the following sense: for every $e, f \in E$ we have $(M/e)\backslash f \cong (M\backslash f)/e$.*
4. *Deletion and contraction are interchanged by duality:*

$$(M\backslash e)^* \cong M^*/e^* \quad \textit{and} \quad (M/e)^* \cong M^*\backslash e^*.$$

The proof is left as an exercise.

Here is what deletion and contraction mean for vector matroids. Let V be a vector space over a field $\Bbbk$, let $E \subseteq V$ be a set of vectors spanning V, let $M = M(E)$, and let $e \in E$. Then:

1. $M\backslash e = M(E - e)$. (If we want to preserve the condition that the ground set spans the ambient space, then e must not be a coloop.)
2. M/e is the matroid represented by the images of $E - e$ in the quotient space $V/\Bbbk e$. (Note that if e is a loop then this quotient space is just V itself.)

Thus both operations preserve representability over $\Bbbk$. For instance, to find an explicit representation of M/e, apply a change of basis to V so that e is the ith standard basis vector, then simply erase the ith coordinate of every vector in $E - e$.

Any matroid M' obtained from M by some sequence of deletions and contractions is called a **minor** of M.

Proposition 3.8.3. *Every minor of a graphic (resp., linear, uniform) matroid is graphic (resp., linear, uniform).*

Proof. The graphic case follows from (3.7), and the linear case from the previous discussion. For uniform matroids, the definitions imply that

$$U_k(n)\backslash e \cong U_k(n-1) \quad \text{and} \quad U_k(n)/e \cong U_{k-1}(n-1)$$

for every ground set element e. □

.

Many invariants of matroids can be expressed recursively in terms of deletion and contraction. The following fact is immediate from Definition 3.8.1.

Proposition 3.8.4. *Let M be a matroid on ground set E, and let $b(M)$ denote the number of bases of M. Let $e \in E$; then*

$$b(M) = \begin{cases} b(M\backslash e) & \text{if } e \text{ is a loop;} \\ b(M/e) & \text{if } e \text{ is a coloop;} \\ b(M\backslash e) + b(M/e) & \text{otherwise.} \end{cases}$$

Example 3.8.5. If $M \cong U_k(n)$, then $b(M) = \binom{n}{k}$, and the recurrence of Proposition 3.8.4 is just the Pascal relation $\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$. ◀

Many other matroid invariants satisfy analogous recurrences involving deletion and contraction. In fact, Proposition 3.8.4 is the tip of an iceberg that we will explore in Chapter 4.

3.9 Exercises

Problem 3.1. Determine, with proof, all pairs of integers $k \leq n$ such that there exists a graph G with $M(G) \cong U_k(n)$. (Here $U_k(n)$ denotes the uniform matroid of rank k on n elements; see Example 3.2.5.). Hint: Use Proposition 3.8.3.

Problem 3.2. Prove the equivalence of the two forms of the basis exchange condition **(B2)** and **(B2′)**. (Hint: Examine $|B \setminus B'|$.)

Problem 3.3. Let $\mathscr{A}$ be a nonempty family of subsets of a finite set E. Prove that $\mathscr{A}$ is a matroid basis system if and only if the following two conditions hold:

(A1) No element of $\mathscr{A}$ is contained in another element.
(A2) If $A_1, A_2 \in \mathscr{A}$ and there exist sets $X, Y \subseteq E$ such that (i) $X \subseteq A_1$, (ii) $A_2 \subseteq Y$, and (iii) $X \subseteq Y$, then there exists $A \in \mathscr{A}$ such that $X \subseteq A \subseteq Y$.

Note: This is a special case of something called a "poset matroid" [BNP98].

Problem 3.4. (Proposed by Kevin Adams.) Let B, B' be bases of a matroid M. Prove that there exists a bijection $\phi : B \setminus B' \to B' \setminus B$ such that $B - e + \phi(e)$ is a basis of M for every $e \in B \setminus B'$.

Problem 3.5. Prove Proposition 3.4.5, which describes the cryptomorphism between matroid independence systems and matroid basis systems.

Problem 3.6. Prove Proposition 3.4.7, which describes the cryptomorphism between matroid independence systems and matroid circuit systems. (Hint: The hardest part is showing that if $\mathscr{C}$ is a matroid circuit system then the family $\mathscr{I}$ of sets containing no circuit satisfies **(I3)**. Under the assumption that **(I3)** fails for some pair I, J with $|I| < |J|$, use circuit exchange to build a sequence of collections of circuits in $I \cup J$ that avoid more and more elements of I, eventually producing a circuit in J and thus producing a contradiction.)

Problem 3.7. Let M be a matroid on ground set E. Suppose there is a partition of E into disjoint sets $E_1, \ldots, E_n$ such that $r(E) = r(E_1) + \cdots + r(E_k)$. Prove that $M = \bigoplus_{i=1}^{n} M_i$, where $M_i = M|_{E_i}$. (Note: This fact provides an algorithm, albeit not necessarily an efficient one, for testing whether a matroid is connected.)

Problem 3.8. Let M be a matroid on ground set E with rank function $r : 2^E \to \mathbb{N}$. Prove that the rank function r^* of the dual matroid M^* is given by $r^*(A) = r(E \setminus A) + |A| - r(E)$ for all $A \subseteq E$.

Problem 3.9. Let M be a matroid on E. A set $S \subseteq E$ is called *spanning* if it contains a basis. Let $\mathscr{S}$ be the set of all spanning sets.

(a) Express $\mathscr{S}$ in terms of (i) the rank function r of M; (ii) its closure operator $A \mapsto \bar{A}$; (iii) its lattice of flats L. (You don't have to prove anything — just give the construction.)
(b) Formulate axioms that could be used to define a matroid via its system of spanning sets. (Hint: Describe spanning sets in terms of the dual matroid M^*.)

Problem 3.10. Let E be a finite set and let Δ be an abstract simplicial complex on E. Let $w : E \to \mathbb{R}_{\geq 0}$ be any function; think of $w(e)$ as the "weight" of e. For $A \subseteq E$, define $w(A) = \sum_{e \in A} w(e)$. Consider the problem of maximizing $w(A)$ over all facets A. A naive approach is the following *greedy algorithm*:

> **Step 1:** Let $A = \emptyset$.
> **Step 2:** If A is a facet of Δ, stop.
> Otherwise, find $e \in E \setminus A$ of maximal weight such that $A + e \in \Delta$
> (if there are several such e, pick one at random), and replace A with $A + e$.
> **Step 3:** Repeat Step 2 until A is a facet of Δ.

This algorithm may or may not work for a given Δ and w. Prove the following facts:

(a) Construct a simplicial complex and a weight function for which this algorithm does not produce a facet of maximal weight. (Hint: The smallest example has $|E| = 3$.)
(b) Prove that the following two conditions are equivalent:
  (i) The greedy algorithm produces a facet of maximal weight for *every* weight function w.
  (ii) Δ is a matroid independence system.

Note: This result does follow from Theorem 6.5.1. However, that is a substantial result, so don't use it unless you first do Problem 6.9. It is possible to do this exercise by working directly with the definition of a matroid independence system.

Problem 3.11. Prove Proposition 3.8.2.

Problem 3.12. Let X and Y be disjoint sets of vertices, and let B be an X, Y-bipartite graph: that is, every edge of B has one endpoint in each of X and Y. For $V = \{x_1, \ldots, x_n\} \subseteq X$, a *transversal* of V is a set $W = \{y_1, \ldots, y_n\} \subseteq Y$ such that $x_i y_i$ is an edge of B. (The set of all edges $x_i y_i$ is called a *matching*.) Let $\mathscr{I}$ be the family of all subsets of X that have a transversal; in particular $\mathscr{I}$ is a simplicial complex.

Prove that $\mathscr{I}$ is in fact a matroid independence system by verifying that the donation condition **(I3)** holds. (Suggestion: Write down an example or two of a pair of independent sets I, J with $|I| < |J|$, and use the corresponding matchings to find a systematic way of choosing a vertex that J can donate to I.) These matroids are called **transversal matroids**; along with linear and graphic matroids, they are the other "classical" examples of matroids in combinatorics.

Problem 3.13. Fix positive integers $n \geq r$.

(a) Let $\mathscr{B}$ be a family of subsets of $\binom{[n]}{r}$ satisfying the following property: if $C, C' \in \binom{[n]}{r} \setminus \mathscr{B}$ and $C \neq C'$, then $|C \triangle C'| \geq 4$. Prove that $\mathscr{B}$ is a matroid basis system. (Such a matroid is called a *sparse paving matroid*.)
(b) More generally, let $j \in \{0, \ldots, n - r\}$ and let $\mathscr{S} \subset \binom{[n]}{k+j}$, such that $|S \triangle S'| \geq 2j + 4$ for all $S, S' \in \mathscr{S}$. Prove that

$$\mathscr{B} = \left\{ B \in \binom{[n]}{k} \;\middle|\; B \not\subseteq S \; \forall S \in \mathscr{S} \right\}$$

is a matroid basis system. (The $j = 0$ case is part (a).) This construction is due to George Nasr.

Problem 3.14. [Oxl92, Prop. 1.15.14] Let M be a matroid on E with basis system $\mathscr{B}$ and rank r. Let $A \subseteq E$ be a circuit that is also a **hyperplane** (that is, a flat of rank $r(E) - 1$). Prove that $\mathscr{B} \cup \{A\}$ is also a matroid basis system. (This is called the **relaxation** of M by A.)

Problem 3.15. (Requires a bit of abstract algebra.) Let n be a positive integer, and let ζ be a primitive nth root of unity. The *cyclotomic matroid* Y_n is represented over $\mathbb{Q}$ by the numbers $1, \zeta, \zeta^2, \ldots, \zeta^{n-1}$, regarded as elements of the cyclotomic field extension $\mathbb{Q}(\zeta)$. Thus, the rank of Y_n is the dimension of $\mathbb{Q}(\zeta)$ as a $\mathbb{Q}$-vector space, which is given by the Euler ϕ function. Prove the following:

(a) if n is prime, then $Y_n \cong U_{n-1}(n)$.
(b) if m is the squarefree part of n (i.e., the product of all the distinct primes dividing n — e.g., the squarefree part of $56 = 2^3 \cdot 7$ is $2 \cdot 7 = 14$) then Y_n is the direct sum of n/m copies of Y_m.
(c) if $n = pq$, where p, q are distinct primes, then $Y_n \cong M(K_{p,q})^*$ — that is, the dual of the graphic matroid of the complete bipartite graph $K_{p,q}$.

This problem is near and dear to my heart; the answer (more generally, a characterization of Y_n for all n) appears in [MR05].

Problem 3.16. (A theorem of Crapo of 1969, brought to my attention by Dania Morales.) Recall the definition of internal and external activity (Defn. 4.3.2). Let M be a matroid, and fix a total order on its ground set E. Prove that every independent set of M can be written in exactly one way as $B \setminus Y$, where B is a basis and Y is a set of internally active elements of B.

Chapter 4

The Tutte Polynomial

Throughout this section, let M be a matroid on ground set E with rank function r, and let $n = |E|$.

4.1 The two definitions of the Tutte polynomial

For each subset $A \subseteq E$, define

$$\begin{aligned} \operatorname{corank} A &= r(E) - r(A), \\ \operatorname{nullity} A &= |A| - r(A). \end{aligned}$$

Corank measures how far A is from being spanning, and nullity measures how far A is from being independent. That is, the corank is the minimum number of elements needed to adjoin to A to produce a spanning set (i.e., to intersect all cocircuits), while the nullity is the minimum number of elements needed to delete from A to produce an independent set (i.e., to break all circuits).

Definition 4.1.1. The **Tutte polynomial** of M is

$$T_M = T_M(x, y) = \sum_{A \subseteq E} (x-1)^{r(E)-r(A)} (y-1)^{|A|-r(A)}. \tag{4.1}$$

Example 4.1.2. If $E = \emptyset$ then $T_M(x, y) = 1$. Mildly less trivially, if every element is a coloop, then $r(A) = |A|$ for all A, so

$$T_M = \sum_{A \subseteq E} (x-1)^{n-|A|} = (x-1+1)^n = x^n$$

by the binomial theorem. If every element is a loop, then the rank function is identically zero and we get

$$T_M = \sum_{A \subseteq E} (y-1)^{|A|} = y^n.$$

◀

Example 4.1.3. For uniform matroids, corank and nullity depend only on cardinality, making their Tutte polynomials easy to compute. $U_1(2)$ has one set with corank 1 and nullity 0 (the empty set), two singleton sets with corank 0 and nullity 0, and one doubleton with corank 0 and nullity 1, so

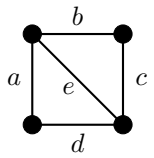
$$T_{U_1(2)} = (x-1) + 2 + (y-1) = x + y.$$

Similarly,

$$T_{U_1(3)} = (x-1)+3+3(y-1)+(y-1)^2 = x+y+y^2,$$
$$T_{U_2(3)} = (x-1)^2+3(x-1)+3+(y-1) = x^2+x+y.$$

◀

Example 4.1.4. Let G be the graph below (known as the "diamond"):



The formula (4.1) gives

A	$\lvert A\rvert$	$r(A)$	$3-r(A)$	$\lvert A\rvert - r(A)$	contribution to (4.1)
1 empty set	0	0	3	0	$1(x-1)^3(y-1)^0 = x^3-3x^2+3x-1$
5 singletons	1	1	2	0	$5(x-1)^2(y-1)^0 = 5x^2-10x+5$
10 doubletons	2	2	1	0	$10(x-1)^1(y-1)^0 = 10x-10$
2 triangles	3	2	1	1	$2(x-1)^1(y-1)^1 = 2xy-2x-2y+2$
8 spanning trees	3	3	0	0	$8(x-1)^0(y-1)^0 = 8$
5 quadrupletons	4	3	0	1	$5(x-1)^0(y-1)^1 = 5y-5$
1 whole set	5	3	0	2	$1(x-1)^0(y-1)^2 = y^2-2y+1$
Total					$x^3+2x^2+x+2xy+y^2+y$

◀

Many invariants of M can be obtained by specializing the variables x, y appropriately. Some easy ones:

1. $T_M(2,2) = \sum_{A\subseteq E} 1 = 2^{|E|}$. (Or, if you like, $|E| = \log_2 T_M(2,2)$.)
2. Consider $T_M(1,1)$. This kills off all summands whose corank is nonzero (i.e., all non-spanning sets) and whose nullity is nonzero (i.e., all non-independent sets). What's left are the bases, each of which contributes a summand of 1. So $T_M(1,1) = b(M)$, the number of bases. We previously observed that this quantity satisfies a deletion/contraction recurrence (Prop. 3.8.4); this will show up again soon.
3. Similarly, $T_M(1,2)$ and $T_M(2,1)$ count respectively the number of spanning sets and the number of independent sets.
4. A little more generally, we can enumerate independent and spanning sets by their cardinality:

$$\sum_{A\subseteq E \text{ independent}} q^{|A|} = q^{r(M)}T(1/q+1,1);$$
$$\sum_{A\subseteq E \text{ spanning}} q^{|A|} = q^{r(M)}T(1,1/q+1).$$

5. $T_M(0,1)$ is (up to a sign) the reduced Euler characteristic (see (6.2)) of the independence complex

of M:

$$\begin{aligned} T_M(0,1) &= \sum_{A\subseteq E} (-1)^{r(E)-r(A)} 0^{|A|-r(A)} = \sum_{A\subseteq E \text{ independent}} (-1)^{r(E)-r(A)} \\ &= (-1)^{r(E)} \sum_{A\in\mathscr{I}(M)} (-1)^{|A|} \\ &= (-1)^{r(E)-1} \tilde{\chi}(\mathscr{I}(M)). \end{aligned} \tag{4.2}$$

The fundamental theorem about the Tutte polynomial is that it satisfies a deletion/contraction recurrence. In a sense it is the most general such invariant — we will give a "recipe theorem" that expresses any deletion/contraction invariant as a Tutte polynomial specialization (more or less).

Theorem 4.1.5. *The Tutte polynomial satisfies (and can be computed by) the following* ***Tutte recurrence****:*

(T1) *If $E = \emptyset$, then $T_M = 1$.*
(T2a) *If $e \in E$ is a loop, then $T_M = yT_{M\setminus e}$.*
(T2b) *If $e \in E$ is a coloop, then $T_M = xT_{M/e}$.*
(T3) *If $e \in E$ is ordinary, then $T_M = T_{M\setminus e} + T_{M/e}$.*

We can use this recurrence to compute the Tutte polynomial, by picking one element at a time to delete and contract. The miracle is that *it doesn't matter what order we choose on the elements of E* — all orders will give the same final result! (In the case that M is a uniform matroid, then it *is* clear at this point that T_M is well-defined by the Tutte recurrence, because, up to isomorphism, $M\setminus e$ and M/e are independent of the choices of $e \in E$.)

Before proving the theorem, here are some examples.

Example 4.1.6. Suppose that $M \cong U_n(n)$, that is, every element is a coloop. By induction, $T_M(x,y) = x^n$. Dually, if $M \cong U_0(n)$ (every element is a loop), then $T_M(x,y) = y^n$. ◀

Example 4.1.7. Let $M \cong U_1(2)$ (the graphic matroid of the "digon", two vertices joined by two parallel edges). Let $e \in E$; then

$$\begin{aligned} T_M &= T_{M\setminus e} + T_{M/e} \\ &= T(U_1(1)) + T(U_0(1)) = x + y. \end{aligned}$$

Next, let $M \cong U_2(3)$ (the graphic matroid of K_3, as well as the matroid associated with the geometric lattice $\Pi_3 \cong M_5$). Applying the Tutte recurrence for any $e \in E$ gives

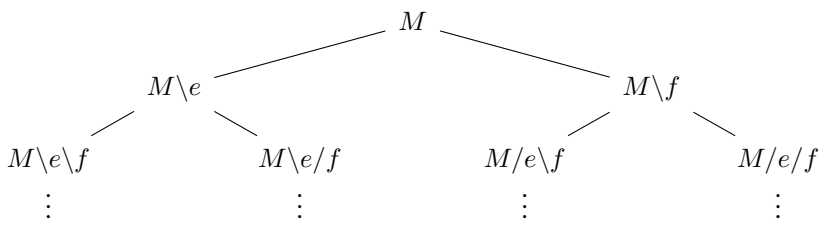
$$T(U_2(3)) = T(U_2(2)) + T(U_1(2)) = x^2 + x + y.$$

On the other hand,

$$T(U_1(3)) = T(U_1(2)) + T(U_0(2)) = x + y + y^2.$$

Note that these calculations agree with those of Example (4.1.3). ◀

The Tutte recurrence says that we can represent a calculation of T_M by a binary tree, with a branch for each deletion/contraction:



Example 4.1.8. Consider the diamond of Example 4.1.4. One possibility is to recurse on edge a (or equivalently on b, c, or d). When we delete a, the edge d becomes a coloop, and contracting it produces a copy of K_3. Therefore

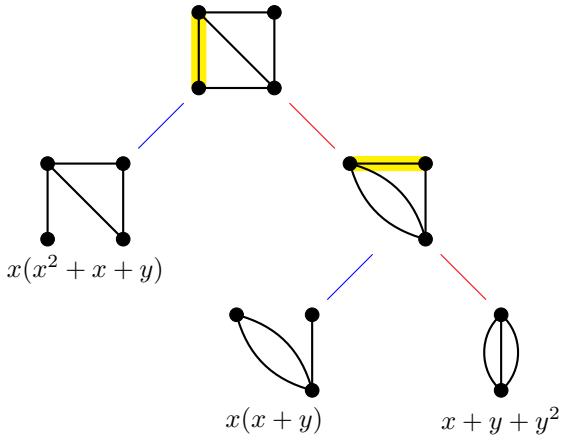
$$T(G\backslash a) \;=\; x(x^2+x+y)$$

by Example 4.1.7. Next, apply the Tutte recurrence to the edge b in G/a. The graph $G/a\backslash b$ has a coloop c, contracting which produces a digon. Meanwhile, $M(G/a/b) \cong U_1(3)$. Therefore

$$T(G/a\backslash b) \;=\; x(x+y) \quad \text{and} \quad T(G/a/b) \;=\; x+y+y^2.$$

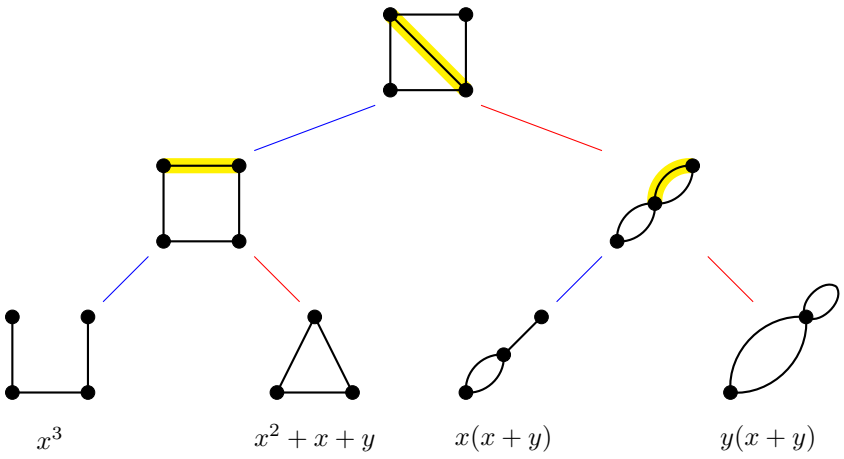
Putting it all together, we get

$$\begin{aligned} T(G) &= T(G\backslash a) + T(G/a) \\ &= T(G\backslash a) + T(G/a\backslash b) + T(G/a/b) \\ &= x(x^2+x+y) \;+\; x(x+y) \;+\; (x+y+y^2) \\ &= x^3+2x^2+2xy+x+y+y^2. \end{aligned}$$



On the other hand, we could have recursed first on e, getting

$$\begin{aligned} T(G) &= T(G\backslash e) + T(G/e) \\ &= T(G\backslash e\backslash c) + T(G\backslash e/c) + T(G/e\backslash c) + T(G/e/c) \\ &= x^3 \;+\; (x^2+x+y) \;+\; x(x+y) \;+\; y(x+y) \\ &= x^3+2x^2+2xy+x+y+y^2. \end{aligned}$$



◀

Proof of Theorem 4.1.5. Let M be a matroid on ground set E, let $e \in E$, and let r' and r'' be the rank functions of $M\backslash e$ and M/e respectively. The definitions of rank function, deletion, and contraction imply the following, for $A \subseteq E - e$:

1. If e is not a coloop, then $r'(A) = r_M(A)$.
2. If e is not a loop, then $r''(A) = r_M(A + e) - 1$.

To save space, set $X = x - 1$, $Y = y - 1$. We already know that if $E = \emptyset$, then $T_M = 1$.

For **(T2a)**, let e be a loop. Then

$$\begin{aligned} T_M &= \sum_{A\subseteq E} X^{r(E)-r(A)}Y^{|A|-r(A)} \\ &= \sum_{A\subseteq E:\ e\notin A} X^{r(E)-r(A)}Y^{|A|-r(A)} + \sum_{B\subseteq E:\ e\in B} X^{r(E)-r(B)}Y^{|B|-r(B)} \\ &= \sum_{A\subseteq E-e} X^{r'(E-e)-r'(A)}Y^{|A|-r'(A)} + \sum_{A\subseteq E-e} X^{r'(E-e)-r'(A)}Y^{|A|+1-r'(A)} \\ &= (1+Y)\sum_{A\subseteq E-e} X^{r'(E-e)-r'(A)}Y^{|A|-r'(A)} \quad = \quad yT_{M\backslash e}. \end{aligned}$$

For **(T2b)**, let e be a coloop. Then

$$\begin{aligned}
T_M &= \sum_{A\subseteq E} X^{r(E)-r(A)}Y^{|A|-r(A)} \\
&= \sum_{e\notin A\subseteq E} X^{r(E)-r(A)}Y^{|A|-r(A)} + \sum_{e\in B\subseteq E} X^{r(E)-r(B)}Y^{|B|-r(B)} \\
&= \sum_{A\subseteq E-e} X^{(r''(E-e)+1)-r''(A)}Y^{|A|-r''(A)} + \sum_{A\subseteq E-e} X^{(r''(E-e)+1)-(r''(A)+1)}Y^{|A|+1-(r''(A)+1)} \\
&= \sum_{A\subseteq E-e} X^{r''(E-e)+1-r''(A)}Y^{|A|-r''(A)} + \sum_{A\subseteq E-e} X^{r''(E-e)-r''(A)}Y^{|A|-r''(A)} \\
&= (X+1)\sum_{A\subseteq E-e} X^{r''(E-e)-r''(A)}Y^{|A|-r''(A)} \quad = \quad xT_{M/e}.
\end{aligned}$$

For **(T3)**, suppose that e is ordinary. Then

$$\begin{aligned}
T_M &= \sum_{A\subseteq E} X^{r(E)-r(A)}Y^{|A|-r(A)} \\
&= \sum_{A\subseteq E-e} \left[X^{r(E)-r(A)}Y^{|A|-r(A)}\right] + \left[X^{r(E)-r(A+e)}Y^{|A+e|-r(A+e)}\right] \\
&= \sum_{A\subseteq E-e} \left[X^{r'(E-e)-r'(A)}Y^{|A|-r'(A)}\right] + \left[X^{(r''(E-e)+1)-(r''(A)+1)}Y^{|A|+1-(r''(A)+1)}\right] \\
&= \sum_{A\subseteq E-e} X^{r'(E-e)-r'(A)}Y^{|A|-r'(A)} + \sum_{A\subseteq E-e} X^{r''(E-e)-r''(A)}Y^{|A|-r''(A)} \\
&= T_{M\setminus e} + T_{M/e}.
\end{aligned}$$

□

Some easy and useful observations (which illustrate, among other things, that both the rank-nullity and recursive forms are valuable tools):

1. The Tutte polynomial is multiplicative on direct sums, i.e., $T_{M_1\oplus M_2} = T_{M_1}T_{M_2}$. This is probably easier to see from the rank-nullity generating function than from the recurrence.
2. Duality interchanges x and y, i.e.,
$$T_M(x,y) = T_{M^*}(y,x). \tag{4.3}$$
The proof is left as an exercise (Problem 4.1). It can be deduced either from the Tutte recurrence (since duality interchanges deletion and contraction; see Prop. (3.8.2)) or from the corank-nullity generating function, by expressing r^* in terms of r (see Problem 3.8).
3. The Tutte recurrence implies that every coefficient of T_M is a *nonnegative* integer, a property which is not obvious from the closed formula (4.1).

4.2 Recipes

The Tutte polynomial is often referred to as "the universal deletion/contraction invariant for matroids": every invariant that satisfies a deletion/contraction-type recurrence can be recovered from the Tutte polynomial. This can be made completely explicit: the results in this section describe how to "reverse-engineer" a general deletion/contraction recurrence for a graph or matroid isomorphism invariant to express it in terms of the Tutte polynomial.

Theorem 4.2.1 (Tutte Recipe Theorem for Matroids)**.** *Let $u(M)$ be a matroid isomorphism invariant that satisfies a recurrence of the form*

$$u(M) = \begin{cases} 1 & \text{if } E = \emptyset, \\ Xu(M/e) & \text{if } e \in E \text{ is a coloop,} \\ Yu(M\backslash e) & \text{if } e \in E \text{ is a loop,} \\ au(M/e) + bu(M\backslash e) & \text{if } e \in E \text{ is ordinary} \end{cases}$$

where E denotes the ground set of M and X, Y, a, b are either indeterminates or numbers, with $a, b \neq 0$. Then

$$u(M) = a^{r(M)} b^{n(M)} T_M(X/a, Y/b).$$

Proof. Denote by $r(M)$ and $n(M)$ the rank and nullity of M, respectively. Note that

$$r(M) = r(M\backslash e) = r(M/e) + 1 \qquad \text{and} \qquad n(M) = n(M\backslash e) + 1 = n(M/e)$$

whenever deletion and contraction are well-defined. Define a new matroid invariant

$$\tilde{u}(M) = a^{-r(M)} b^{-n(M)} u(M)$$

and rewrite the recurrence in terms of $\tilde{u}$, abbreviating $r = r(M)$ and $n = n(M)$, to obtain

$$a^r b^n \tilde{u}(M) = \begin{cases} 1 & \text{if } E = \emptyset, \\ Xa^{r-1} b^n \tilde{u}(M/e) & \text{if } e \in E \text{ is a coloop,} \\ Ya^r b^{n-1} \tilde{u}(M\backslash e) & \text{if } e \in E \text{ is a loop,} \\ a^r b^n \tilde{u}(M/e) + a^r b^n \tilde{u}(M\backslash e) & \text{if } e \in E \text{ is ordinary.} \end{cases}$$

Setting $X = xa$ and $Y = yb$, we see that $\tilde{u}(M) = T_M(x, y) = T_M(X/a, Y/b)$ by Theorem 4.1.5, and rewriting in terms of $u(M)$ gives the desired formula. □

Bollobás [Bol98, p.340] gives an analogous result for graphs:

Theorem 4.2.2 (Tutte Recipe Theorem for Graphs)**.** *Let $u(G)$ be a graph isomorphism invariant that satisfies a recurrence of the form*

$$u(G) = \begin{cases} a^{|V|} & \text{if } E = \emptyset, \\ Xu(G\backslash e) & \text{if } e \in E \text{ is a coloop,} \\ Yu(G\backslash e) & \text{if } e \in E \text{ is a loop,} \\ bu(G\backslash e) + cu(G/e) & \text{if } e \in E \text{ is ordinary,} \end{cases}$$

where $G = (V, E)$ and X, Y, a, b, c are either indeterminates or numbers (with $b, c \neq 0$). Then

$$u(G) = a^{k(G)} b^{n(G)} c^{r(G)} T_G(aX/c, Y/b).$$

We omit the proof, which is similar to that of the previous result. A couple of minor complications are that many deletion/contraction graph invariants involve the numbers of vertices or components, which cannot be deduced from the matroid of a graph. Also, while deletion and contraction of a cut-edge of a graph produce two isomorphic *matroids*, they do not produce two isomorphic *graphs* (so, no, that's not a misprint in the coloop case of Theorem 4.2.2). The invariant U is described by Bollobás as "the universal form of the Tutte polynomial."

4.3 Basis activities

We know that $T_M(x,y)$ has nonnegative integer coefficients and that $T_M(1,1)$ is the number of bases of M. These observations suggest that we should be able to interpret the Tutte polynomial as a generating function for bases: that is, there should be combinatorially defined functions $i, e : \mathscr{B}(M) \to \mathbb{N}$ such that

$$T_M(x,y) = \sum_{B \in \mathscr{B}(M)} x^{i(B)} y^{e(B)}.$$

In fact, this is the case. The tricky part is that $i(B)$ and $e(B)$ must be defined with respect to a total order $e_1 < \cdots < e_n$ on the ground set E, so they are not really invariants of B itself. However, another miracle occurs: the Tutte polynomial itself is independent of the choice of total order.

Definition 4.3.1. Let M be a matroid on E with basis system $\mathscr{B}$ and let $B \in \mathscr{B}$. For $e \in B$, the **fundamental cocircuit** of e with respect to B, denoted $C^*(e,B)$, is the unique cocircuit in $(E \setminus B) + e$. That is,

$$C^*(e,B) = \{e' \colon B - e + e' \in \mathscr{B}\}.$$

Dually, for $e \notin B$, then the **fundamental circuit** of e with respect to B, denoted $C(e,B)$, is the unique circuit in $B + e$. That is,

$$C(e,B) = \{e' \colon B + e - e' \in \mathscr{B}\}.$$

In other words, the fundamental cocircuit consists of e together with all elements outside B that can replace e in a basis exchange, while the fundamental circuit consists of e together with all elements outside B that can be replaced by e.

Suppose that $M = M(G)$, where G its a connected graph, and B is a spanning tree. For all $e \in B$, the graph $B - e$ has two components, say X and Y, and $C^*(e,B)$ is the set of all edges with one endpoint in each of X and Y. Dually, if $e \notin B$, then $B + e$ has exactly one cycle, and that cycle is $C(e,B)$.

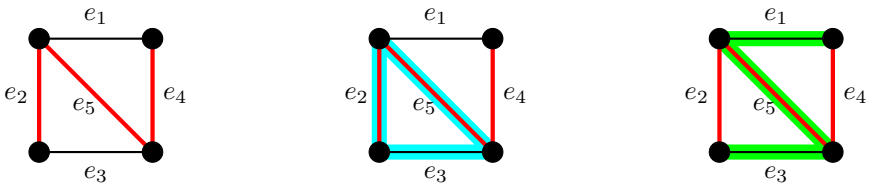
If M is a vector matroid, then $C^*(e,B)$ consists of all vectors not in the codimension-1 subspace spanned by $B - e$, and $C(e,B)$ is the unique linearly dependent subset of $B + e$.

Definition 4.3.2. Let M be a matroid on a totally ordered vertex set $E = \{e_1 < \cdots < e_n\}$, and let B be a basis of M. An element $e \in B$ is **internally active** with respect to B if e is the minimal element of $C^*(e,B)$. An element $e \notin B$ is **externally active** with respect to B if e is the minimal element of $C(e,B)$. We set

$$\begin{aligned} i(B) &= \#\{e \in B \colon e \text{ is internally active with respect to } B\} \\ &= \#\{\text{edges of } B \text{ that cannot be replaced by anything smaller outside } B\}, \\ e(B) &= \#\{e \in E \setminus B \colon e \text{ is externally active with respect to } B\} \\ &= \#\{\text{edges of } E \setminus B \text{ that cannot replaced anything smaller inside } B\}. \end{aligned}$$

Note that these numbers depend on the choice of ordering of E.

Example 4.3.3. Let G be the graph with edges labeled as shown below, and let B be the spanning tree $\{e_2, e_4, e_5\}$ shown in red. The middle figure shows $C(e_1,B)$, and the right-hand figure shows $C^*(e_5,B)$.



Here are some fundamental circuits and cocircuits:

$$\begin{aligned} C(e_1, B) &= \{e_1, e_4, e_5\} && \text{so } e_1 \text{ is externally active;} \\ C(e_3, B) &= \{e_2, e_3, e_5\} && \text{so } e_3 \text{ is not externally active;} \\ C^*(e_2, B) &= \{e_2, e_3\} && \text{so } e_2 \text{ is internally active;} \\ C^*(e_4, B) &= \{e_1, e_4\} && \text{so } e_4 \text{ is not internally active;} \\ C^*(e_5, B) &= \{e_1, e_3, e_5\} && \text{so } e_5 \text{ is not internally active.} \end{aligned}$$

◀

Theorem 4.3.4 (Tutte). *Let M be a matroid on E. Fix a total ordering of E and let $e(B)$ and $i(B)$ denote respectively the number of externally active and internally active elements with respect to B. Then*

$$T_M(x, y) \;=\; \sum_{B \in \mathscr{B}(M)} x^{i(B)} y^{e(B)}. \tag{4.4}$$

For instance, in Example 4.3.3, the spanning tree B contributes the monomial $xy = x^1 y^1$ to $T(G;\, x, y)$.

Tutte's original paper [Tut54] actually defined the Tutte polynomial (which he called the "dichromate") as $\sum_{B \in \mathscr{B}(M)} x^{i(B)} y^{e(B)}$ (rather than the corank-nullity generating function), then proved it that obeys the deletion/contraction recurrence. Like the proof of Theorem 4.1.5, this result requires careful bookkeeping but is not conceptually difficult. Note in particular that if e is a loop (resp. coloop), then $e \notin B$ (resp. $e \in B$) for every basis B, and $C(e, B) = \{e\}$ (resp. $C^*(e, B) = \{e\}$), so e is externally (resp. internally) active with respect to B, so the generating function (4.4) is divisible by y (resp. x).

4.4 The characteristic and chromatic polynomials

We first show that the characteristic polynomial of a geometric lattice is a specialization of the Tutte polynomial of the corresponding matroid.

Theorem 4.4.1. *Let M be a simple matroid on E with rank function r and lattice of flats $!L$. Then*

$$\chi(L; k) \;=\; (-1)^{r(M)} T_M(1 - k, 0).$$

Proof. Let $A \mapsto \bar{A}$ be the matroid closure operator of M. Observe that

$$\begin{aligned} (-1)^{r(M)} T_M(1-k, 0) \;&=\; (-1)^{r(M)} \sum_{A \subseteq E} (-k)^{r(M) - r(A)} (-1)^{|A| - r(A)} \\ &=\; \sum_{A \subseteq E} (-1)^{|A|} k^{r(M) - r(A)} \\ &=\; \sum_{K \in L} \underbrace{\left(\sum_{A \subseteq E:\ \bar{A} = K} (-1)^{|A|} \right)}_{f(K)} k^{r(M) - r(K)}. \end{aligned}$$

We now claim that $f(K) = \mu_L(\hat{\mathbf{0}}, K)$. For each flat $K \in L$, let

$$g(K) \;=\; \sum_{J \in L:\ J \subseteq K} f(J)$$

so that by Möbius inversion (2.3a)

$$f(K) \;=\; \sum_{J\in L:\ J\subseteq K} \mu(J,K)g(J). \tag{4.5}$$

But on the other hand

$$g(J) \;=\; \sum_{I\in L:\ I\subseteq J}\ \sum_{A\subseteq E:\ \bar{A}=I} (-1)^{|A|} \;=\; \sum_{A\subseteq E:\ \bar{A}\subseteq J} (-1)^{|A|} \;=\; \sum_{A\subseteq J} (-1)^{|A|} = (1-1)^{|J|} = \delta_{J,\emptyset}.$$

So (4.5) simplifies to $f(K) = \mu(\hat{\mathbf{0}}, K)$, completing the proof. □

Theorem 4.4.1 gives another proof that the Möbius function of a semimodular lattice L weakly alternates in sign, or specifically that $(-1)^{r(L)}\mu(L) \geq 0$ (Theorem 2.4.7). First, if L is not geometric, or equivalently not atomic, then by Corollary 2.4.10 $\mu(L) = 0$. Second, if L is geometric, then by (2.7) and Theorem 4.4.1

$$(-1)^{r(L)}\mu(L) = (-1)^{r(L)}\chi(L;0) = T_M(1,0). \tag{4.6}$$

But $T_M(1,0) \geq 0$ for every matroid M, because $T_M \in \mathbb{N}[x,y]$.

The characteristic polynomial of a graphic matroid has a classical combinatorial interpretation in terms of colorings. Let $G = (V,E)$ be a connected graph. Recall that a ***k*-coloring** of G is a function $f : V \to [k]$, and a coloring is **proper** if $f(v) \neq f(w)$ whenever vertices v and w are adjacent. We showed in Example 2.3.5 that the function

$$p_G(k) = \text{number of proper } k\text{-colorings of } G$$

is a polynomial in k, called the **chromatic polynomial** of G. In fact $p_G(k) = k \cdot \chi_{K(G)}(k)$. We can also prove this fact via deletion/contraction.

First, note some important special cases:

- If G has a loop, then its endpoints automatically have the same color, so it's impossible to color G properly and $p_G(k) = 0$.
- If $G = K_n$, then all vertices must have different colors. There are k choices for $f(1)$, $k-1$ choices for $f(2)$, etc., so $p_{K_n}(k) = k(k-1)(k-2)\cdots(k-n+1)$.
- At the other extreme, the graph $G = \overline{K_n}$ with n vertices and no edges has chromatic polynomial k^n, since every coloring is proper.
- If T is a tree with n vertices, then pick any vertex as the root; this imposes a partial order on the vertices in which the root is $\hat{\mathbf{1}}$ and each non-root vertex v is covered by exactly one other vertex $p(v)$ (its "parent"). There are k choices for the color of the root, and once we know $f(p(v))$ there are $k-1$ choices for $f(v)$. Therefore $p_T(k) = k(k-1)^{n-1}$.
- If G has connected components $G_1, \ldots, G_s$, then $p_G(k) = \prod_{i=1}^{s} p_{G_i}(k)$. Equivalently, $p_{G+H}(k) = p_G(k)p_H(k)$, where $+$ denotes disjoint union of graphs.

Theorem 4.4.2. *For every graph G*

$$p_G(k) \;=\; (-1)^{n-c}k^c \cdot T_G(1-k,0)$$

where n is the number of vertices of G and c is the number of components. In particular, $p_G(k)$ is a polynomial function of k.

Proof. First, we show that the chromatic function satisfies the recurrence

$$p_G(k) = k^n \qquad \text{if } E = \emptyset; \tag{4.7}$$
$$p_G(k) = 0 \qquad \text{if } G \text{ has a loop;} \tag{4.8}$$
$$p_G(k) = (k-1)p_{G/e}(k) \qquad \text{if } e \text{ is a coloop;} \tag{4.9}$$
$$p_G(k) = p_{G\setminus e}(k) - p_{G/e}(k) \qquad \text{otherwise.} \tag{4.10}$$

We already know (4.7) and (4.8). Suppose $e = xy$ is not a loop. Let f be a proper k-coloring of $G \setminus e$. If $f(x) = f(y)$, then we can identify x and y to obtain a proper k-coloring of G/e. If $f(x) \neq f(y)$, then f is a proper k-coloring of G. So (4.10) follows.

This argument applies even if e is a coloop. In that case, however, the component H of G containing e becomes two components H' and H'' of $G \setminus e$, whose colorings can be chosen independently of each other. So the probability that $f(x) = f(y)$ in any proper coloring is $1/k$, implying (4.9).

The graph $G \setminus e$ has n vertices and either $c+1$ or c components, according as e is or is not a coloop. Meanwhile, G/e has $n-1$ vertices and c components. By induction,

$$(-1)^{n-c}k^c T_G(1-k,0) = \begin{cases} k^n & \text{if } E = \emptyset, \\ 0 & \text{if } e \text{ is a loop,} \\ (1-k)(-1)^{n+1-c}\,k^c\,T_{G/e}(1-k,0) & \text{if } e \text{ is a coloop,} \\ (-1)^{n-c}k^c\left(T_{G\setminus e}(1-k,0) + T_{G/e}(1-k,0)\right) & \text{otherwise} \end{cases}$$
$$= \begin{cases} k^n & \text{if } E = \emptyset, \\ 0 & \text{if } e \text{ is a loop,} \\ (k-1)p_{G/e}(k) & \text{if } e \text{ is a coloop,} \\ p_{G\setminus e}(k) - p_{G/e}(k) & \text{otherwise} \end{cases}$$

which is exactly the recurrence satisfied by the chromatic polynomial. □

Remark 4.4.3. It is also possible to prove Theorem 4.4.2 by invoking the Tutte Recipe Theorem for Graphs (Theorem 4.2.2). To do so, one would need to replace case (4.9) of the chromatic recurrence with the statement $p_G(k) = \frac{k-1}{k}p_{G-e}(k)$.

More generally, if G is a graph with n vertices and c components, then its graphic matroid $M = M(G)$ has rank $n-c$, whose associated geometric lattice is the connectivity lattice $K(G)$. Combining Theorems 4.4.1 and 4.4.2 gives

$$p_G(k) = k^c\,\chi(K(G);k).$$

4.5 Acyclic orientations

An **orientation** $\mathcal{O}$ of a graph $G = (V, E)$ is an assignment of a direction to each edge $xy \in E$ (either $\overline{xy}$ or $\overline{yx}$). A **directed cycle** is a sequence $(x_0, x_1, \ldots, x_{n-1})$ of vertices such that $\overline{x_i x_{i+1}}$ is a directed edge for every i. (Here the indices are taken modulo n.)

An orientation is **acyclic** if it has no directed cycles. Let $A(G)$ be the set of acyclic orientations of G, and let $a(G) = |A(G)|$.

For example:

1. If G has a loop then $a(G) = 0$.
2. If G has no loops, then every total order on the vertices gives rise to an acyclic orientation: orient each edge from smaller to larger vertex. Of course, different total orders can produce the same a.o.
3. If G has no edges than $a(G) = 1$. Otherwise, $a(G)$ is even, since reversing all edges is a fixed-point free involution on $A(G)$.
4. Removing parallel copies of an edge does not change $a(G)$, since all parallel copies would have to be oriented in the same direction to avoid any 2-cycles.
5. If G is a forest then every orientation is acyclic, so $a(G) = 2^{|E(G)|}$.
6. If $G = K_n$ then the acyclic orientations are in bijection with the total orderings, so $a(G) = n!$.
7. If $G = C_n$ (the cycle of graph of length n) then it has 2^n orientations, of which exactly two are not acyclic, so $a(C_n) = 2^n - 2$.

Colorings and orientations are intimately connected. Given a proper coloring $f : V(G) \to [k]$, one can naturally define an acyclic orientation by directing each edge from the smaller to the larger color. (So #2 in the above list is a special case of this.) The connection between them is the prototypical example of what is called *combinatorial reciprocity*.

A **(compatible) k-pair** for a graph $G = (V, E)$ is a pair $(\mathcal{O}, f)$, where $\mathcal{O}$ is an acyclic orientation of G and $f : V \to [k]$ is a coloring such that $f(x) \leq f(y)$ for every directed edge $x \to y$ in D. Let $\mathsf{C}(G) = \mathsf{C}(G, k)$ be the set of compatible k-pairs of G (we can safely drop k from the notation)

Theorem 4.5.1 (Stanley's Acyclic Orientation Theorem)**.** *For every graph G and positive integer k,*

$$|\mathsf{C}(G, k)| = (-1)^n p_G(-k) = k^c T_G(1 + k, 0). \tag{4.11}$$

Proof. The second equality follows from Theorem 4.4.2, so we prove the first one. Let $n = |G|$.

If G has no edges then $|\mathsf{C}(G)| = k^n = (-1)^n(-k)^n = (-1)^n p_G(-k)$, confirming (4.11).

If G has a loop then it has no acyclic orientations, hence no k-pairs for any k, so both sides of (4.11) are zero.

Let $e = xy$ be an edge of G that is not a loop. Denote the left-hand side of (4.11) by $\bar{p}_G(k)$. Then

$$\begin{aligned}\bar{p}_G(k) = (-1)^n p_G(-k) &= (-1)^n (p_{G\setminus e}(-k) - p_{G/e}(-k)) \\ &= (-1)^n ((-1)^n \bar{p}_{G\setminus e}(k) - (-1)^{n-1} \bar{p}_{G/e}(k)) \\ &= \bar{p}_{G\setminus e}(k) + \bar{p}_{G/e}(k)\end{aligned}$$

so we need to show that $|\mathsf{C}(G)|$ satisfies the same recurrence. Write

Say that a pair $(\mathcal{O}, f) \in \mathsf{C}(G)$ is **reversible** (with respect to e) if reversing e produces a compatible pair $(\mathcal{O}', f)$; otherwise it is **irreversible**. (Reversibility is equivalent to saying that $f(x) = f(y)$ and that G does not contain a directed path from either endpoint of e to the other.) Let $\mathsf{C}_{\text{rev}}(G)$ and $\mathsf{C}_{\text{irr}}(G)$ denote the sets of reversible and irreversible compatible pairs, respectively.

If e is reversible, then contracting it to a vertex z and defining $f(z) - f(x) - f(y)$ produces a compatible pair of G/e. (The resulting orientation is acyclic because any directed cycle lifts to either a directed cycle in G, or an oriented path between the endpoints of e, neither of which exists.) This defines a map $\psi : \mathsf{C}_{\text{rev}}(G) \to \mathsf{C}(G/e)$, which is 2-to-1 because $\psi(\mathcal{O}, f) = \psi(\mathcal{O}', f)$. Moreover, ψ is onto: any $(\mathcal{O}, f) \in \mathsf{C}(G/e)$ can be lifted to $(\tilde{\mathcal{O}}\tilde{f}) \in \mathsf{C}(G)$ by defining $\tilde{f}(x) = \tilde{f}(y) = f(z)$ and orienting e in either direction (the acyclicity of $\mathcal{O}$ means that there is no oriented path from either x or y to the other in $\tilde{\mathcal{O}}$). We conclude that

$$|\mathsf{C}(G/e)| = \frac{|\mathsf{C}_{\text{rev}}(G)|}{2}. \tag{4.12}$$

There is a map $\omega : \mathsf{C}(G) \to \mathsf{C}(G-e)$ given by deleting e. I claim that ω is surjective, which is equivalent to saying that it is always possible to extend any element of $\mathsf{C}(G-e)$ to $\mathsf{C}(G)$ by choosing an appropriate orientation for e. (If $f(x) < f(y)$, then $\mathcal{O}$ has no y,x-path by compatibility. If $f(x) = f(y)$ and neither orientation of e is acyclic, then $\mathcal{O}$ must contain a directed path from each of x, y to the other, hence is not acyclic.) The map ω is 1-to-1 on $\mathsf{C}_{\text{irr}}(G)$ but 2-to-1 on $\mathsf{C}_{\text{rev}}(G)$ (for the same reason as ψ). Therefore,

$$|\mathsf{C}(G-e)| = |\mathsf{C}_{\text{irr}}(G)| + \frac{|\mathsf{C}_{\text{rev}}(G)|}{2}. \tag{4.13}$$

Combining (4.12) and (4.13) gives

$$|\mathsf{C}(G-e)| + |\mathsf{C}(G/e)| = |\mathsf{C}_{\text{irr}}(G)| + |\mathsf{C}_{\text{rev}}(G)| = |\mathsf{C}(G)|$$

as desired. □

In particular, if $k = 1$ then there is only one choice for f and every acyclic orientation is compatible with it, which produces the following striking corollary (often referred to as "Stanley's theorem on acyclic orientations," although Stanley himself prefers that name for the more general Theorem 4.5.1).

Theorem 4.5.2. *The number of acyclic orientations of G is $|p_G(-1)| = T_G(2,0)$.*

Combinatorial reciprocity can be viewed geometrically. For more detail, look ahead to Section 5.5 and/or see a source such as Beck and Robins [BR15], but here is a brief taste.

Let G be a simple graph on n vertices. The **graphic arrangement** $\mathcal{A}_G$ is the union of all hyperplanes in $\mathbb{R}^n$ defined by the equations $x_i = x_j$ where ij is an edge of G. The complement $\mathbb{R}^n \setminus \mathcal{A}_G$ consists of finitely many disjoint open polyhedra (the "regions" of the arrangement), each of which is defined by a set of inequalities, including either $x_i < x_j$ or $x_i > x_j$ for each edge. Thus each region naturally gives rise to an orientation of G, and it is not hard to see that the regions are in fact in bijection with the acyclic orientations. Meanwhile, a k-coloring of G can be regarded as an integer point in the cube $[1,k]^n \subseteq \mathbb{R}^n$, and a proper coloring corresponds to a point that does not lie on any hyperplane in $\mathcal{A}_G$. In this setting, Stanley's theorem is an instance of something more general called *Ehrhart reciprocity* (which I will add notes on at some point).

4.6 The Tutte polynomial and linear codes

Definition 4.6.1. A **linear code** $\mathscr{C}$ is a subspace of $(\mathbb{F}_q)^n$, where q is a prime power and $\mathbb{F}_q$ is the field of order q. The number n is the **length** of $\mathscr{C}$. The elements $c = (c_1, \ldots, c_n) \in \mathscr{C}$ are called **codewords**. The **support** of a codeword is $\operatorname{supp}(c) = \{i \in [n] \colon c_i \neq 0\}$, and its **weight** is $\operatorname{wt}(c) = |\operatorname{supp}(c)|$. The **weight enumerator** of $\mathscr{C}$ is the polynomial

$$W_{\mathscr{C}}(t) = \sum_{c \in \mathscr{C}} t^{\operatorname{wt}(c)}.$$

For example, let $\mathscr{C}$ be the subspace of $\mathbb{F}_2^3$ generated by the rows of the matrix

$$X = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \in (\mathbb{F}_2)^{3\times 2}.$$

So $\mathscr{C} = \{000, 101, 011, 110\}$, and $W_{\mathscr{C}}(t) = 1 + 3t^2$.

The **dual code** $\mathscr{C}^\perp$ is the orthogonal complement under the standard inner product. This inner product is nondegenerate, i.e., $\dim \mathscr{C}^\perp = n - \dim \mathscr{C}$. (Note, though, that a subspace and its orthogonal complement can intersect nontrivially. A space can even be its own orthogonal complement, such as $\{00, 11\} \subseteq \mathbb{F}_2^2$. This does not happen over $\mathbb{R}$, where the inner product is not only nondegenerate but also positive-definite, but "positive" does not make sense over a finite field.) In this case, $\mathscr{C}^\perp = \{000, 111\}$ and $W_{\mathscr{C}^\perp}(t) = 1 + t^3$.

Theorem 4.6.2 (Curtis Greene, 1976). *Let $\mathscr{C}$ be a linear code of length n and dimension r over $\mathbb{F}_q$, and let M be the matroid represented by the columns of a matrix X whose rows are a basis for $\mathscr{C}$. Then*

$$W_{\mathscr{C}}(t) = t^{n-r}(1-t)^r\, T_M\left(\frac{1+(q-1)t}{1-t},\ \frac{1}{t}\right)$$

The proof is a deletion-contraction argument. As an example, if $\mathscr{C} = \{000, 101, 011, 110\} \subseteq \mathbb{F}_2^3$ as above, then the matroid M is $U_2(3)$. Its Tutte polynomial is $x^2 + x + y$, and Greene's theorem gives $W_{\mathscr{C}}(t) = t(1-t)^2\, T_M\left(\frac{1+t}{1-t},\ \frac{1}{t}\right) = 1 + 3t^2$ as noted above (calculation omitted).

If $X^\perp$ is a matrix whose rows are a basis for the dual code, then the corresponding matroid $M^\perp$ is precisely the dual matroid to M. We know that $T_M(x, y) = T_{M^\perp}(y, x)$ by (4.3), so setting $s = (1-t)/(1+(q-1)t)$ (so $t = (1-s)/(1+(q-1)s)$; isn't that convenient?) gives

$$\begin{aligned} W_{\mathscr{C}^\perp}(t) &= t^r(1-t)^{n-r}\, T_M\left(\frac{1+(q-1)s}{1-s},\ \frac{1}{s}\right) \\ &= t^r(1-t)^{n-r}s^{r-n}(1-s)^{-r}W_{\mathscr{C}}(s), \end{aligned}$$

or rewriting in terms of t,

$$W_{\mathscr{C}^\perp}(t) = \frac{1+(q-1)t^n}{q^r} W_{\mathscr{C}}\left(\frac{1-t}{1+(q-1)t}\right)$$

which is known as the *MacWilliams identity* and is important in coding theory.

4.7 Exercises

Problem 4.1. Give two proofs of the fact that $T_{M^*}(x, y) = T_M(y, x)$, following the hints following equation (4.3).

Problem 4.2. An orientation of a graph is called **totally cyclic** if every edge belongs to a directed cycle. Prove that the number of totally cyclic orientations of G is $T_G(0, 2)$.

Problem 4.3. Let G be a finite graph with n vertices, r edges, and k components. Fix an orientation $\mathcal{O}$ on $E(G)$. Let $I(v)$ (resp., $O(v)$) denote the set of edges entering (resp., leaving) each vertex v. Let q be a positive integer and $\mathbb{Z}_q = \mathbb{Z}/q\mathbb{Z}$. A **nowhere-zero q-flow** (or q-NZF) on G (with respect to $\mathcal{O}$) is a function $\phi : E(G) \to \mathbb{Z}_q \setminus \{0\}$ satisfying the **conservation law**

$$\sum_{e \in I(v)} \phi(e) = \sum_{e \in O(v)} \phi(e)$$

for every $v \in V(G)$. Let $F_G^{\mathcal{O}}(q)$ denote the set of nowhere-zero q-flows and $f_G^{\mathcal{O}}(q) = |F_G^{\mathcal{O}}(q)|$.

(i) Prove that $f_G^{\mathcal{O}}(q)$ depends only on the graph G, not on the choice of orientation (so we are justified in writing $f_G(q)$).

(ii) Prove that $f_G(q) = (-1)^{r-n+k} T_G(0, 1-q)$.

(Interestingly, $\mathbb{Z}/q\mathbb{Z}$ can be replaced with any abelian group of cardinality q without affecting the result.)

Problem 4.4. Let $G = (V, E)$ be a graph with n vertices and c components. For a vertex coloring $f : V \to \mathbb{P}$, let $i(f)$ denote the number of "improper" edges, i.e., whose endpoints are assigned the same color. The *(Crapo) coboundary polynomial* of G is

$$\bar{\chi}_G(q;t) \;=\; q^{-c} \sum_{f:V\to[q]} t^{i(f)}.$$

This is evidently a stronger invariant than the chromatic polynomial of G, which can be obtained as $q\bar{\chi}_G(q,0)$. In fact, the coboundary polynomial provides the same information as the Tutte polynomial. Prove that

$$\bar{\chi}_G(q;t) \;=\; (t-1)^{n-c} T_G\left(\frac{q+t-1}{t-1},\, t\right)$$

by finding a deletion/contraction recurrence for the coboundary polynomial.

Problem 4.5. Let M be a matroid on E and let $0 \le p \le 1$. The **reliability polynomial** $R_M(p)$ is the probability that the rank of M stays the same when each ground set element is independently retained with probability p and deleted with probability $1-p$. (In other words, we have a family of i.i.d. random variables $\{X_e : e \in E\}$, each of which is 1 with probability p and 0 with probability $1-p$. Let $A = \{e \in E : X_e = 1\}$. Then $R_M(p)$ is the probability that $r(A) = r(E)$.) Give a formula for $R_M(p)$ in terms of the Tutte polynomial, using

(a) the definition of the Tutte polynomial as the corank/nullity generating function;
(b) the Tutte Recipe Theorem.

Problem 4.6. Prove Merino's theorem on critical configurations of the chip-firing game. (This needs details!)

Problem 4.7. Prove Theorem 4.3.4.

Problem 4.8. Prove Theorem 4.6.2.

Much, much more about the Tutte polynomial can be found in [BO92], the MR review of which begins, "The reviewer, having once worked on that polynomial himself, is awed by this exposition of its present importance in combinatorial theory." (The reviewer was one W.T. Tutte.)

Chapter 5

Hyperplane Arrangements

An excellent source for the combinatorial theory of hyperplane arrangements is Stanley's book chapter [Sta07], which is accessible to newcomers, and includes a self-contained treatment of topics such as the Möbius function and characteristic polynomial. Another canonical (but harder) source is the monograph by Orlik and Terao [OT92].

5.1 Basic definitions

Definition 5.1.1. Let $\mathbb{k}$ be a field, typically either $\mathbb{R}$ or $\mathbb{C}$, and let $n \geq 1$. A **linear hyperplane** in $\mathbb{k}^n$ is a vector subspace of codimension 1. An **affine hyperplane** is a translate of a linear hyperplane. A **hyperplane arrangement** $\mathcal{A} \subseteq \mathbb{k}^n$ is a finite set of (distinct) hyperplanes $H_1, \ldots, H_k \subseteq \mathbb{k}^n$. The number n is called the **dimension** of $\mathcal{A}$, and the space $\mathbb{k}^n$ is its **ambient space**. The **intersection poset** $L(\mathcal{A})$ is the poset of all nonempty intersections of subsets of $\mathcal{A}$, ordered by reverse inclusion. If $\mathcal{B} \subseteq \mathcal{A}$ is a subset of hyperplanes, we write $\cap\mathcal{B}$ for $\bigcap_{H \in \mathcal{B}} H$. The **characteristic polynomial** of $\mathcal{A}$ is

$$\chi_{\mathcal{A}}(t) = \sum_{x \in L(\mathcal{A})} \mu(\hat{\mathbf{0}}, x) t^{\dim x}. \tag{5.1}$$

This is essentially the same as the characteristic polynomial of the poset $L(\mathcal{A})$, up to a correction factor that we will explain soon.

Example 5.1.2. Two line arrangements in $\mathbb{R}^2$ are shown in Figure 5.1. The arrangement $\mathcal{A}_1$ consists of the lines $x = 0$, $y = 0$, and $x = y$. The arrangement $\mathcal{A}_2$ consists of the four lines $\ell_1, \ell_2, \ell_3, \ell_4$ given by the equations $y = 1$, $x = y$, $x = -y$, $y = -1$ respectively. The intersection posets $L(\mathcal{A}_1)$ and $L(\mathcal{A}_2)$ are shown in Figure 5.2; the characteristic polynomials are $t^2 - 3t + 2$ and $t^2 - 4t + 5$ respectively. ◀

Example 5.1.3. The **Boolean arrangement** Bool_n (or **coordinate arrangement**) consists of the n coordinate hyperplanes in n-space. Its intersection poset is the Boolean lattice Bool_n. (I make no apologies for abusing notation by referring to the arrangement and the poset with the same symbol.) More generally, any arrangement whose intersection poset is Boolean might be referred to as a Boolean arrangement. ◀

Example 5.1.4. The **braid arrangement** Br_n consists of the $\binom{n}{2}$ hyperplanes $x_i = x_j$ in $\mathbb{R}^n$. Its intersection poset is naturally identified with the partition lattice Π_n. This is simply because any set of equalities among $x_1, \ldots, x_n$ defines an equivalence relation on $[n]$, and certainly every equivalence relation can be obtained

Figure 5.1: Two line arrangements in $\mathbb{R}^2$.

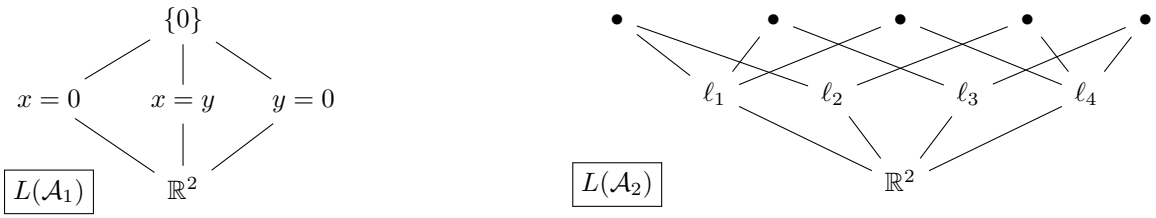


Figure 5.2: Intersection posets of the arrangements in Fig. 5.1.

in this way. For instance, the intersection poset of Br_3 and the isomorphism to Π_3 is shown in Figure 5.3. One can check that Br_3 has characteristic polynomial $t^3 - 3t^2 + 2t$.

◀

Figure 5.4 shows some hyperplane arrangements in $\mathbb{R}^3$. Note that every hyperplane in Br_n contains the line

$$x_1 = x_2 = \cdots = x_n,$$

so projecting $\mathbb{R}^4$ along that line allows us to picture Br_4 as an arrangement $\mathrm{ess}(\mathsf{Br}_4)$ in $\mathbb{R}^3$. (The symbol "ess" means *essentialization*, to be defined precisely soon.) The second two figures were produced using the computer algebra system `Sage` [S+14].

The poset $L(\mathcal{A})$ is the fundamental combinatorial invariant of $\mathcal{A}$. Some easy observations:

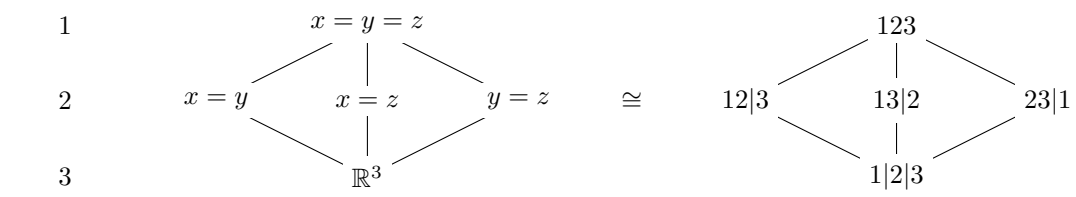


Figure 5.3: The isomorphism $L(\mathsf{Br}_3) \cong \Pi_3$.

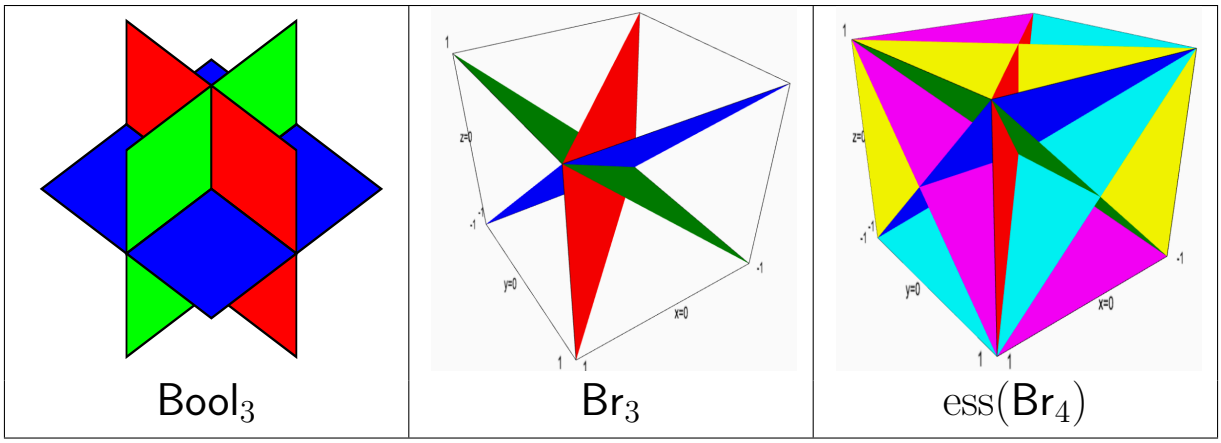


Figure 5.4: Three hyperplane arrangements in $\mathbb{R}^3$.

1. If $T : \mathbb{R}^n \to \mathbb{R}^n$ is an invertible linear transformation, then $L(T(\mathcal{A})) \cong L(\mathcal{A})$, where $T(\mathcal{A}) = \{T(H) \colon H \in \mathcal{A}\}$. In fact, the intersection poset is invariant under any *affine* transformation. (The group of affine transformations is generated by the invertible linear transformations together with translations.)

2. The poset $L(\mathcal{A})$ is a meet-semilattice, with meet given by $\cap\mathcal{B} \ \wedge \ \cap\mathcal{C} = \cap(\mathcal{B} \cap \mathcal{C})$ for all $\mathcal{B}, \mathcal{C} \subseteq \mathcal{A}$. Its $\hat{\mathbf{0}}$ element is $\cap\emptyset$, which by convention is $\Bbbk^n$.

3. $L(\mathcal{A})$ is ranked, with rank function $r(X) = n - \dim X$. To see this, observe that each covering relation $X \lessdot Y$ comes from intersecting an affine linear subspace X with a hyperplane H that neither contains nor is disjoint from X, so that $\dim(X \cap H) = \dim X - 1$.

4. $L(\mathcal{A})$ has a $\hat{\mathbf{1}}$ element if and only if the **center** $\cap\mathcal{A}$ is nonempty. Such an arrangement is called **central**. In this case $L(\mathcal{A})$ is a lattice (and may be referred to as the **intersection lattice** of $\mathcal{A}$). Since translation does not affect whether an arrangement is central (or indeed any of its combinatorial structure), we will typically assume that $\cap\mathcal{A}$ contains the zero vector, which is to say that every hyperplane in $\mathcal{A}$ is a *linear* hyperplane in $\Bbbk^n$. (So an arrangement is central if and only if it is a translation of an arrangement of linear hyperplanes.)

5. When $\mathcal{A}$ is central, the lattice $L(\mathcal{A})$ is geometric. It is atomic by definition, and it is submodular because it is a sublattice of the chain-finite modular lattice $L(\Bbbk^n)^*$ (the lattice of all subspaces of $\Bbbk^n$ ordered by reverse inclusion). The associated matroid $M(\mathcal{A}) = M(L(\mathcal{A}))$ is represented over $\Bbbk$ by any family of vectors $\{\mathbf{n}_H \colon H \in \mathbb{A}\}$ where $\mathbf{n}_H$ is normal to H. (That is, $H^\perp = \Bbbk\langle \mathbf{n}_H \rangle$ with respect to some fixed non-degenerate bilinear form on $\Bbbk^n$.) Any normals will do, since the matroid is unchanged by scaling the $\mathbf{n}_H$ independently.

Therefore, all of the tools we have developed for looking at posets, lattices and matroids can be applied to study hyperplane arrangements.

The dimension of an arrangement is not a combinatorial invariant; that is, it cannot be extracted from the intersection poset. If Br_4 were a "genuine" 4-dimensional arrangement then we would not be able to represent it in $\mathbb{R}^3$. However, we can do so because the center of Br_4 has positive dimension, so squashing the center to a point reduces the ambient dimension without changing the intersection poset. This observation motivates the following definition.

Definition 5.1.5. Let $\mathcal{A} \subseteq \Bbbk^n$ be an arrangement and let $N(\mathcal{A}) = \Bbbk\langle \mathbf{n}_H \colon H \in \mathcal{A} \rangle$, where $\mathbf{n}_H$ is normal to H. The **essentialization** of $\mathcal{A}$ is the arrangement

$$\operatorname{ess}(\mathcal{A}) = \{H \cap N(\mathcal{A}) \colon H \in \mathcal{A}\} \ \subseteq \ N(\mathcal{A}).$$

We say that $\mathcal{A}$ is **essential** if $\operatorname{ess}(\mathcal{A}) = \mathcal{A}$, or equivalently if $N(\mathcal{A}) = \Bbbk^n$. Note that $L(\operatorname{ess}(\mathcal{A})) \cong L(\mathcal{A})$ as posets. The **rank** of $\mathcal{A}$ is the dimension of its essentialization.

This is why there is an extra power of k in the characteristic polynomial of the arrangement (as opposed to the intersection poset), so that it can record the dimension of $\mathcal{A}$. Specifically,

$$\chi_{\mathcal{A}}(t) = t^{\dim N(\mathcal{A})}\chi_{L(\mathcal{A})}(t) = t^{\dim \mathcal{A} - \operatorname{rank}\mathcal{A}}\chi_{L(\mathcal{A})}(t). \tag{5.2}$$

The two polynomials coincide for essential arrangements. For example, $\operatorname{rank}\mathsf{Br}_n = \dim \operatorname{ess}(\mathsf{Br}_n) = n-1$, and $\operatorname{rank}\mathcal{A}_G = r(G) = |V(G)| - c$, where c is the number of connected components of G.

If $\mathcal{A}$ is linear, then we could define the essentialization by setting $V = N(\mathcal{A})^\perp = \cap\mathcal{A}$, and then defining $\operatorname{ess}(\mathcal{A}) = \{H/V : H \in \mathcal{A}\} \subseteq \Bbbk^n/V$. Thus $\mathcal{A}$ is essential if and only if $\cap\mathcal{A} = 0$. Moreover, if $\mathcal{A}$ is linear then $\operatorname{rank}(\mathcal{A})$ is the rank of its intersection lattice — so rank is a combinatorial invariant, unlike dimension.

Example 5.1.6. If $G = (V, E)$ is a simple graph on vertex set $V = [n]$, then the corresponding **graphic arrangement** $\mathcal{A}_G$ is the subarrangement of Br_n consisting of those hyperplanes $x_i = x_j$ for which $ij \in E$. Thus Br_n itself is the graphic arrangement of the complete graph K_n.

Moreover, the intersection poset of $\mathcal{A}_G$ is precisely the connectivity lattice $K(G)$ defined in Example 1.2.3. Precisely, each subset $\mathcal{B} \subseteq \mathcal{A}_G$ corresponds to a set of edges $E_{\mathcal{B}}$ that induce some partition $\pi \in K(G)$, and the corresponding intersection is

$$\cap\mathcal{B} = \{(x_1, \ldots, x_n) \in \Bbbk^n \mid x_i = x_j \text{ whenever } i, j \text{ belong to the same component of } E_{\mathcal{B}}\}.$$

In particular $\dim \cap\mathcal{B} = |\pi|$, so the definition (5.1) of the characteristic polynomial of $\mathcal{A}_G$ becomes

$$\chi_{\mathcal{A}}(t) = \sum_{\pi \in K(G)} \mu(\hat{\mathbf{0}}, \pi) t^{|\pi|} = p_G(t),$$

the chromatic polynomial of G (by (2.5)). In §5.4, we will see a more explicit way in which the characteristic polynomial enumerates colorings. ◀

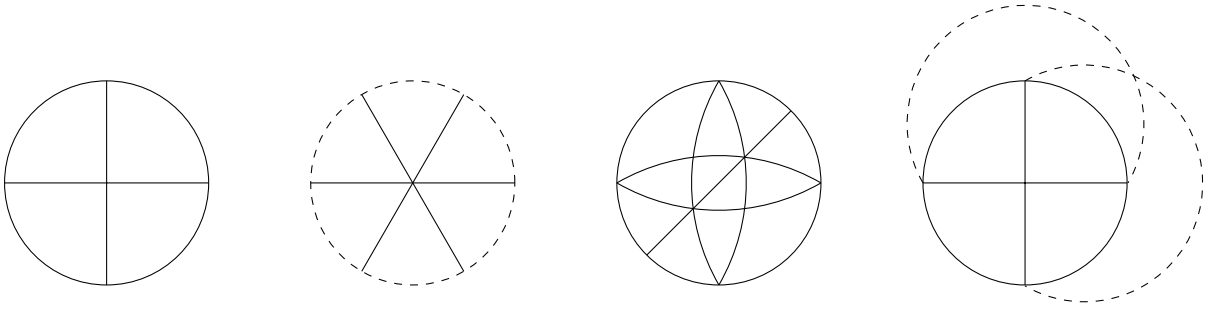
There are two natural operations that go back and forth between central and non-central arrangements, called **projectivization** and **coning**.

Let $\Bbbk$ be a field and $n \geq 1$. The set of lines through the origin in $\Bbbk^n$ is called ***n*-dimensional projective space** over $\Bbbk$ and denoted by $\mathbb{P}^{n-1}\Bbbk$. For example, if $\Bbbk = \mathbb{R}$, we can regard $\mathbb{P}^{n-1}\mathbb{R}$ as the unit sphere $\mathbb{S}^{n-1}$ with opposite points identified. (In particular, it is an $(n-1)$-dimensional manifold, although it is orientable only if n is even.)

Algebraically, write $x \sim y$ if x and y are nonzero scalar multiples of each other. Then $\sim$ is an equivalence relation on $\Bbbk^n \setminus \{0\}$, and $\mathbb{P}^{n-1}$ is the set of equivalence classes. In particular, each linear hyperplanes $H \subset \Bbbk^n$ correspond to a set of equivalence classes that form an affine hyperplane $\operatorname{proj}(H) \subseteq \mathbb{P}^{n-1}\Bbbk$.

Definition 5.1.7. Let $\mathcal{A} \subseteq \Bbbk^n$ be a central arrangement. Its **projectivization** $\operatorname{proj}(\mathcal{A})$ is the affine arrangement $\{\operatorname{proj}(H) \mid H \in \mathcal{A}\}$ in $\mathbb{P}^{n-1}\Bbbk$.

Projectivization supplies a nice way to draw central 3-dimensional real arrangements. Let S be the unit sphere, so that $H \cap S$ is a great circle for every $H \in \mathcal{A}$; then regard $H_0 \cap S$ as the equator and project the northern hemisphere into your piece of paper. Several examples as shown below. Of course, a diagram of $\operatorname{proj}(\mathcal{A})$ only shows the upper half of $\mathcal{A}$; we can recover $\mathcal{A}$ from $\operatorname{proj}(\mathcal{A})$ by "reflecting the interior of the disc to the exterior" (Stanley); e.g., for the Boolean arrangement $\mathcal{A} = \mathsf{Bool}_3$, the picture is as shown in the fourth figure below. In general, $r(\operatorname{proj}(\mathcal{A})) = \frac{1}{2}r(\mathcal{A})$.



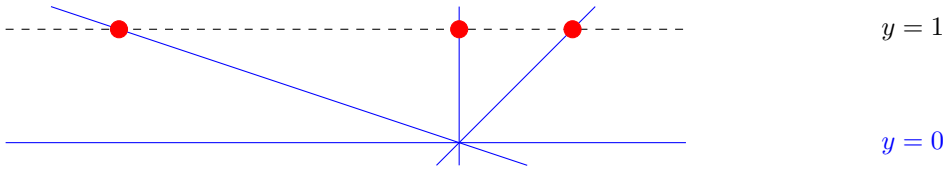
$\mathrm{proj}(\mathsf{Bool}_3)$ $\mathrm{proj}(\mathsf{Br}_3)$ $\mathrm{proj}(\mathsf{ess}(\mathsf{Br}_4))$ Bool_3 from $\mathrm{proj}(\mathsf{Bool}_3)$

The operation of *coning* is a sort of inverse of projectivization. It lets us turn a non-central arrangement into a central arrangement, at the price of increasing the dimension by 1.

Definition 5.1.8. Let $\mathcal{A} \subseteq \Bbbk^n$ be a hyperplane arrangement, not necessarily central. The **cone** $c\mathcal{A}$ is the central arrangement in $\Bbbk^{n+1}$ defined as follows:

- *Geometrically:* Make a copy of $\mathcal{A}$ in $\Bbbk^{n+1}$, choose a point p not in any hyperplane of $\mathcal{A}$, and replace each $H \in \mathcal{A}$ with the affine span H' of p and H (which will be a hyperplane in $\Bbbk^{n+1}$). Then, toss in one more hyperplane containing p and in general position with respect to every H'.
- *Algebraically:* For $H = \{x \colon L(x) = a_i\} \in \mathcal{A}$ (with L a homogeneous linear form on $\Bbbk^n$ and $a_i \in \Bbbk$), construct a hyperplane $H' = \{(x_1, \ldots, x_n, y) \colon L(x) = a_i y\} \subseteq \Bbbk^{n+1}$ in $c\mathcal{A}$. Then, toss in the hyperplane $y = 0$.

For example, if $\mathcal{A}$ consists of the points $x = 0$, $x = -3$ and $x = 1$ in $\mathbb{R}^1$ (shown in red), then $c\mathcal{A}$ consists of the lines $x = y$, $x = -5y$, $x = 3y$, and $y = 0$ in $\mathbb{R}^2$ (shown in blue).



Proposition 5.1.9. $\chi_{c\mathcal{A}}(k) = (k-1)\chi_{\mathcal{A}}(k)$.

5.2 Counting regions: examples

Let $\mathcal{A} \subseteq \mathbb{R}^n$ be a real hyperplane arrangement. The **regions** of $\mathcal{A}$ are the connected components of $\mathbb{R}^n \setminus \mathcal{A}$. Each component is the interior of a (bounded or unbounded) polyhedron; in particular, it is homeomorphic to $\mathbb{R}^n$. We call a region **relatively bounded** if the corresponding region in $\mathrm{ess}(\mathcal{A})$ is bounded. (If $\mathcal{A}$ is not essential then every region is unbounded, because it contains a translate of $W^\perp$, where W is the space defined in Definition 5.1.5. Therefore passing to the essentialization is necessary to make the problem of counting bounded regions nontrivial for all arrangements.) Let

$$\begin{aligned} r(\mathcal{A}) &= \text{ number of regions of } \mathcal{A}, \\ b(\mathcal{A}) &= \text{ number of relatively bounded regions of } \mathcal{A}. \end{aligned}$$

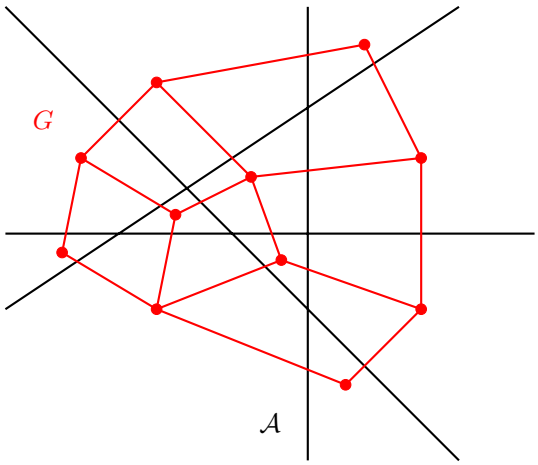
Example 5.2.1. For the arrangements $\mathcal{A}_1$ and $\mathcal{A}_2$ shown in Example 5.1.2,

$$\begin{aligned} r(\mathcal{A}_1) &= 6, & r(\mathcal{A}_2) &= 10, \\ b(\mathcal{A}_1) &= 0, & b(\mathcal{A}_2) &= 2. \end{aligned}$$

◀

Example 5.2.2. The Boolean arrangement Bool_n consists of the n coordinate hyperplanes in $\mathbb{R}^n$. It is a central, essential arrangement whose intersection lattice is the Boolean lattice of rank n; accordingly, $\chi_{\mathsf{Bool}_n}(t) = (t-1)^n$. The complement $\mathbb{R}^n \setminus \mathsf{Bool}_n$ is $\{(x_1, \ldots, x_n) \colon x_i \neq 0 \text{ for all } i\}$, and the connected components are the open orthants, specified by the signs of the n coordinates. Therefore, $r(\mathsf{Bool}_n) = 2^n$ and $b(\mathsf{Bool}_n) = 0$. ◀

Example 5.2.3. Let $\mathcal{A}$ consist of m lines in $\mathbb{R}^2$ in *general position*: that is, no two lines are parallel and no three are coincident. Draw the **dual graph** G, whose vertices are the regions of $\mathcal{A}$, with an edge between every two regions that share a common border.



Let $r = r(\mathcal{A})$ and $b = b(\mathcal{A})$, and let v, e, f denote the numbers of vertices, edges and faces of G, respectively. (In the example above, $(v, e, f) = (11, 16, 7)$.) Each bounded face of G is a quadrilateral that contains exactly one point where two lines of $\mathcal{A}$ meet, and the unbounded face is a cycle of length $r - b$. Therefore,

$$v = r, \tag{5.3a}$$

$$f = 1 + \binom{m}{2} = \frac{m^2 - m + 2}{2} \tag{5.3b}$$

$$4(f-1) + (r-b) = 2e. \tag{5.3c}$$

Moreover, the number $r - b$ of unbounded regions of $\mathcal{A}$ is just $2m$. (Take a walk around a very large circle. You will enter each unbounded region once, and will cross each line twice.) Therefore, from (5.3c) and (5.3b) we obtain

$$e = m + 2(f-1) = m^2. \tag{5.3d}$$

Euler's formula for planar graphs says that $v - e + f = 2$. Substituting in (5.3a), (5.3b) and (5.3d) and solving for r gives

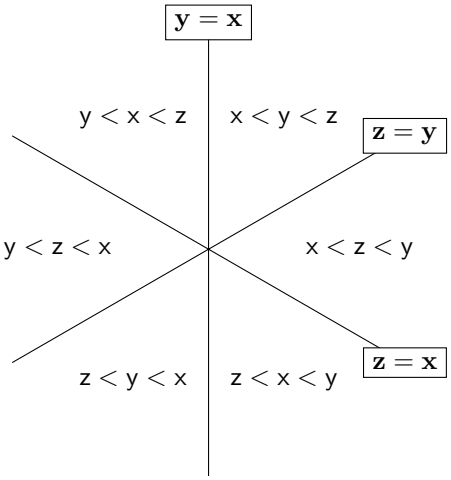
$$r = \frac{m^2 + m + 2}{2}$$

and therefore

$$b = r - 2m = \frac{m^2 - 3m + 2}{2} = \binom{m-1}{2}.$$

◀

Example 5.2.4. The braid arrangement Br_n consists of the $\binom{n}{2}$ hyperplanes $H_{ij} = \{\mathbf{x}\colon x_i = x_j\}$ in $\mathbb{R}^n$. The complement $\mathbb{R}^n \setminus \mathsf{Br}_n$ consists of all vectors in $\mathbb{R}^n$ with no two coordinates equal, and the connected components of this set are specified by the ordering of the set of coordinates as real numbers:



Therefore, $r(\mathsf{Br}_n) = n!$. (Stanley: "Rarely is it so easy to compute the number of regions!") Furthermore,

$$\chi_{\mathsf{Br}_n}(t) = t(t-1)(t-2)\cdots(t-n+1).$$

Note that the braid arrangement is central but not essential; its center is the line $x_1 = x_2 = \cdots = x_n$, so its rank is $n-1$. ◀

Example 5.2.5. Let $G = (V, E)$ be a simple graph with $V = [n]$, and let $\mathcal{A}_G$ be its graphic arrangement (see Example 5.1.6). The characteristic polynomial of $L(\mathcal{A}_G)$ is precisely the chromatic polynomial of G (see Section 4.4). We will see another explanation for this fact later; see Example 5.4.4.

The regions of $\mathbb{R}^n \setminus \mathcal{A}_G$ are the open polyhedra whose defining inequalities include either $x_i < x_j$ or $x_i > x_j$ for each edge $ij \in E$. Those inequalities give rise to an orientation of G, and it is not hard to check that this correspondence is a bijection between regions and acyclic orientations. Hence

$$r(\mathcal{A}_G) \;=\; \text{number of acyclic orientations of } G \;=\; |\chi_{L(\mathcal{A}_G)}(-1)|. \qquad ◀$$

5.3 Zaslavsky's theorems

Example 5.2.5 motivates the main result of this section, historically the first major theorem about hyperplane arrangements.

Theorem 5.3.1 (Zaslavsky's Theorem [Zas75])**.** *Let $\mathcal{A}$ be a real hyperplane arrangement, and let $\chi_{\mathcal{A}}$ be the characteristic polynomial of its intersection poset. Then*

$$r(\mathcal{A}) = (-1)^{\dim \mathcal{A}} \chi_{\mathcal{A}}(-1) \text{ and} \tag{5.4}$$

$$b(\mathcal{A}) = (-1)^{\operatorname{rank} \mathcal{A}} \chi_{\mathcal{A}}(1). \tag{5.5}$$

The proof has the following parts:

1. Show that r and b satisfy restriction/contraction recurrences in terms of associated hyperplane arrangements $\mathcal{A}'$ and $\mathcal{A}''$ (Prop. 5.3.3).
2. Rewrite the characteristic polynomial $\chi_{\mathcal{A}}(k)$ as a sum over central subarrangements of $\mathcal{A}$ (the "Whitney formula", Prop. 5.3.4).
3. Show that the Whitney formula obeys a restriction/contraction recurrence (Prop. 5.3.5) and compare it with those for r and b.

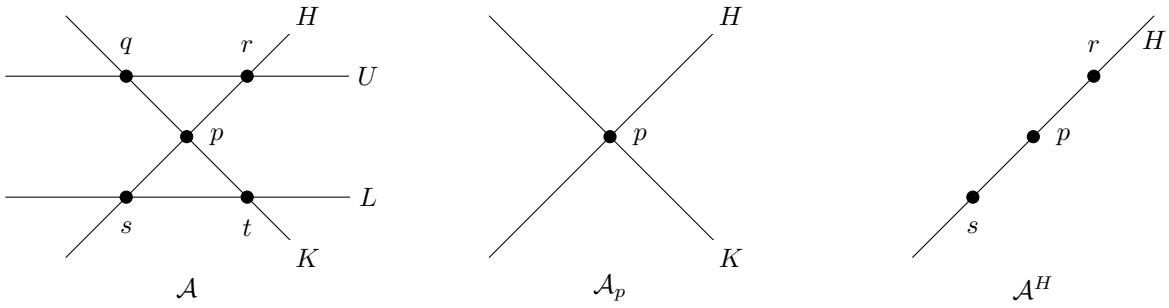
Let $x \in L(\mathcal{A})$, i.e., x is a nonempty affine space formed by intersecting some of the hyperplanes in $\mathcal{A}$. Define

$$\begin{aligned} \mathcal{A}_x &= \{H \in \mathcal{A} \colon H \supseteq x\}, \\ \mathcal{A}^x &= \{H \cap x \colon H \in \mathcal{A} \setminus \mathcal{A}_x\}. \end{aligned} \tag{5.6}$$

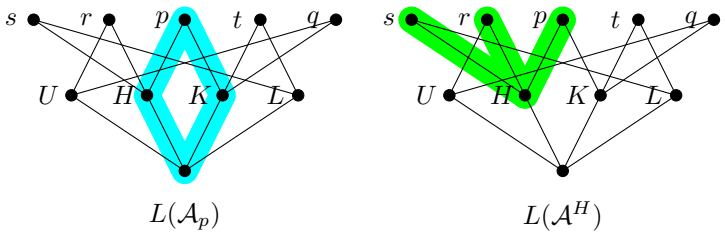
In other words, $\mathcal{A}_x$ is obtained by deleting the hyperplanes not containing x, while $\mathcal{A}^x$ is obtained by restricting $\mathcal{A}$ to x so as to get an arrangement whose ambient space is x itself. The notation is mnemonic: $L(\mathcal{A}_x)$ and $L(\mathcal{A}^x)$ are isomorphic respectively to the principal order ideal and principal order filter generated by x in $L(\mathcal{A})$. That is,

$$L(\mathcal{A}_x) \cong \{y \in L(\mathcal{A}) \colon y \leq x\}, \qquad L(\mathcal{A}^x) \cong \{y \in L(\mathcal{A}) \colon y \geq x\}.$$

Example 5.3.2. Let $\mathcal{A}$ be the 2-dimensional arrangement shown on the left, with the line H and point p as shown. Then $\mathcal{A}_p$ and $\mathcal{A}^H$ are shown on the right.



The lattice $L(\mathcal{A})$ and its subposets (in this case, sublattices) $L(\mathcal{A}_p)$ and $L(\mathcal{A}^H)$ are shown below.



◀

Let $M(\mathcal{A})$ be the matroid represented by normal vectors $\{\mathbf{n}_H : H \in \mathcal{A}\}$. Fix a hyperplane $H \in \mathcal{A}$ and let

$$\mathcal{A}' = \mathcal{A} \setminus H, \qquad \mathcal{A}'' = \mathcal{A}^H. \tag{5.7}$$

Proposition 5.3.3. *The invariants r and b satisfy the following recurrences:*

1. $r(\mathcal{A}) = r(\mathcal{A}') + r(\mathcal{A}'')$.
2. $b(\mathcal{A}) = \begin{cases} 0 & \text{if } \operatorname{rank}\mathcal{A} = \operatorname{rank}\mathcal{A}' + 1 \quad \textit{(i.e., if } \mathbf{n}_H \textit{ is a coloop in } M(\mathcal{A})\textit{)}, \\ b(\mathcal{A}') + b(\mathcal{A}'') & \text{if } \operatorname{rank}\mathcal{A} = \operatorname{rank}\mathcal{A}' \quad \textit{(i.e., if it isn't)}. \end{cases}$

Proof. (1) Consider what happens when we add H to $\mathcal{A}'$ to obtain $\mathcal{A}$. Some regions of $\mathcal{A}'$ will remain the same, while others will be split into two regions.

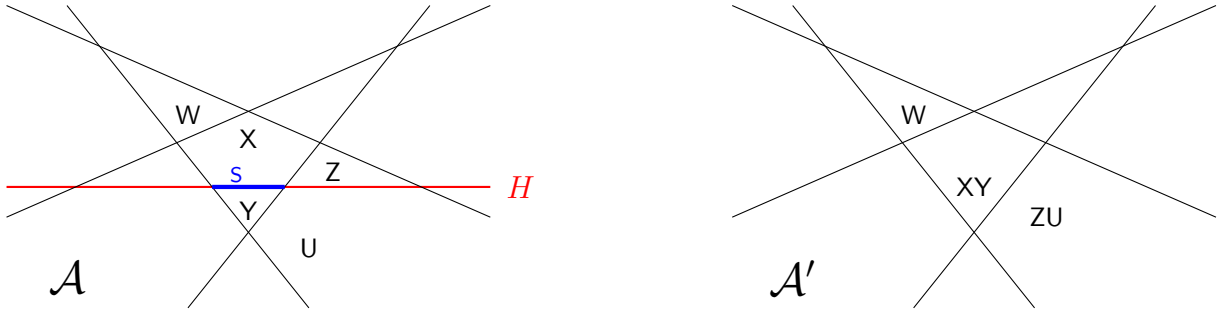
Let s and u be the numbers of split and unsplit regions of $\mathcal{A}'$ (so in the figure above, $s = 2$ and $u = 4$). The unsplit regions each contribute 1 to $r(\mathcal{A})$. The split regions each contribute 2 to $r(\mathcal{A})$, but they also correspond bijectively to the regions of $\mathcal{A}''$. (See, e.g., Example 5.3.2.) So

$$r(\mathcal{A}') = s + u, \qquad r(\mathcal{A}) = 2s + u, \qquad r(\mathcal{A}'') = s$$

and so $r(\mathcal{A}) = r(\mathcal{A}') + r(\mathcal{A}'')$, proving the first assertion of Proposition 5.3.3. By the way, if (and only if) H is a coloop then it borders *every* region of $\mathcal{A}$, so $r(\mathcal{A}) = 2r(\mathcal{A}')$ in this case.

(2) Now we count bounded regions.

- If $\operatorname{rank}\mathcal{A} = \operatorname{rank}\mathcal{A}' + 1$, then $N(\mathcal{A}') \subsetneq \mathbb{R}^n$, i.e., $\mathcal{A}'$ is not essential. In that case, every region of $\mathcal{A}'$ must contain a translate of the nonempty vector space $N(\mathcal{A}')^\perp$, which gets squashed down to a point upon essentialization. In particular, every region of $\mathcal{A}$ contains half of that translate, hence is unbounded, so $b(\mathcal{A}) = 0$.
- If $\operatorname{rank}\mathcal{A} = \operatorname{rank}\mathcal{A}'$, then the relatively bounded regions of $\mathcal{A}$ come in a few different flavors.
  - If W is a bounded region not bordered by H, then it is also a bounded region of $\mathcal{A}'$.
  - If X, Y are two bounded regions separated by part of H, then they merge into one bounded region XY in $\mathcal{A}'$. Also, the part of H that separates them (namely $\mathsf{S} = \partial\mathsf{X} \cap H = \partial\mathsf{Y} \cap H$) is itself a bounded region of $\mathcal{A}''$.
  - If Z is a bounded region of $\mathcal{A}$ whose neighbor across H is an unbounded region U, then the merged region ZU is unbounded in $\mathcal{A}'$, but $\partial\mathsf{Z} \cap H$ is nevertheless bounded in $\mathcal{A}''$.



In short, here are the possibilities:

	Description	Contributions to… $b(\mathcal{A})$	$b(\mathcal{A}')$	$b(\mathcal{A}'')$
(W)	bounded regions that don't touch H	1	1	0
(X, Y)	pairs of bounded regions separated by H	2	1	1
(Z)	bounded, neighbor across H is unbounded	1	0	1

In all cases the contribution to $b(\mathcal{A})$ equals the sum of those to $b(\mathcal{A}')$ and $b(\mathcal{A}'')$, establishing the second desired recurrence. □

Proposition 5.3.3 looks a lot like a Tutte polynomial deletion/contraction recurrence. This suggests that we should be able to extract $r(\mathcal{A})$ and $b(\mathcal{A})$ from the characteristic polynomial $\chi_{\mathcal{A}}$. The first step is to find a more convenient form for the characteristic polynomial.

Proposition 5.3.4 (Whitney formula for $\chi_{\mathcal{A}}$). *For any hyperplane arrangement $\mathcal{A}$,*

$$\chi_{\mathcal{A}}(t) = \sum_{\text{central } \mathcal{B} \subseteq \mathcal{A}} (-1)^{|\mathcal{B}|} t^{\dim \mathcal{A} - \operatorname{rank} \mathcal{B}}.$$

Whitney [Whi32b, Whi32a] gave an equivalent formula for the chromatic polynomial of a graph, well before anyone was talking about crosscuts or hyperplane arrangements in those terms. So the more general version gets his name on it as well. I don't know whether the modern proof given below specializes to Whitney's proof.

Proof. For $x \in L(\mathcal{A})$, consider the interval $[\hat{\mathbf{0}}, x]$ as a sublattice of $L(\mathcal{A})$. Its atoms are the hyperplanes of $\mathcal{A}$ containing x, and they form a lower crosscut of $[\hat{\mathbf{0}}, x]$. Therefore,

$$\begin{aligned} \chi_{\mathcal{A}}(t) &= \sum_{x \in L(\mathcal{A})} \mu(\hat{\mathbf{0}}, x) t^{\dim x} \\ &= \sum_{x \in L(\mathcal{A})} \left(\sum_{\mathcal{B} \subseteq \mathcal{A}:\ x = \bigcap \mathcal{B}} (-1)^{|\mathcal{B}|} \right) t^{\dim x} \end{aligned}$$

(by the second form of Rota's crosscut theorem (Thm. 2.4.9); note that $\hat{\mathbf{1}}_{[\hat{\mathbf{0}},x]} = x$)

$$\begin{aligned} &= \sum_{\mathcal{B}\subseteq\mathcal{A}:\ \bigcap\mathcal{B}\neq 0} (-1)^{|\mathcal{B}|} t^{\dim(\bigcap\mathcal{B})} \\ &= \sum_{\text{central } \mathcal{B}\subseteq\mathcal{A}} (-1)^{|\mathcal{B}|} t^{\dim\mathcal{A}-\operatorname{rank}\mathcal{B}} \end{aligned}$$

as desired. Note that the empty subarrangement is considered central for the purpose of this formula (since by convention its intersection is $\mathbb{R}^{\dim\mathcal{A}}$), corresponding to the summand $x = \hat{\mathbf{0}}$ and giving rise to the leading term $t^{\dim\mathcal{A}}$ of $\chi_{\mathcal{A}}(t)$. □

Proposition 5.3.5. *Let $\mathcal{A}$ be a hyperplane arrangement in $\Bbbk^n$. As before, let $H \in \mathcal{A}$ and define $\mathcal{A}', \mathcal{A}''$ as in (5.7). Then $\chi_{\mathcal{A}}(t) = \chi_{\mathcal{A}'}(t) - \chi_{\mathcal{A}''}(t)$.*

Sketch of proof. Splitting up Whitney's formula gives

$$\chi_{\mathcal{A}}(t) \;=\; \underbrace{\sum_{\text{central } \mathcal{B}\subseteq\mathcal{A}:\ H\notin\mathcal{B}} (-1)^{|\mathcal{B}|} t^{n-\operatorname{rank}\mathcal{B}}}_{SUM1} \;+\; \underbrace{\sum_{\text{central } \mathcal{B}\subseteq\mathcal{A}:\ H\in\mathcal{B}} (-1)^{|\mathcal{B}|} t^{n-\operatorname{rank}\mathcal{B}}}_{SUM2}.$$

Then $SUM1 = \chi_{\mathcal{A}'}(t)$ (it is just Whitney's formula for $\mathcal{A}'$), so it remains to show that $SUM2 = -\chi_{\mathcal{A}''}(t)$. This is a little trickier, because different hyperplanes in $\mathcal{A}$ can have the same intersection with H, which means that multiple subarrangements of $\mathcal{A}$ can give rise to the same subarrangement of $\mathcal{A}''$.

Label the hyperplanes of $\mathcal{A}''$ (which, remember, are codimension-1 subspaces of H) as $K_1, \ldots, K_s$. Then $\mathcal{A}''$ is the union of the pairwise-disjoint sets $\mathcal{A}_i = \{J \in \mathcal{A} :\ J \cap H = K_i\}$, for $i \in [s]$. Each arrangement $\mathcal{B}$ arising as a summand of $SUM2$ gives rise to a central subarrangement of $\mathcal{A}''$, namely

$$\pi(\mathcal{B}) = \{J \cap H :\ J \in \mathcal{B}\},$$

that depends only on the values of i for which $\mathcal{A}_i \cap \mathcal{B} \neq \emptyset$ (so that $K_i \in \pi(\mathcal{B})$). That is, for each central subarrangement $\mathcal{B}'' \subseteq \mathcal{A}''$, the summands $\mathcal{B}$ of $SUM2$ such that $\pi(\mathcal{B}) = \mathcal{B}''$ are precisely the arrangements of the form

$$\{H\} \cup \bigcup_{i:\ K_i\in\mathcal{B}''} \mathcal{B}_i$$

where $\emptyset \subsetneq \mathcal{B}_i \subseteq \mathcal{A}_i$ for all i. Moreover, in this case,

$$n - \operatorname{rank}\mathcal{B} \;=\; \dim(\cap\mathcal{B}) \;=\; \dim(\cap\pi(\mathcal{B})) \;=\; \dim H - \operatorname{rank}\mathcal{B}''. \tag{5.8}$$

Now we break up $SUM2$ into subsums depending on $\pi(\mathcal{B})$:

$$\begin{aligned} SUM2 \;&=\; \sum_{\text{central } \mathcal{B}''\subseteq\mathcal{A}''} \;\sum_{\mathcal{B}\in\pi^{-1}(\mathcal{B}'')} (-1)^{|\mathcal{B}|} t^{n-\operatorname{rank}\mathcal{B}} \\ &= \sum_{\mathcal{B}''} \left(\sum_{\mathcal{B}\in\pi^{-1}(\mathcal{B}'')} (-1)^{|\mathcal{B}|} \right) t^{\dim H-\operatorname{rank}\mathcal{B}''} \\ &= -\sum_{\mathcal{B}''} \left(\prod_{i:\ K_i\in\mathcal{B}''} \sum_{\emptyset\neq\mathcal{B}_i\subseteq\mathcal{A}_i} (-1)^{|\mathcal{B}_i|} \right) t^{\dim H-\operatorname{rank}\mathcal{B}''} \end{aligned}$$

(to see this, expand the product and observe that equals the inner sum in the previous line; the outer minus sign is contributed by H, which is an element of $\mathcal{B}$). But $\sum_{\emptyset \neq \mathcal{B}_i \subseteq \mathcal{A}_i} (-1)^{|\mathcal{B}_i|} = -1$, because it is the binomial expansion of $(1-1)^{|\mathcal{A}_i|} = 0$, with one $+1$ term (namely $\mathcal{B}_i = \emptyset$) removed. (Note that $\mathcal{A}_i \neq \emptyset$.) Therefore, the whole thing boils down to

$$-\sum_{\mathcal{B}''} (-1)^{|\mathcal{B}''|} t^{\dim H - \operatorname{rank} \mathcal{B}''}$$

which is just Whitney's formula for $-\chi_{\mathcal{A}''}(t)$. □

Remark 5.3.6. This recurrence is strongly reminiscent of the chromatic recurrence (4.10). Indeed, if $\mathcal{A} = \mathcal{A}_G$ is a graphic arrangement in $\mathbb{R}^n$, e is an edge of G, and H_e is the corresponding hyperplane in $\mathcal{A}_G$, then it is clear that $\mathcal{A}_{G \setminus e} = \mathcal{A}_G \setminus \{H_e\}$. In addition, two hyperplanes $H_f, H_{f'}$ will have the same intersection with H_e if and only if f, f' become parallel upon contracting e, so $\mathcal{A}_{G/e}$ can be identified with $(\mathcal{A}_G)^{H_e}$ (where the coordinates on $H_e \cong \mathbb{R}^{n-1}$ are given by equating the coordinates for the two endpoints of e).

We can now finish the proof of the main result. We have already done the hard work, and just need to put all the pieces together.

Proof of Zaslavsky's Theorem 5.3.1. Let $\tilde{r}(\mathcal{A})$ and $\tilde{b}(\mathcal{A})$ denote the numbers on the right-hand sides of (5.4) and (5.5).

If $|\mathcal{A}| = 1$, then $L(\mathcal{A})$ is the lattice with two elements, namely $\mathbb{R}^n$ and a single hyperplane H, and its characteristic polynomial is $t^n - t^{n-1}$. Thus $\tilde{r}(\mathcal{A}) = (-1)^n((-1)^n - (-1)^{n-1}) = 2$ and $\tilde{b}(\mathcal{A}) = -(1-1) = 0$, which match $r(\mathcal{A})$ and $b(\mathcal{A})$.

For the general case, we just need to show that $\tilde{r}$ and $\tilde{b}$ satisfy the same recurrences as r and b (see Prop. 5.3.3). First,

$$\begin{aligned}
\tilde{r}(\mathcal{A}) &= (-1)^{\dim \mathcal{A}} \chi_{\mathcal{A}}(-1) \\
&= (-1)^{\dim \mathcal{A}} \big(\chi_{\mathcal{A}'}(-1) - \chi_{\mathcal{A}''}(-1)\big) && \text{(by Prop. 5.3.5)} \\
&= (-1)^{\dim \mathcal{A}'} \chi_{\mathcal{A}'}(-1) + (-1)^{\dim \mathcal{A}''} \chi_{\mathcal{A}''}(-1) && \text{(since } \dim \mathcal{A}'' = \dim \mathcal{A} - 1\text{)} \\
&= \tilde{r}(\mathcal{A}') + \tilde{r}(\mathcal{A}'').
\end{aligned}$$

As for $\tilde{b}$, if $\operatorname{rank} \mathcal{A} = \operatorname{rank} \mathcal{A}' + 1$, then in fact $\mathcal{A}'$ and $\mathcal{A}''$ have the same essentialization, hence the same rank, and their characteristic polynomials only differ by a factor of t. The deletion/restriction recurrence (Prop. 5.3.5) therefore implies $\tilde{b}(\mathcal{A}) = 0$.

On the other hand, if $\operatorname{rank} \mathcal{A} = \operatorname{rank} \mathcal{A}'$, then $\operatorname{rank} \mathcal{A}'' = \operatorname{rank} \mathcal{A} - 1$ and a calculation similar to that for $\tilde{r}$ (replacing dimension with rank) shows that $\tilde{b}(\mathcal{A}) = \tilde{b}(\mathcal{A}') + \tilde{b}(\mathcal{A}'')$. □

Corollary 5.3.7. *Let $\mathcal{A} \subseteq \mathbb{R}^n$ be a central hyperplane arrangement and let $M = M(\mathcal{A})$ be the matroid represented by normals. Then $r(\mathcal{A}) = T_M(2,0)$ and $b(\mathcal{A}) = 0$.*

Proof. Combine Zaslavsky's theorem with the formula $\chi_{\mathcal{A}}(t) = (-1)^n T_M(1-t, 0)$ **which needs to be proved!**, and use the fact that $T_M(0,0) = 0$ for any matroid M with nonempty ground set. □

Remark 5.3.8. The formula for $r(\mathcal{A})$ could be obtained from the Tutte Recipe Theorem (Thm. 4.2.1). But this would not work for $b(\mathcal{A})$, which is not an invariant of $M(\mathcal{A})$. (The matroid $M(\mathcal{A})$ is not as meaningful when $\mathcal{A}$ is not central, which is precisely the case that $b(\mathcal{A})$ is interesting.)

Example 5.3.9. Let $s \geq n$, and let $\mathcal{A}$ be an arrangement of s linear hyperplanes in general position in $\mathbb{R}^n$; that is, every k hyperplanes intersect in a space of dimension $n-k$ (or 0 if $k > n$). Equivalently, the corresponding matroid M is $U_n(s)$, whose rank function $r : 2^{[s]} \to \mathbb{N}$ is given by $r(A) = \min(n, |A|)$. Therefore,

$$\begin{aligned} r(\mathcal{A}) \;=\; T_M(2,0) \;&=\; \sum_{A \subseteq [s]} (2-1)^{n-r(A)}(0-1)^{|A|-r(A)} \\ &= \sum_{A \subseteq [s]} (-1)^{|A|-r(A)} \\ &= \sum_{k=0}^{s} \binom{s}{k} (-1)^{k-\min(n,k)} \\ &= \sum_{k=0}^{n} \binom{s}{k} + \sum_{k=n+1}^{s} \binom{s}{k} (-1)^{k-n} \\ &= \sum_{k=0}^{n} \binom{s}{k} (1-(-1)^{k-n}) + \underbrace{\sum_{k=0}^{s} \binom{s}{k} (-1)^{k-n}}_{=0} \\ &= 2\left(\binom{s}{n-1} + \binom{s}{n-3} + \binom{s}{n-5} + \cdots \right). \end{aligned}$$

For instance, if $n = 3$ then

$$r(\mathcal{A}) = 2\binom{s}{2} + 2\binom{s}{0} = s^2 - s + 2.$$

◀

Notice that this is not the same as the number of regions formed by s *affine* lines in general position in $\mathbb{R}^2$. The calculation of $r(\mathcal{A})$ and $b(\mathcal{A})$ for that arrangement is left to the reader (Problem 5.1).

Corollary 5.3.10. *Let $\mathcal{A}$ be an arrangement in which no two hyperplanes are parallel. Then $\mathcal{A}$ has at least one relatively bounded region if and only if it is noncentral.* ***Prove this and find a place for it — assuming it is true. The non-parallel assumption is necessary since the conclusion fails for the arrangement with hyperplanes $x = 0, y = 0, y = 1$.***

5.4 The finite field method

The following very important result is implicit in the work of Crapo and Rota [CR70] and was stated explicitly by Athanasiadis [Ath96]:

Theorem 5.4.1. *Let $\mathbb{F}_q$ be the finite field of order q, and let $\mathcal{A} \subseteq \mathbb{F}_q^n$ be a hyperplane arrangement. Then*

$$|\mathbb{F}_q^n \setminus \mathcal{A}| = \chi_{\mathcal{A}}(q).$$

This result gives a combinatorial interpretation of the values of the characteristic polynomial. In practice, it is often used to calculate the characteristic polynomial of a hyperplane arrangement by counting points in its complement over $\mathbb{F}_q$ (which can be regarded as regions of the complement, if you endow $\mathbb{F}_q^n$ with the discrete topology).

Proof #1. By inclusion-exclusion,

$$|\mathbb{F}_q^n \setminus \mathcal{A}| \;=\; \sum_{\mathcal{B}\subseteq\mathcal{A}} (-1)^{|\mathcal{B}|} \left|\bigcap \mathcal{B}\right|.$$

If $\mathcal{B}$ is not central, then by definition $|\bigcap \mathcal{B}| = 0$. Otherwise, $|\bigcap \mathcal{B}| = q^{n-\operatorname{rank}\mathcal{B}}$. So the sum becomes

$$|\mathbb{F}_q^n \setminus \mathcal{A}| \;=\; \sum_{\text{central } \mathcal{B}\subseteq\mathcal{A}} (-1)^{|\mathcal{B}|} q^{n-\operatorname{rank}\mathcal{B}}$$

which is just Whitney's formula for $\chi_{\mathcal{A}}(q)$ (Prop. 5.3.4). □

Proof #2. Start with the definition of the characteristic polynomial, letting r be the rank function in $L(\mathcal{A})$:

$$\begin{aligned}
\chi_{\mathcal{A}}(q) &= \sum_{x\in L(\mathcal{A})} \mu(\hat{\mathbf{0}}, x) q^{n-r(x)} \\
&= \sum_{x\in L(\mathcal{A})} \mu(\hat{\mathbf{0}}, x) q^{\dim x} \\
&= \sum_{x\in L(\mathcal{A})} \mu(\hat{\mathbf{0}}, x) |x| \\
&= \sum_{\mathbf{p}\in\mathbb{F}_q^n} \sum_{x\in L(\mathcal{A}):\ \mathbf{p}\in x} \mu(\hat{\mathbf{0}}, x) \\
&= \sum_{\mathbf{p}\in\mathbb{F}_q^n} \left(\sum_{x\in[\hat{\mathbf{0}}, y_{\mathbf{p}}]} \mu(\hat{\mathbf{0}}, x) \right)
\end{aligned}$$

where $y_{\mathbf{p}} = \bigcap_{H\supseteq\mathbf{p}} H$. By definition of the Möbius function, the parenthesized sum is 1 if $y_{\mathbf{p}} = \hat{\mathbf{0}}$ and 0 otherwise. Therefore

$$\chi_A(q) = \#\{\mathbf{p}\in\mathbb{F}_q^n :\ y_{\mathbf{p}} = \hat{\mathbf{0}}\} = \#\{\mathbf{p}\in\mathbb{F}_q^n :\ \mathbf{p}\notin H\ \ \forall H\in\mathcal{A}\} = |\mathbb{F}_q^n \setminus \mathcal{A}|.$$ □

This fact has a much more general application, which was systematically mined by Athanasiadis, e.g., [Ath96].

Definition 5.4.2. Let $\mathcal{A}\subseteq\mathbb{R}^n$ be an integral hyperplane arrangement (i.e., whose hyperplanes are defined by equations with integer coefficients). For a prime p, let $\mathcal{A}_p = \mathcal{A}\otimes\mathbb{F}_p$ be the arrangement in $\mathbb{F}_p^n$ defined by regarding the equations in $\mathcal{A}$ as equations over $\mathbb{F}_p$. We say that $\mathcal{A}$ **reduces correctly modulo p** if $L(\mathcal{A}_p)\cong L(\mathcal{A})$. (We need only consider the prime case, since if q is a power of p, then $L(\mathcal{A}_q) = L(\mathcal{A}_p)$.)

A sufficient condition for correct reduction is that no minor of the matrix of normal vectors is a nonzero multiple of p (so that rank calculations are the same over $\mathbb{F}_p$ as over $\mathbb{Z}$). In particular, if we choose p larger than the absolute value of any minor of M, then each set of columns of M is linearly independent over $\mathbb{F}_p$ iff it is independent over $\mathbb{Q}$. There are infinitely many such primes, implying the following highly useful result:

Theorem 5.4.3 (The finite field method). *Let $\mathcal{A}\subseteq\mathbb{R}^n$ be an integral hyperplane arrangement and q a power of a large enough prime. Then $\chi_{\mathcal{A}}(q)$ is the polynomial that counts points in the complement of $\mathcal{A}_q$.*

Example 5.4.4. Let $G = ([n], E)$ be a simple graph and let $\mathcal{A}_G$ be the corresponding graphic arrangement in $\mathbb{R}^n$. Note that $\mathcal{A}_G$ reduces correctly over every finite field $\mathbb{F}_q$ (because graphic matroids are regular). A point $(x_1,\dots,x_n)\in\mathbb{F}_q^n$ can be regarded as the q-coloring of G that assigns color x_i to vertex i. The

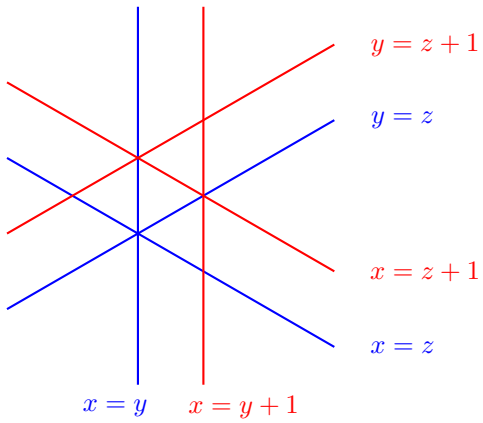
proper q-colorings are precisely the points of $\mathbb{F}_q^n \setminus \mathcal{A}_G$. The number of such colorings is $p_G(q)$ (the chromatic polynomial of G evaluated at q). On the other hand, by Theorem 5.4.1, it is also the characteristic polynomial $\chi_{\mathcal{A}_G}(q)$. Since $p_G(q) = \chi_{\mathcal{A}_G}(q)$ for infinitely many q (namely, all integer prime powers), the polynomials must be equal. In particular, by Zaslavsky's theorems, the number of regions is $p_G(-1)$, and on the other hand we know that regions of $\mathcal{A}_G$ are in bijection with acyclic orientations of G (see Example 5.2.5), so we now have a geometric proof of Theorem 4.5.2. ◀

Example 5.4.5. The **Shi arrangement** is the arrangement of $n(n-1)$ hyperplanes in $\mathbb{R}^n$ defined by

$$\mathsf{Shi}_n = \{x_i = x_j,\ x_i = x_j + 1 \mid 1 \le i < j \le n\}.$$

In other words, take the braid arrangement, clone it, and nudge each of the cloned hyperplanes a little bit in the direction of the bigger coordinate. The Shi arrangement has rank $n-1$ (every hyperplane in it contains a line parallel to the all-ones vector), so we may project along that line to obtain the essentialization in $\mathbb{R}^{n-1}$. Thus $\mathrm{ess}(\mathsf{Shi}_2)$ consists of two points on a line, while $\mathrm{ess}(\mathsf{Shi}_3)$ is shown below.

$\mathsf{ess}(\mathsf{Shi}_3)$



We will prove that the characteristic polynomial of the Shi arrangement is

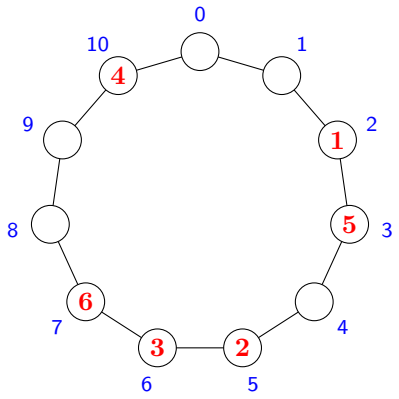
$$\chi_{\mathsf{Shi}_n}(q) = q(q-n)^{n-1}. \tag{5.9}$$

In particular, the numbers of regions and bounded regions are respectively

$$r(\mathsf{Shi}_n) = |\chi(-1)| = (n+1)^{n-1}, \qquad b(\mathsf{Shi}_n) = |\chi(1)| = (n-1)^{n-1}. \tag{5.10}$$

(The number $(n+1)^{n-1}$ may look familiar; by Cayley's formula, it is the number of spanning trees of the complete graph K_{n+1}. It also counts many other things of combinatorial interest, including parking functions.)

The following proof is from [Sta07, §5.2]. By Theorem 5.4.3, it suffices to count the points in $\mathbb{F}_q^n \setminus \mathsf{Shi}_n$ for a large enough prime q. Let $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{F}_q^n \setminus \mathsf{Shi}_n$. Draw a necklace with q beads labeled by the elements $0, 1, \ldots, q-1 \in \mathbb{F}_q$, and for each $k \in [n]$, put a big red $\boldsymbol{k}$ on the x_k-th bead. For example, let $n = 6$ and $q = 11$. Then the necklace for $\mathbf{x} = (2, 5, 6, 10, 3, 7)$ is as follows:



The requirement that $\mathbf{x}$ avoids the hyperplanes $x_i = x_j$ implies that the red numbers are all on different beads. If we read the red numbers clockwise, starting at **1** and putting in a divider sign | for each bead without a red number, we get

$$15 \mid 236 \mid \mid 4 \mid$$

which can be regarded as the *ordered weak partition* (or OWP)

$$\Pi(x) = 15, 236, \emptyset, 4, \emptyset$$

that is, a $(q-n)$-tuple $B_1, \ldots, B_{q-n}$, where the B_i are pairwise disjoint sets (possibly empty; that's what the "weak" means) whose union is $[n]$, and $1 \in B_1$. (We've omitted the divider corresponding to the bead just counterclockwise of **1**; stay tuned.)

Note that each block of $\Pi(\mathbf{x})$ corresponds to a contiguous set of values among the coordinates of $\mathbf{x}$. For example, the block 236 occurs because the values 5,6,7 occur in coordinates x_2, x_3, x_6. In order to avoid the hyperplanes $x_i = x_j + 1$ for $i < j$, each contiguous block of beads must have its red numbers in strictly increasing order counterclockwise. (In particular the bead just counterclockwise of **1** must be unlabeled, which is why we could omit that divider.)

Thus we have bijections

$$\mathbb{F}_q^n \setminus \mathsf{Shi}_n \quad \longleftrightarrow \quad \{\text{necklaces}\} \quad \longleftrightarrow \quad \{\text{OWPs of } [n] \text{ with } q-n \text{ blocks}\}.$$

To get a necklace from an OWP, write out each block in increasing order, with bars between successive blocks.

Meanwhile, an OWP is given by a function $f : [n] \to [q-n]$, where $f(i)$ is the index of the block containing i (so $f(1) = 1$). There are $(q-n)^{n-1}$ such things. Since there are q choices for the bead containing the red **1**, we obtain

$$\left|\mathbb{F}_q^n \setminus \mathsf{Shi}_n\right| = q(q-n)^{n-1} = \chi_{\mathsf{Shi}_n}(q).$$

This proves (5.9), and (5.10) follows from Zaslavsky's theorems. ◀

5.5 Supersolvable lattices and arrangements

We have seen that for a simple graph $G = ([n], E)$, the chromatic polynomial $p_G(k)$ is precisely the characteristic polynomial of the graphic arrangement $\mathcal{A}_G$. For some graphs, the chromatic polynomial factors

into linear terms over $\mathbb{Z}$. For example, if $G = K_n$, then $p_G(k) = k(k-1)(k-2)\cdots(k-n+1)$, and if G is a forest with n vertices and c components, then $p_G(k) = k^c(k-1)^{n-c}$. This property does not hold for all graphs. For example, it is easy to work out that the chromatic polynomial of C_4 (the cycle with four vertices and four edges) is $k^4 - 4k^3 + 6k^2 - 3k = k(k-1)(k^2-3k+k)$, which does not factor further over $\mathbb{Z}$. Is there a structural condition on a graph or a central arrangement (or really, on a geometric lattice) that will guarantee that its characteristic polynomial factors completely? It turns out that *supersolvable* geometric lattices have this good property.

Definition 5.5.1. Let L be a ranked lattice. An element $x \in L$ is a **modular element** if $r(x) + r(y) = r(x \vee y) + r(x \wedge y)$ for every $y \in L$.

For example:

- By Theorem 1.5.6, a ranked lattice L is modular iff all elements are modular.
- The elements $\hat{\mathbf{0}}$ and $\hat{\mathbf{1}}$ are clearly modular in any lattice.
- If L is geometric, then every atom x is modular. Indeed, for $y \in L$, if $y \geq x$, then $y = x \vee y$ and $x = x \wedge y$, while if $y \not\geq x$ then $y \wedge x = \hat{\mathbf{0}}$ and $y \vee x \succ y$.
- The coatoms of a geometric lattice need not be modular. For example, let $L = \Pi_n$, and recall that Π_n has rank function $r(\pi) = n - |\pi|$. Let $x = 12|34$, $y = 13|24 \in \Pi_4$. Then $r(x) = r(y) = 2$, but $r(x \vee y) = r(\hat{\mathbf{1}}) = 3$ and $r(x \wedge y) = r(\hat{\mathbf{0}}) = 0$. So x is not a modular element.

Proposition 5.5.2. *The modular elements of Π_n are exactly the partitions with at most one nonsingleton block.*

Proof. Suppose that $\pi \in \Pi_n$ has one nonsingleton block B. For $\sigma \in \Pi_n$, let

$$X = \{C \in \sigma : C \cap B \neq \emptyset\}, \qquad Y = \{C \in \sigma : C \cap B = \emptyset\}.$$

Then

$$\pi \wedge \sigma = \Big\{C \cap B : C \in X\Big\} \cup \Big\{\{i\} : i \notin B\Big\}, \qquad \pi \vee \sigma = \left\{\bigcup_{C \in X} C\right\} \cup Y$$

so

$$\begin{aligned} |\pi \wedge \sigma| + |\pi \vee \sigma| &= (|X| + n - |B|) + (1 + |Y|) \\ &= (n - |B| + 1) + (|X| + |Y|) = |\pi| + |\sigma|, \end{aligned}$$

proving that π is a modular element.

For the converse, suppose B, C are nonsingleton blocks of π, with $i, j \in B$ and $k, \ell \in C$. Let σ be the partition with exactly two nonsingleton blocks $\{i,k\}, \{j,\ell\}$. Then $r(\sigma) = 2$ and $r(\pi \wedge \sigma) = r(\hat{\mathbf{0}}) = 0$, but

$$r(\pi \vee \sigma) = r(\pi) + 1 < r(\pi) + r(\sigma) - r(\pi \wedge \sigma)$$

so π is not a modular element. □

Modular elements are useful because they lead to factorizations of the characteristic polynomial of L.

Theorem 5.5.3. *Let L be a geometric lattice of rank n, and let $z \in L$ be a modular element. Then*

$$\chi_L(k) = \chi_{[\hat{\mathbf{0}},z]}(k) \sum_{y:\ y \wedge z = \hat{\mathbf{0}}} \mu_L(\hat{\mathbf{0}}, y) k^{n - r(z) - r(y)}. \tag{5.11}$$

Here is a sketch of the proof; for the full details, see [Sta07, pp. 440–441]. We work in the dual Möbius algebra $A^*(L) = A(L^*)$; that is, the vector space of $\mathbb{C}$-linear combinations of elements of L, with multiplication given by join (rather than meet as in §2.4). Thus the "algebraic" basis of $A^*(L)$ is

$$\{\sigma_y \stackrel{\text{def}}{=} \sum_{x:\, x \geq y} \mu(y,x)x : y \in L\}.$$

First, show by direct calculation that

$$\sigma_{\hat{\mathbf{0}}} \;=\; \sum_{x \in L} \mu(x)x \;=\; \left(\sum_{v:\, v \leq z} \mu(v)v \right) \left(\sum_{y:\, y \wedge z = \hat{\mathbf{0}}} \mu(y)y \right) \tag{5.12}$$

for any $z \in L$. Second, for $z, y, v \in L$ such that z is modular, $v \leq z$, and $y \wedge z = 0$, one shows first that $z \wedge (v \vee y) = v$ (by rank considerations) and then that $\operatorname{rank}(v \vee y) = \operatorname{rank}(v) + \operatorname{rank}(y)$. Third, make the substitutions $v \mapsto k^{\operatorname{rank} z - \operatorname{rank} v}$ and $y \mapsto k^{n - \operatorname{rank} y - \operatorname{rank} z}$ in the two sums on the RHS of (5.12). Since $vy = v \vee y$, the last observation implies that substituting $x \mapsto k^{n - \operatorname{rank} x}$ on the LHS preserves the product, and the equation becomes (5.11).

In particular, every atom a is modular, so

$$\chi_L(k) \;=\; (k-1) \sum_{x:\, x \not\geq a} \mu_L(\hat{\mathbf{0}}, x) k^{r(L) - 1 - r(x)}.$$

This does not really tell us anything new, because we already knew that $k-1$ had to be a factor of $\chi_L(k)$, because $\chi_L(1) = \sum_{x \in L} \mu_L(\hat{\mathbf{0}}, x) = 0$. Also, the sum in the expression is not the characteristic polynomial of a lattice.

On the other hand, if we have a modular *coatom*, then Theorem 5.5.3 is much more useful, since we can identify an interesting linear factor and describe what is left after factoring it out.

Corollary 5.5.4. *Let L be a geometric lattice, and let $z \in L$ be a coatom that is a modular element. Then*

$$\chi_L(k) \;=\; (k - e)\chi_{[\hat{\mathbf{0}}, z]}(k),$$

where e is the number of atoms $a \in L$ such that $a \not\leq z$.

If we are extremely lucky, then L will have a saturated chain of modular elements

$$\hat{\mathbf{0}} = x_0 \lessdot x_1 \lessdot \cdots \lessdot x_{n-1} \lessdot x_n = \hat{\mathbf{1}}.$$

In this case, we can apply Corollary 5.5.4 successively with $z = x_{n-1}$, $z = x_{n-2}$, …, $z = x_1$ to split the characteristic polynomial completely into linear factors:

$$\begin{aligned} \chi_L(k) \;&=\; (k - e_{n-1})\chi_{[\hat{\mathbf{0}}, x_{n-1}]}(k) \\ &=\; (k - e_{n-1})(k - e_{n-2})\chi_{[\hat{\mathbf{0}}, x_{n-2}]}(k) \\ &=\; \cdots \\ &=\; (k - e_{n-1})(k - e_{n-2}) \cdots (k - e_0), \end{aligned}$$

where

$$\begin{aligned} e_i \;&=\; \#\{\text{atoms } a \text{ of } [\hat{\mathbf{0}}, x_{i+1}] : a \not\leq x_i\} \\ &=\; \#\{a \in A : a \leq x_{i+1},\ a \not\leq x_i\}. \end{aligned}$$

Definition 5.5.5. A geometric lattice L is **supersolvable** if it has a **modular chain**, that is, a maximal chain $\hat{\mathbf{0}} = x_0 \lessdot x_1 \lessdot \cdots \lessdot x_n = \hat{\mathbf{1}}$ such that every x_i is a modular element. A central hyperplane arrangement $\mathcal{A}$ is called supersolvable if $L(\mathcal{A})$ is supersolvable.

Example 5.5.6. Every modular lattice is supersolvable, because every maximal chain is modular. In particular, the characteristic polynomial of every modular lattice splits into linear factors. ◀

Example 5.5.7. The partition lattice Π_n (and therefore the associated hyperplane arrangement Br_n) is supersolvable by induction. Let z be the coatom with blocks $[n-1]$ and $\{n\}$, which is a modular element by Proposition 5.5.2. There are $n-1$ atoms $a \not\leq z$, namely the partitions whose non-singleton block is $\{i, n\}$ for some $i \in [n-1]$, so we obtain

$$\chi_{\Pi_n}(k) \;=\; (k-n+1)\chi_{\Pi_{n-1}}(k)$$

and by induction

$$\chi_{\Pi_n}(k) \;=\; (k-1)(k-2)\cdots(k-n+1).$$

◀

Example 5.5.8. Let $G = C_4$ (a cycle with four vertices and four edges), and let $\mathcal{A} = \mathcal{A}_G$. Then $L(\mathcal{A})$ is the lattice of flats of the matroid $U_3(4)$; i.e.,

$$L = \{F \subseteq [4] :\; |F| \neq 3\}$$

with $r(F) = \min(|F|, 3)$. This lattice is not supersolvable, because no element at rank 2 is modular. For example, let $x = 12$ and $y = 34$; then $r(x) = r(y) = 2$ but $r(x \vee y) = 3$ and $r(x \wedge y) = 0$. (We have already seen that the characteristic polynomial of L does not split.) ◀

Theorem 5.5.9. *Let $G = (V, E)$ be a simple graph. Then $\mathcal{A}_G$ is supersolvable if and only if the vertices of G can be ordered $v_1, \ldots, v_n$ such that for every $i > 1$, the set*

$$C_i \;:=\; \{v_j : j \leq i,\; v_i v_j \in E\}$$

forms a clique in G.

Such an ordering is called a **perfect elimination ordering**. The proof of Theorem 5.5.9 is left as an exercise (see Stanley, pp. 55–57). An equivalent condition is that G is a **chordal graph**: if $C \subseteq G$ is a cycle of length ≥ 4, then some pair of vertices that are not adjacent in C are in fact adjacent in G. This equivalence is sometimes known as *Dirac's theorem*. It is fairly easy to prove that supersolvable graphs are chordal, but the converse is somewhat harder; see, e.g., [Wes96, pp. 224–226]. There are other graph-theoretic formulations of this property; see, e.g., [Dir61]. See the recent paper [HS15] for much more about factoring the characteristic polynomial of lattices in general.

If G satisfies the condition of Theorem 5.5.9, then we can see directly why its chromatic polynomial $\chi(G;\, k)$ splits into linear factors. Consider what happens when we color the vertices in order. When we color vertex v_i, it has $|C_i|$ neighbors that have already been colored, and they all have received different colors because they form a clique. Therefore, there are $k - |C_i|$ possible colors available for v_i, and we see that

$$\chi(G;\, k) \;=\; \prod_{i=1}^{n} (k - |C_i|).$$

5.6 Beyond real hyperplane arrangements

One can also study *complex* hyperplane arrangements $\mathcal{A} \subseteq \mathbb{C}^n$. Since the hyperplanes of $\mathcal{A}$ have codimension 2 as real vector subspaces, the complement $X = \mathbb{C}^n \setminus \mathcal{A}$ is a connected topological space, but not simply

connected. Thus instead of counting regions, we should count holes, as expressed by the homology groups. Brieskorn [Bri73] solved this problem completely:

Theorem 5.6.1 (Brieskorn [Bri73])**.** *The homology groups $H_i(X, \mathbb{Z})$ are free abelian, and the Poincáre polynomial of X is the characteristic polynomial backwards:*

$$\sum_{i=0}^{n} \operatorname{rank}_{\mathbb{Z}} H_i(X, \mathbb{Z}) q^i \;=\; (-q)^n \chi_{L(\mathcal{A})}(-1/q).$$

In a very famous paper, Orlik and Solomon [OS80] strengthened Brieskorn's result by giving a presentation of the cohomology ring $H^*(X, \mathbb{Z})$ in terms of $L(\mathcal{A})$, thereby proving that the cohomology is a combinatorial invariant of $\mathcal{A}$. (Brieskorn's theorem says only that the *additive* structure of $H^*(X, \mathbb{Z})$ is a combinatorial invariant.) By the way, the homotopy type of X is *not* a combinatorial invariant; Rybnikov [Ryb11] constructed arrangements with isomorphic lattices of flats but different fundamental groups. There is much more to say on this topic!

In another direction, one can study arrangements of subspaces of $\mathbb{R}^n$ or $\mathbb{C}^n$ that are not hyperplanes, i.e., have codimension greater than 1. This topic is much more difficult, in particular because one does not have the nice combinatorial model of matroid theory in the background. A starting point is Björner's survey article [Bjö94].

5.7 Faces and the big face lattice

Consider the two arrangements $\mathcal{A}_1, \mathcal{A}_2 \subset \mathbb{R}^2$ shown in Figure 5.5. Their intersection posets are isomorphic, so, by Zaslavsky's theorems they have the same numbers of regions and bounded regions (this can of course be checked directly). However, there is good reason *not* to consider the two arrangements isomorphic. For example, both bounded regions in $\mathcal{A}_1$ are triangles, while $\mathcal{A}_2$ has a triangle and a trapezoid. Also, the point $H_1 \cap H_2 \cap H_4$ lies *between* the lines H_3 and H_5 in $\mathcal{A}_1$, while it lies *below* both of them in $\mathcal{A}_2$. The intersection poset lacks the power to model geometric data like "between," "below," "triangle" and "trapezoid." Accordingly, we need to define a stronger combinatorial invariant.

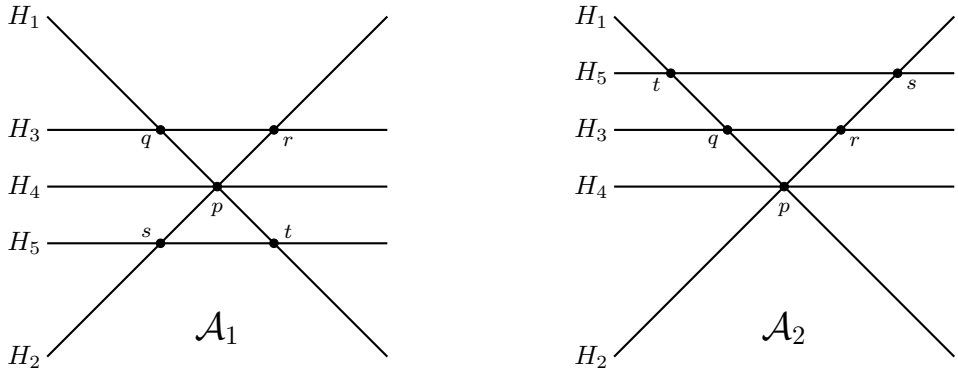


Figure 5.5: Two different arrangements with isomorphic intersection posets.

First we fix notation. Let $\mathcal{A} = \{H_1, \ldots, H_n\}$ be an essential hyperplane arrangement in $\mathbb{R}^d$, with normal vectors $\mathbf{n}_1, \ldots, \mathbf{n}_n$. For each i, let λ_i be an affine linear functional on $\mathbb{R}^n$ such that $H_i = \{\mathbf{x} \in \mathbb{R}^d \colon \lambda_i(\mathbf{x}) = 0\}$. (If $\cap\mathcal{A} = \{\vec{\mathbf{0}}\}$ then we may define $\lambda_i(\mathbf{x}) = \mathbf{n}_i \cdot \mathbf{x}$.)

The intersections of hyperplanes in $\mathcal{A}$, together with its regions, decompose $\mathbb{R}^d$ as a **polyhedral cell complex**: a disjoint union of polyhedra, each homeomorphic to $\mathbb{R}^e$ for some $e \leq d$ (that's what "cell" means), such that the boundary of any cell is a union of other cells. We can encode each cell by recording whether the linear functionals $\lambda_1, \ldots, \lambda_n$ are positive, negative or zero on it. Specifically, for $\mathbf{k} = (k_1, \ldots, k_n) \in \{+, -, 0\}^n$, define a (possibly empty) subset of $\mathbb{R}^d$ by

$$F = F(\mathbf{k}) = \left\{ \mathbf{x} \in \mathbb{R}^d \;\middle|\; \begin{array}{lll} \lambda_i(\mathbf{x}) > 0 & \text{if } k_i = + & \iff \quad i \in \mathbf{k}_+ \\ \lambda_i(\mathbf{x}) < 0 & \text{if } k_i = - & \iff \quad i \in \mathbf{k}_- \\ \lambda_i(\mathbf{x}) = 0 & \text{if } k_i = 0 & \iff \quad i \in \mathbf{k}_0 \end{array} \right\}.$$

This formula can be taken as the definition of $\mathbf{k}_+$, $\mathbf{k}_-$, and $\mathbf{k}_0$. A convenient shorthand ("digital notation") is to represent $\mathbf{k}$ by the list of digits i for which $k_i \neq 0$, placing a bar over the digits for which $k_i < 0$. For instance, $\mathbf{k} = 0 + -00 - +0$ would be abbreviated $2\bar{3}\bar{6}7$; here $\mathbf{k}_+ = \{2, 7\}$ and $\mathbf{k}_- = \{3, 6\}$.

If $F \neq \emptyset$ then it is called a **face** of $\mathcal{A}$, and $\mathbf{k} = \mathbf{k}(F)$ is the corresponding **covector**. The set of all faces is denoted $\mathscr{F}(\mathcal{A})$. The poset $\hat{\mathscr{F}}(\mathcal{A}) = \mathscr{F}(\mathcal{A}) \cup \{\hat{\mathbf{0}}, \hat{\mathbf{1}}\}$, ordered by containment of closures ($F \leq F'$ if $\bar{F} \subseteq \bar{F}'$), is a lattice, called the **(big) face lattice**[1] of $\mathcal{A}$. If $\mathcal{A}$ is central, then $\mathscr{F}(\mathcal{A})$ already has a unique minimal element and we don't add an extra one. For example, the big face lattice of Bool_2 is shown in Figure 5.6.

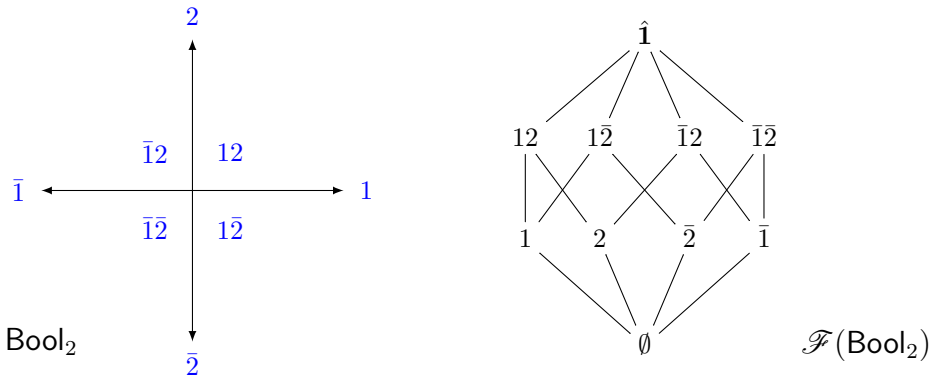


Figure 5.6: The Boolean arrangement Bool_2 and its big face lattice.

Combinatorially, the order relation in $\mathscr{F}(\mathcal{A})$ is given by $\mathbf{k} \leq \mathbf{l}$ if $\mathbf{k}_+ \subseteq \mathbf{l}_+$ and $\mathbf{k}_- \subseteq \mathbf{l}_-$. (This is very easy to read off using digital notation.) The maximal covectors (or **topes**) are precisely those with no zeroes; they correspond to the regions of $\mathcal{A}$.

The big face lattice captures more of the geometry of $\mathcal{A}$ than the intersection poset; for instance, the two arrangements $\mathcal{A}_1, \mathcal{A}_2$ shown above have isomorphic intersection posets but non-isomorphic face lattices. (This may be clear to you now; there are lots of possible explanations and we will see one soon.)

Example 5.7.1. An especially important example is the braid arrangement Br_n (see Example 5.1.4), whose faces have an explicit combinatorial description in terms of set compositions. If F is a face, then F lies either below, above, or on each hyperplane H_{ij} — i.e., either $x_i < x_j$, $x_i = x_j$, or $x_i > x_j$ holds on F — and this data describes F exactly. In fact, we can record F by a **set composition** of $[n]$, i.e., an ordered list A of nonempty sets $A_1 | \ldots | A_k$ whose disjoint union is $[n]$. (We write $A \models [n]$ for short.) For example, the set composition

$$569 \,|\, 3 \,|\, 14 \,|\, 28 \,|\, 7$$

[1] That is, the big lattice of faces, not the lattice of big faces.

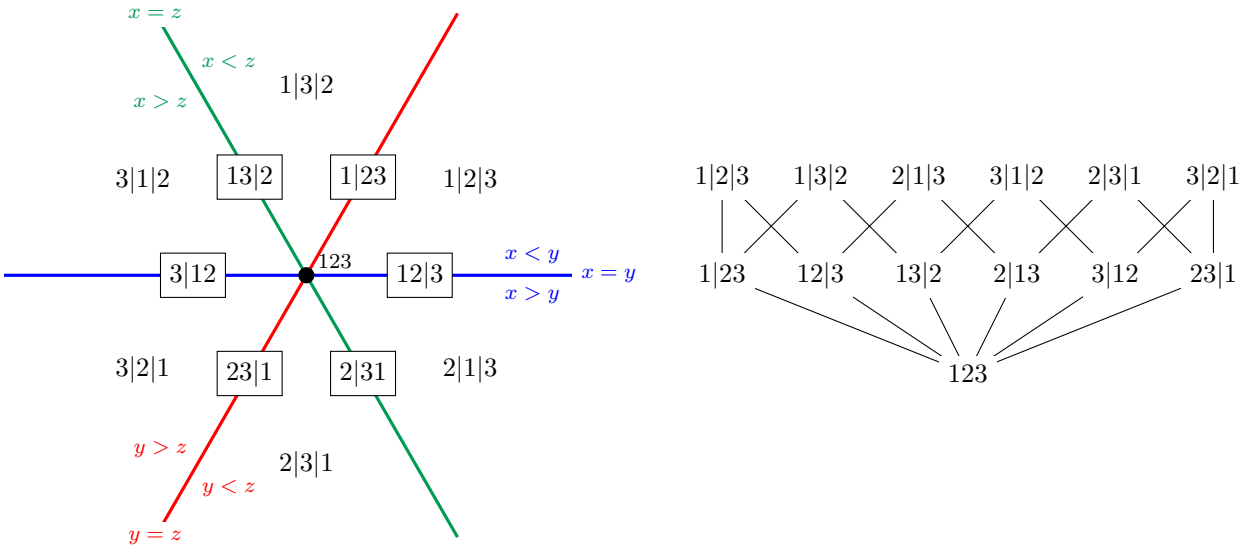


Figure 5.7: Br_3 and its big face lattice (the lattice of set compositions).

represents the face F_A of Br_9 defined by the equations and inequalities

$$x_5 = x_6 = x_9 \;<\; x_3 \;<\; x_1 = x_4 \;<\; x_2 = x_8 \;<\; x_7.$$

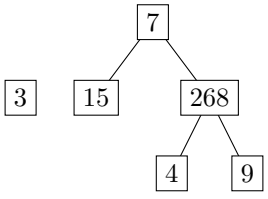
Note that the number of blocks of A (in this case, 5) equals $\dim F_A$, since that is the number of free coordinates. In fact, F_A is linearly equivalent to a maximal region of Br_5, say the principal region, under the linear transformation $\mathbb{R}^5 \to \mathbb{R}^9$ given by $(a, b, c, d, e) \mapsto (c, d, b, c, a, a, e, d, a)$; in particular it is a simplicial polyhedron. In the extreme case that $\dim A = n$, the set composition has only singleton parts, hence is equivalent to a permutation (this confirms what we already know, that Br_n has $n!$ regions).

The correspondence between faces of Br_n and set compositions $A \models [n]$ is a bijection. In fact, the big face lattice of Br_n is isomorphic to the lattice of set compositions ordered by refinement; see Figure 5.7. ◀

More generally, consider a system of linear equalities and inequalities of the form $x_i = x_j$ and $x_i < x_j$. If such a system is consistent, it gives rise to a nonempty polyhedron that is a convex union of faces of Br_9. Such a system can be described by a **preposet**, which is a relation $<$ on $[n]$ that is reflexive, transitive, but not necessarily antisymmetric (compare Defn. 1.1.1). In other words, $x \leq y$ and $y \leq x$ does not imply $x = y$. This relation has a Hasse diagram, just like a poset, except that multiple elements of the ground set can be put in the same "box" (whenever there is a failure of antisymmetry). For example, the system

$$x_1 = x_5,\ x_4 < x_6,\ x_5 < x_7,\ x_6 = x_8,\ x_2 = x_6,\ x_9 < x_8,\ x_2 < x_7.$$

corresponds to the preposet whose Hasse diagram is



and this gives rise to a 6-dimensional convex polyhedron P consisting of faces of Br_9. (Each box in the Hasse diagram represents a coordinate that can vary (locally) freely, which is why the dimension is 6.) The maximal faces in P correspond to the linear extensions of the preposet, expressed as set compositions: $15|3|4|9|268|7$, $4|9|268|3|15|7$, etc. For more on the "cone/preposet dictionary", see [PRW08].

5.8 Oriented matroids

Oriented matroids are a vast topic; these notes just scratch the surface. The canonical resource is the book [BLVS+99]; an excellent free source is Reiner's lecture notes [Rei] and another good brief reference is [RGZ97].

5.8.1 Oriented matroid covectors from hyperplane arrangements

Consider the linear forms λ_i that were used in representing each face by a covector. Recall that specifying λ_i is equivalent to specifying a normal vector $\mathbf{n}_i$ to the hyperplane H_i (with $\lambda_i(\mathbf{x}) = \mathbf{n}_i \cdot \mathbf{x}$). As we know, the vectors $\mathbf{n}_i$ represent a matroid whose lattice of flats is precisely $L(\mathcal{A})$. Scaling $\mathbf{n}_i$ (equivalently, λ_i) by a nonzero constant $c \in \mathbb{R}$ has no effect on the matroid represented by the $\mathbf{n}_i$'s, but what does it do to the covectors? If $c > 0$, then nothing happens, but if $c < 0$, then we have to switch $+$ and $-$ signs in the ith position of every covector. So, in order to figure out the covectors, we need not just the normal vectors $\mathbf{n}_i$, but an **orientation** for each one — hence the term "oriented matroid". Equivalently, for each hyperplane H_i, we are designating one of the two corresponding halfspaces (i.e., connected components of $\mathbb{R}^d \setminus H_i$) as positive and the other as negative.

See Figure 5.8 for examples. (The normal vectors all have positive z-coordinate, so "above" means "above.") For instance, the trapezoidal bounded region in $\mathcal{A}_2$ has covector $++++-$ because it lies above hyperplanes H_1, H_2, H_3, H_4 but below H_5. Its top side has covector $++++0$, its bottom $++0+-$, etc.

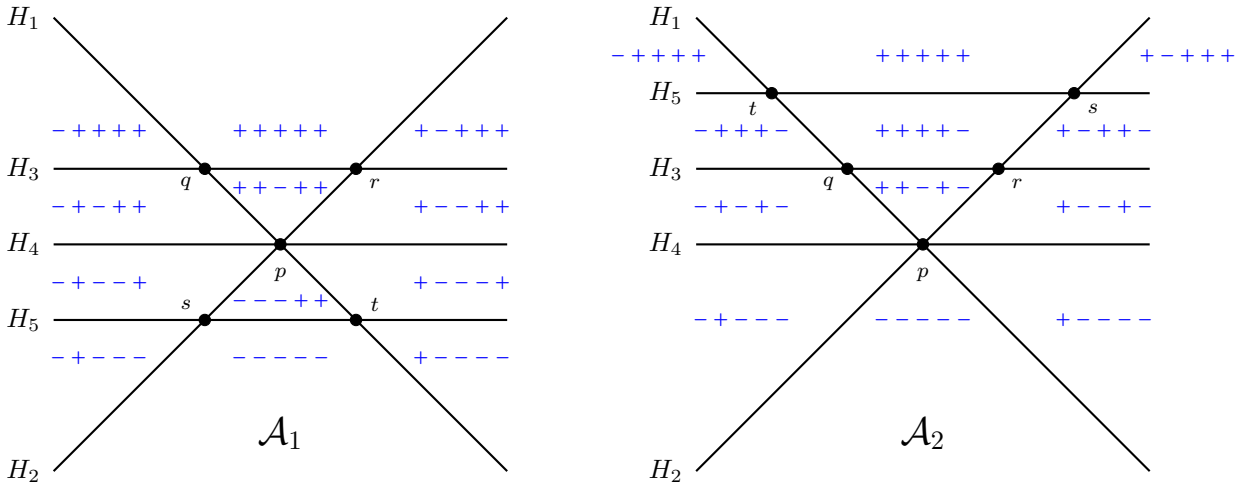


Figure 5.8: Covectors correspond to faces of a hyperplane arrangement.

Proposition 5.8.1. *Suppose that no two hyperplanes in $\mathcal{A}$ are parallel. Then the maximal covectors whose negatives are also covectors are precisely those that correspond to relatively-unbounded faces. In particular, $\mathcal{A}$ is central if and*

only if every negative of a covector is a covector.

Proof. Assume without loss of generality that $\mathcal{A}$ is essential.

Suppose R is an unbounded region, with $\mathbf{k}$ the corresponding covector. Fix a point $\mathbf{x} \in R$ and choose a direction $\mathbf{v}$ in which R is unbounded. By perturbing $\mathbf{v}$ slightly, we can assume that $\mathbf{v}$ is not orthogonal to any normal vector $\mathbf{n}_i$ for which $\mathbf{k}_i = 0$. (This perturbation step is where we use the assumption that no two hyperplanes are parallel.) In other words, if we walk in the direction of $\mathbf{v}$ then the values of λ_i increase without bound, decrease without bound, or remain zero according as i belongs to $\mathbf{k}_+$, $\mathbf{k}_-$, or $\mathbf{k}_0$. But then if we walk in the direction of $-\mathbf{n}_i$, then "increase" and "decrease" are reversed. Therefore, walking sufficiently far in that direction arrives in an (unbounded) region with covector $-\mathbf{k}$.

Conversely, suppose that $\mathbf{k}$ and $-\mathbf{k}$ are covectors of regions R and S. Pick points $\mathbf{x} \in R$ and $\mathbf{y} \in S$ and consider the line ℓ joining $\mathbf{x}$ and $\mathbf{y}$. The functionals λ_i are identically zero on ℓ for $i \in \mathbf{k}_0 = (-\mathbf{k})_0$, but otherwise increase or decrease (necessarily without bound). Therefore the ray pointing from $\mathbf{x}$ away from $\mathbf{y}$ (resp., from $\mathbf{y}$ away from $\mathbf{x}$) is contained in R (resp., S). It follows that both R and S are unbounded.

The second assertion now follows from Corollary 5.3.10. **WHICH IS FALSE** □

(It would be nice to modify the statement to handle the case that $\mathcal{A}$ has parallel hyperplanes. Here the conclusion fails, since for example in $\mathcal{A}_1$ or $\mathcal{A}_2$ above, every ray in the region with covector $+---+$ is horizontal, hence orthogonal to the normals to $H_3, H_4.H_5$, so the functionals λ_3, λ_4 are constant and positive — hence do not become negative upon walking in the other direction; the "opposite" unbounded region has covector $-+--+$. It is still true that any pair of opposite covectors correspond to opposite unbounded regions, but I think this condition holds only for unbounded regions that contain more than one direction's worth of rays.)

Just like circuits, bases, etc., of a matroid, oriented matroid covectors can be axiomatized purely combinatorially. First some preliminaries. For $\mathbf{k}, \mathbf{l} \in \{+, 0, -\}^n$, define the **composition** $\mathbf{k} \circ \mathbf{l}$ by

$$(\mathbf{k} \circ \mathbf{l})_i = \begin{cases} k_i & \text{if } k_i \neq 0, \\ l_i & \text{if } k_i = 0. \end{cases}$$

This somewhat odd-looking definition models something geometric: if $\mathbf{k}, \mathbf{l}$ correspond to faces R, R' of a hyperplane arrangement then $\mathbf{k} \circ \mathbf{l}$ corresponds to the face you are in if you start at a point in R and take one very small step towards R'. Also, define $S(\mathbf{k}, \mathbf{l}) = \{i \in [n] : \ k_i = -l_i \neq 0\}$; this corresponds to the set of (pseudo)hyperplanes separating R and R'.

The axioms are as follows [RGZ97, §7.2.1]: a collection $\mathscr{K} \subseteq \{+, -, 0\}^n$ is a covector system if for all $\mathbf{k}, \mathbf{l} \in \mathscr{K}$:

(K1) $\vec{\mathbf{0}} = (0, 0, \ldots, 0) \in \mathscr{K}$;
(K2) $-\mathbf{k} \in \mathscr{K}$;
(K3) $\mathbf{k} \circ \mathbf{l} \in \mathscr{K}$;
(K4) If $i \in S(\mathbf{k}, \mathbf{l})$ then there exists $\mathbf{m} \in \mathscr{K}$ with (a) $m_i = 0$ and (b) $m_j = (\mathbf{k} \circ \mathbf{l})_j$ for $j \in [n] \setminus S(\mathbf{k}, \mathbf{l})$.

Note that **(K1)** and **(K2)** are really properties of *central* hyperplane arrangements. However, any non-central arrangement $\mathcal{A}$ can be turned into a central one by coning (see Definition 5.1.8), and if $\mathscr{K}(\mathcal{A})$ is the set of covectors of $\mathcal{A}$ then

$$\mathscr{K}(c\mathcal{A}) = \{(\mathbf{k}, +) \colon \mathbf{k} \in \mathscr{K}(\mathcal{A})\} \cup \{(-\mathbf{k}, -) : \ \mathbf{k} \in \mathscr{K}(\mathcal{A})\} \cup \{\vec{\mathbf{0}}\}$$

and by the way, $\mathscr{K}(\mathcal{A}) = \{\mathbf{k} \colon (\mathbf{k}, +) \in \mathscr{K}(c\mathcal{A})\}$.

5.8.2 Oriented matroid circuits

The cones over the arrangements $\mathcal{A}_1$ and $\mathcal{A}_2$ (not including the new hyperplane introduced in coning) are central, essential arrangements in $\mathbb{R}^3$, whose matroids of normals can be represented respectively by the matrices

$$X_1 = \begin{array}{c} \begin{matrix} \mathbf{n}_1 & \mathbf{n}_2 & \mathbf{n}_3 & \mathbf{n}_4 & \mathbf{n}_5 \end{matrix} \\ \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & -1 \end{bmatrix} \end{array}, \qquad X_2 = \begin{array}{c} \begin{matrix} \mathbf{n}'_1 & \mathbf{n}'_2 & \mathbf{n}'_3 & \mathbf{n}'_4 & \mathbf{n}'_5 \end{matrix} \\ \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 2 \end{bmatrix} \end{array}$$

Evidently the matroids represented by X_1 and X_2 are isomorphic, with circuit system $\{124, 345, 1235\}$. However, they are not isomorphic as oriented matroids. The minimal linear dependencies realizing the circuits in each case are

$$\begin{aligned} \mathbf{n}_1 + \mathbf{n}_2 - 2\mathbf{n}_4 &= 0 \\ \mathbf{n}_3 - 2\mathbf{n}_4 + \mathbf{n}_5 &= 0 \\ \mathbf{n}_1 + \mathbf{n}_2 - \mathbf{n}_3 - \mathbf{n}_5 &= 0 \end{aligned} \qquad \begin{aligned} \mathbf{n}'_1 + \mathbf{n}'_2 - 2\mathbf{n}'_4 &= 0 \\ 2\mathbf{n}'_3 - \mathbf{n}'_4 - \mathbf{n}'_5 &= 0 \\ \mathbf{n}'_1 + \mathbf{n}'_2 - 4\mathbf{n}'_3 + 2\mathbf{n}'_5 &= 0 \end{aligned}$$

An **oriented circuit** keeps track not just of minimal linear dependencies, but of how to orient the vectors in the circuit so that all the signs are positive. Thus $12\bar{4}$ is a oriented circuit in both cases. However, in the first case $3\bar{4}5$ is a circuit, while in the second it is $3\bar{4}\bar{5}$. Note that if $\mathbf{c}$ is a circuit then so is $-\mathbf{c}$, where, e.g., $-12\bar{4} = \bar{1}\bar{2}4$. In summary, the oriented circuit systems for $C\mathcal{A}_1$ and $C\mathcal{A}_2$ are respectively

$$\begin{aligned} \vec{\mathscr{C}_1} &= \{12\bar{4},\ \bar{1}\bar{2}4,\ 3\bar{4}5,\ \bar{3}4\bar{5},\ 12\bar{3}\bar{5},\ \bar{1}\bar{2}35\}, \\ \vec{\mathscr{C}_2} &= \{12\bar{4},\ \bar{1}\bar{2}4,\ 3\bar{4}\bar{5},\ \bar{3}45,\ 12\bar{3}5,\ \bar{1}\bar{2}3\bar{5}\}. \end{aligned}$$

Oriented circuits are minimal obstructions to covector-ness. For example, $12\bar{4}$ is a circuit of $\mathcal{A}_1$ because the linear functionals defining its hyperplanes satisfy $\lambda_1 + \lambda_2 - 2\lambda_4 = 0$. But if a covector of $\mathcal{A}_1$ contains $12\bar{4}$, then any point in the corresponding face of $\mathcal{A}$ would have $\lambda_1, \lambda_2, -\lambda_4$ all positive, which is impossible.

Oriented circuits can be axiomatized too:

Definition 5.8.2. Let n be a positive integer. A **oriented circuit system** is a collection $\vec{\mathscr{C}}$ of n-tuples $\mathbf{c} \in \{+, -, 0\}^n$ satisfying the following properties for all $\mathbf{c}, \mathbf{c}' \in \mathscr{C}$ with $\mathbf{c} \neq \mathbf{c}'$:

(OC1) $\vec{\mathbf{0}} \notin \vec{\mathscr{C}}$.
(OC2) $-\mathbf{c} \in \vec{\mathscr{C}}$.
(OC3) Either $\mathbf{c}_+ \not\subseteq \mathbf{c}'_+$ or $\mathbf{c}_- \not\subseteq \mathbf{c}'_-$.
(OC4) If $c_i = +$ and $c'_i = -$, then there exists $\mathbf{d} \in \vec{\mathscr{C}}$ such that (a) $d_i = 0$ and (b) for all $j \neq i$, $\mathbf{d}_+ \subseteq \mathbf{c}_+ \cup \mathbf{c}'_+$ and $\mathbf{d}_- \subseteq \mathbf{c}_- \cup \mathbf{c}'_-$.

Again, the idea is to record not just the linearly dependent subsets of a set $\{\lambda_i, \ldots, \lambda_n\}$ of linear forms, but also the sign patterns of the corresponding linear dependences ("syzygies"). The first two are elementary: **(OC1)** says that the empty set is linearly independent and **(OC2)** says that multiplying any syzygy by -1 gives a syzygy. Condition **(OC3)** must hold if we want circuits to record signed syzygies with minimal support, as for circuits in an unoriented matroid,

(OC4) is the oriented version of circuit exchange. Suppose that we have two syzygies

$$\sum_{j=1}^{n} \gamma_j \lambda_j \;=\; \sum_{j=1}^{n} \gamma'_j \lambda_j \;=\; 0$$

with $\gamma_i > 0$ and $\gamma'_i < 0$ for some i. Multiplying by positive scalars if necessary (hence not changing the sign patterns), we may assume that $\gamma_i = -\gamma'_i$. Then adding the two syzygies gives

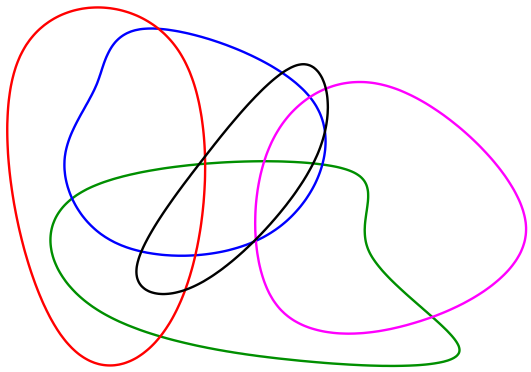
$$\sum_{j=1}^{n} \delta_j \lambda_j \;=\; 0,$$

where $\delta_j = \gamma_j + \gamma'_j$. In particular, $\delta_i = 0$, and δ_j is positive (resp., negative) if and only if at least one of γ_j, γ'_j is positive (resp., negative).

Remark 5.8.3. If $\vec{\mathscr{C}}$ is an oriented circuit system, then $\mathscr{C} = \{\mathbf{c}_+ \cup \mathbf{c}_- : \mathbf{c} \in \vec{\mathscr{C}}\}$ is a circuit system for an ordinary matroid with ground set $[n]$. (I.e., just erase all the bars.) This is called the **underlying matroid** of the oriented matroid with circuit system $\vec{\mathscr{C}}$.

As in the unoriented setting, the circuits of an oriented matroid represent minimal obstructions to being a covector. That is, every real hyperplane arrangement $\mathcal{A}$ gives rise to an oriented circuit system $\vec{\mathscr{C}}$ such that if $\mathbf{k}$ is a covector of $\mathcal{A}$ and $\mathbf{c}$ is a circuit, then it is not the case that $\mathbf{k}_+ \supseteq \mathbf{c}_+$ and $\mathbf{k}_- \supseteq \mathbf{c}_-$.

More generally, one can construct an oriented matroid from any real *pseudosphere arrangement*, or collection of homotopy $(d-1)$-spheres embedded in $\mathbb{R}^n$ such that the intersection of the closures of the spheres in any subcollection is either connected or empty — i.e., a thing like this:

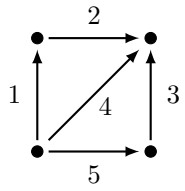


Again this arrangement gives rise to a cellular decomposition of $\mathbb{R}^n$, and each cell corresponds to a covector which describes whether the cell is inside, outside, or on each pseudocircle.

In fact, the Topological Representation Theorem of Folkman and Lawrence (1978) says that *every* combinatorial oriented matroid can be represented by such a pseudosphere arrangement. However, there exist oriented matroids that cannot be represented as hyperplane arrangements. For example, recall the construction of the non-Pappus matroid (Example 3.5.7). If we bend the line xyz a little so that it meets x and y but not z (and no other points), the result is a pseudoline arrangement whose oriented matroid $\mathcal{M}$ cannot be represented by means of a line arrangement.

5.8.3 Oriented matroids from graphs

Recall (§3.3) that every graph $G = (V, E)$ gives rise to a graphic matroid $M(G)$ with ground set E. Correspondingly, every *directed* graph $\vec{G}$ gives rise to an *oriented* matroid, whose circuit system $\vec{\mathscr{C}}$ is the family of oriented cycles. This is best shown by an example.



Oriented circuits

$12\bar{4}$ $3\bar{4}5$ $12\bar{3}\bar{5}$
$\bar{1}\bar{2}4$ $\bar{3}4\bar{5}$ $\bar{1}\bar{2}35$

For example, $13\bar{5}$ is a circuit because the clockwise orientation of the northwest triangle in G includes edges 1 and 3 forward, and edge 5 backward. In fact, this circuit system is identical to the circuit system $\mathscr{C}_1$ seen previously. More generally, for every oriented graph $\vec{G}$, the signed set system $\vec{\mathscr{C}}$ formed in this way satisfies the axioms of Definition 5.8.2. To understand axiom (4) of that definition, suppose e is an edge that occurs forward in $\mathbf{c}$ and backward in $\mathbf{c}'$. Then $\mathbf{c} - e$ and $\mathbf{c}' - e$ are paths between the two endpoints of e, with opposite starting and ending points, so when concatenated, they form an closed walk in $\vec{G}$, which must contain an oriented cycle.

Reversing the orientation of edge e corresponds to interchanging e and $\bar{e}$ in the circuit system; this is called a **reorientation**. For example, reversing edge 5 produces the previously seen oriented circuit system $\mathscr{C}_2$.

An oriented matroid is called **acyclic** if every circuit has at least one barred and at least one unbarred element; this is equivalent to $\vec{G}$ having no directed cycles (i.e., being an acyclic orientation of its underlying graph G). In fact, for any ordinary unoriented matroid M, one can define an **orientation** of M as an oriented matroid whose underlying matroid is M; the number of acyclic orientations is $T_M(2, 0)$ [Rei, §3.1.6, p.29], just as for graphs.

The covectors of the circuit system for a directed graph are in fact the faces of the (essentialization of) the graphic arrangement associated to $\vec{G}$, in which the orientation of each edge determines the orientation of the corresponding normal vector — if $\vec{ij}$ is an edge in $\vec{G}$ then the hyperplane $x_i = x_j$ is assigned the normal vector $\mathbf{e}_i - \mathbf{e}_j$. The maximal covectors are precisely the regions of the graphic arrangement.

5.9 Exercises

Problem 5.1. Let $m > n$, and let $\mathcal{A}$ be the arrangement of m affine hyperplanes in general position in $\mathbb{R}^n$. Here "general position" means that every k of the hyperplanes intersect in an affine linear space of dimension $n - k$; if $k > n$ then the intersection is empty. (Compare Example 5.3.9, where the hyperplanes are linear.) Calculate $\chi_{\mathcal{A}}(k)$, $r(\mathcal{A})$, and $b(\mathcal{A})$.

Problem 5.2. (Stanley, HA, 2.5) Let G be a graph on n vertices, let $\mathcal{A}_G$ be its graphic arrangement in $\mathbb{R}^n$, and let $\mathcal{B}_G = \mathsf{Bool}_n \cup \mathcal{A}_G$. (That is, $\mathcal{B}$ consists of the coordinate hyperplanes $x_i = 0$ in $\mathbb{R}^n$ together with the hyperplanes $x_i = x_j$ for all edges ij of G.) Calculate $\chi_{\mathcal{B}_G}(q)$ in terms of $\chi_{\mathcal{A}_G}(q)$.

Problem 5.3. (Stanley, EC2, 3.115) Determine the characteristic polynomial and the number of regions of

the *type B braid arrangement* and the *type D braid arrangement* $\mathcal{B}_n, \mathcal{D}_n \subset \mathbb{R}^n$, which are defined by

$$\mathcal{B}_n = \{x_i = x_j : 1 \le i < j \le n\} \cup \{x_i = -x_j : 1 \le i < j \le n\} \cup \{x_i = 0 : 1 \le i \le n\},$$
$$\mathcal{D}_n = \{x_i = x_j : 1 \le i < j \le n\} \cup \{x_i = -x_j : 1 \le i < j \le n\}.$$

(Hint: Work out $\mathcal{B}_n$ first and use the result to understand $\mathcal{D}_n$.)

Problem 5.4 (Stanley [Sta07], Exercise 5.9(a))**.** Find the characteristic polynomial and number of regions of the arrangement $\mathcal{A}_n \subseteq \mathbb{R}^n$ with hyperplanes $x_i = 0$, $x_i = x_j$, and $x_i = 2x_j$, for all $1 \le i \ne j \le n$.

Problem 5.5. Recall that each permutation $w = (w_1, \ldots, w_n) \in \mathfrak{S}_n$ corresponds to a region of the braid arrangement Br_n, namely the open cone $C_w = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : x_{w_1} < x_{w_2} < \cdots < x_{w_n}\}$. Denote its closure by $\overline{C_w}$. For any set $W \subseteq \mathfrak{S}_n$, consider the closed fan

$$F(W) = \bigcup_{w \in W} \overline{C_w} = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : x_{w_1} \le \cdots \le x_{w_n} \text{ for some } w \in W\}.$$

Prove that $F(W)$ is a convex set if and only if W is the set of linear extensions of some poset P on $[n]$. (A *linear extension* of P is a total ordering $\prec$ consistent with the ordering of P, i.e., if $x <_P y$ then $x \prec y$.)

Problem 5.6. The runners in a sprint are seeded $1, \ldots, n$ (stronger runners are assigned higher numbers). To even the playing field, the rules specify that you earn one point for each higher-ranked opponent you beat, and one point for each lower-ranked opponent you beat *by at least one second*. (If a higher-ranked runner beats a lower-ranked runner by less than 1 second, no one gets the point for that matchup.) Let s_i be the number of points scored by the ith player and let $s = (s_1, \ldots, s_n)$ be the *score vector*.

(a) Show that the possible score vectors are in bijection with the regions of the Shi arrangement.
(b) Work out all possible score vectors in the cases of 2 and 3 players. Conjecture a necessary and sufficient condition for $(s_1, \ldots, s_n)$ to be a possible score vector for n players. Prove it if you can.

Problem 5.7. Prove Theorem 5.5.9.

Chapter 6

Simplicial Complexes

The canonical references for this material are [Sta96], [BH93, Ch. 5]. See also [MS05] (for the combinatorics and algebra) and [Hat02] (for the topology).

6.1 Basic definitions and terminology

Definition 6.1.1. Let V be a finite set of **vertices**. An **(abstract) simplicial complex** Δ on V is a nonempty family of subsets of V with the property that if $\sigma \in \Delta$ and $\tau \subseteq \sigma$, then $\tau \in \Delta$. Equivalently, Δ is an order ideal in the Boolean lattice 2^V. The elements of Δ are called its **faces** or **simplices**. A face that is maximal with respect to inclusion is called a **facet**.

The **dimension** of a face σ is $\dim \sigma = |\sigma| - 1$. A face of dimension k is a ***k*-face** or ***k*-simplex**. The **dimension** of a non-void simplicial complex Δ is $\dim \Delta = \max\{\dim \sigma \colon \sigma \in \Delta\}$. (Sometimes we write Δ^{d-1} to indicate that $\dim \Delta = d - 1$; this is a common convention since then d is the maximum number of vertices in a face.) A complex is **pure** if all its facets have the same dimension.

The simplest simplicial complexes are the **void complex** $\Delta = \emptyset$ (which is often excluded from consideration) and the **irrelevant complex** $\Delta = \{\emptyset\}$. In some contexts, there is the additional requirement that every singleton subset of V is a face (since if $v \in V$ and $\{v\} \notin \Delta$, then $v \notin \sigma$ for all $\sigma \in \Delta$, so you might as well replace V with $V \setminus \{v\}$). A simplicial complex with a single facet is also called a **simplex**.

The simplicial complex **generated** by a list of faces $\sigma_1, \ldots, \sigma_r$ is

$$\langle \sigma_1, \ldots, \sigma_r \rangle = \bigcup_{i=1}^{r} 2^{\sigma_i}.$$

The set of facets of a complex is the unique minimal set of generators for it.

Simplicial complexes are combinatorial models for compact topological spaces. The vertices $V = [n]$ can be regarded as the points $\mathbf{e}_1, \ldots, \mathbf{e}_n \in \mathbb{R}^n$, and a simplex $\sigma = \{v_1, \ldots, v_r\}$ is then the convex hull of the corresponding points:

$$|\sigma| = \operatorname{conv}\{\mathbf{e}_{v_1}, \ldots, \mathbf{e}_{v_r}\} = \{c_1 \mathbf{e}_{v_1} + \cdots + c_{v_r} \mathbf{e}_r \,:\, 0 \le c_i \le 1 \ (\forall i),\ \ c_1 + \cdots + c_n = 1\}\,.$$

For example, faces of sizes 1, 2, and 3 correspond respectively to vertices, line segments, and triangles. (This explains why $\dim \sigma = |\sigma| - 1$.) Taking $\{\mathbf{e}_i\}$ to be the standard basis of $\mathbb{R}^n$ gives the **standard geometric realization** $|\Delta|$ of Δ:

$$|\Delta| = \bigcup_{\sigma \in \Delta} \operatorname{conv}\{\mathbf{e}_i : i \in \sigma\}.$$

It is usually possible to realize Δ geometrically in a space of much smaller dimension. For example, every graph can be realized in $\mathbb{R}^3$, and planar graphs can be realized in $\mathbb{R}^2$. It is common to draw geometric pictures of simplicial complexes, just as we draw pictures of graphs. We sometimes use the notation $|\Delta|$ to denote any old geometric realization (i.e., any topological space homeomorphic to the standard geometric realization). Typically, it is easiest to ignore the distinction between Δ and $|\Delta|$; if we want to be specific we will use terminology like "geometric realization of Δ" or "face poset of Δ". A **triangulation** of a topological space X is a simplicial complex whose geometric realization is homeomorphic to X.

Figure 6.1 shows geometric realizations of the simplicial complexes $\Delta_1 = \langle 124, 23, 24, 34 \rangle$ and $\Delta_2 = \langle 12, 14, 23, 24, 34 \rangle$.

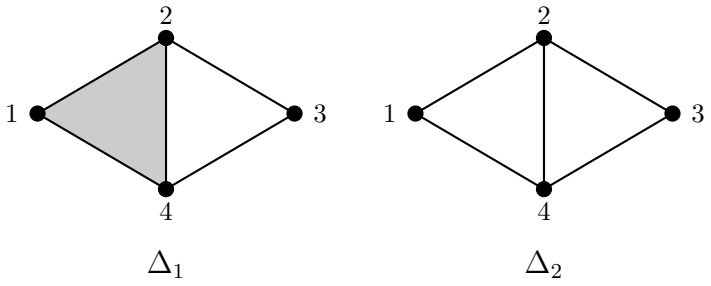


Figure 6.1: Two simplicial complexes.

The filled-in triangle indicates that 124 is a face of Δ_1, but not of Δ_2. Note that Δ_2 is the subcomplex of Δ_1 consisting of all faces of dimensions ≤ 1 — that is, it is the **1-skeleton** of Δ_1.

Some basic operations on simplicial complexes are as follows:

1. For $X \subseteq V$, the **induced subcomplex** is $\Delta|_X = \{\sigma \in \Delta : \sigma \subseteq X\}$.
2. The **deletion** $\operatorname{del}_\Delta(v)$ is the subcomplex of all faces not containing v. Equivalently, $\operatorname{del}_\Delta(v) = \Delta|_{V-v}$. More generally,
$$\operatorname{del}_\Delta(\sigma) = \{\tau \in \Delta : \tau \cap \sigma = \emptyset\} = \Delta|_{V \setminus \sigma}.$$
3. The **link** of a face σ is defined as
$$\operatorname{link}_\Delta(\sigma) = \{\tau \in \Delta : \tau \cap \sigma = \emptyset,\ \tau \cup \sigma \in \Delta\}.$$
The link can be thought of as "what you see if you stand in σ in look outward"; for example, if Δ is a triangulation of a $(d-1)$-dimensional manifold, then the link of every vertex is a $(d-2)$-sphere, and more generally the link of every k-dimensional face is a $(d-k-2)$-sphere.
4. The **join** of two complexes Δ, Δ' on disjoint vertex sets is
$$\Delta * \Delta' = \{\sigma \cup \sigma' : \sigma \in \Delta,\ \sigma' \in \Delta'\}.$$
Combinatorially, join behaves like a product; for example, it is multiplicative on f-vectors. On the other hand, it is not a product in the topological sense: $[\Delta * \Delta']$ is *not* homeomorphic to $[\Delta] \times [\Delta']$.

And here is the basic numerical invariant of a simplicial complex.

Definition 6.1.2. Let Δ^{d-1} be a simplicial complex. The ***f*-vector** of Δ is $(f_{-1}, f_0, f_1, \ldots, f_{d-1})$, where $f_i = f_i(\Delta)$ is the number of faces of dimension i. The term f_{-1} is often omitted, because $f_{-1} = 1$ unless Δ is the void complex. The ***f*-polynomial** is the generating function for the nonnegative f-numbers (essentially the rank-generating function of Δ as a poset):

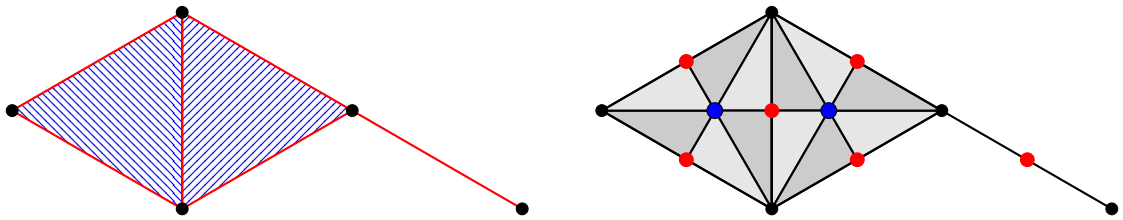
$$f(\Delta, q) = f_0 + f_1 q + f_2 q^2 + \cdots + f_{d-1} q^{d-1}.$$

For instance, if Δ_1, Δ_2 are the simplicial complexes pictured above, then

$$f(\Delta_1) = (4, 5, 1) \quad \text{and} \quad f(\Delta_2) = (4, 5).$$

Example 6.1.3. Let P be a finite poset and let $\Delta(P)$ be the set of chains in P. Every subset of a chain is a chain, so $\Delta(P)$ is a simplicial complex, called the **order complex** of P. The minimal nonfaces of $\Delta(P)$ are precisely the pairs of incomparable elements of P; in particular every minimal nonface has size two, which is to say that $\Delta(P)$ is a **flag complex**. Note that $\Delta(P)$ is pure if and only if P is ranked.

If P itself is the set of faces of a simplicial complex Δ, then $\Delta(P(\Delta))$ is the **barycentric subdivision** of that complex. Combinatorially, the vertices of $\mathrm{Sd}(\Delta)$ correspond to the faces of Δ; a collection of vertices of $\mathrm{Sd}(\Delta)$ forms a face if the corresponding faces of Δ are a chain in its face poset. Topologically, $\mathrm{Sd}(\Delta)$ can be constructed by drawing a vertex in the middle of each face of Δ and connecting them — this is best illustrated by a picture.



Δ $\qquad\qquad$ $\mathrm{Sd}(\Delta)$

Each vertex (black, red, blue) of $\mathrm{Sd}(\Delta)$ corresponds to a (vertex, edge, triangle) face of Δ. Note that barycentric subdivision does not change the topological space itself, only the triangulation of it. ◀

6.2 Simplicial homology

Simplicial complexes are models of topological spaces, and combinatorialists use tools from algebraic topology to study them, in particular the machinery of *simplicial homology*. Here we give a "user's guide" to the subject that assumes as little topology background as possible. Readers familiar with the subject will know that I am leaving many things out. For a full theoretical treatment, I recommend Chapter 2 of Hatcher [Hat02].

Let Δ be a simplicial complex on vertex set $[n]$. The ***k*th simplicial chain group** of Δ over a field[1], say $\mathbb{R}$, is the vector space $C_k(\Delta)$ of formal linear combinations of k-simplices in Δ. Thus $\dim C_k(\Delta) = f_k(\Delta)$. The

[1] More generally, this could be any commutative ring, but let's keep things simple for the moment.

elements of $C_k(\Delta)$ are called ***k*-chains**. The **(simplicial) boundary map** $\partial_k : C_k(\Delta) \to C_{k-1}(\Delta)$ is defined as follows: if $\sigma = \{v_0, \ldots, v_k\}$ is a k-face, with $1 \le v_0 < \cdots < v_k \le n$, then

$$\partial_k[\sigma] = \sum_{i=0}^{k} (-1)^i [v_0, \ldots, \widehat{v_i}, \ldots, v_k] \in C_{k-1}(\Delta)$$

where the hat denotes removal. The map is then extended linearly to all of $C_k(\Delta)$.

Recall that each $\sigma \in \Delta$ of cardinality $k+1$ is realized geometrically by a simplex, which is homeomorphic to a k-dimensional ball. The chain $\partial[\sigma]$ should be thought of as the $(k-1)$-sphere that is its boundary, expressed as a sum of $(k-1)$-simplices with consistent orientations (as represented by the signs). Often it is convenient to abbreviate ∂_k by ∂, since either the subscript is clear from context or else we want to say something about all boundary maps at once.

The entire collection of data $\{C_k(\Delta, \partial_k)\}$ is called the **simplicial chain complex** of Δ. For example, if $\Delta = \langle 123, 14, 24 \rangle$, then the simplicial chain complex is as follows:

$$C_2 = \mathbb{R}^1 \xrightarrow{\partial_2} C_1 = \mathbb{R}^5 \xrightarrow{\partial_1} C_0 = \mathbb{R}^4 \xrightarrow{\partial_0} C_{-1} = \mathbb{R}$$

$$\begin{array}{c} \\ 12 \\ 13 \\ 14 \\ 23 \\ 24 \end{array} \begin{array}{c} \begin{array}{c} 123 \end{array} \\ \begin{bmatrix} 1 \\ -1 \\ 0 \\ 1 \\ 0 \end{bmatrix} \end{array} \qquad \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccc} 12 & 13 & 14 & 23 & 24 \end{array} \\ \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ -1 & 0 & 0 & 1 & 1 \\ 0 & -1 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 & -1 \end{bmatrix} \end{array} \qquad \begin{array}{c} \\ \emptyset \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ [1 \quad 1 \quad 1 \quad 1] \end{array}$$

The fundamental fact about boundary maps is that $\partial_k \circ \partial_{k+1}$ for all k, a fact that is frequently written without subscripts:

$$\partial^2 = 0.$$

(This can be checked directly from the definition of ∂, and is a calculation that everyone should do for themselves once.) This is precisely what the term "chain complex" means in algebra.

An equivalent condition is that $\ker \partial_k \supseteq \operatorname{im} \partial_{k+1}$ for all k. In particular, we can define the **reduced simplicial homology groups**[2]

$$\tilde{H}_k(\Delta) \;=\; \ker \partial_k / \operatorname{im} \partial_{k+1}.$$

The $\tilde{H}_k(\Delta)$ are just $\mathbb{R}$-vector spaces, so they can be described up to isomorphism by their dimensions[3], which are called the **Betti numbers** $\beta_k(\Delta)$. They can be calculated using the rank-nullity formula: in general

$$\beta_k(\Delta) = \dim \tilde{H}_k(\Delta) = \dim \ker \partial_k - \dim \operatorname{im} \partial_{k+1} = f_k - \operatorname{rank} \partial_k - \operatorname{rank} \partial_{k+1}.$$

In the example above, this formula gives

$$\tilde{\beta}_0(\Delta) = 4 - 1 - 3 = 0, \qquad \tilde{\beta}_1(\Delta) = 5 - 3 - 1 = 1, \qquad \tilde{\beta}_2(\Delta) = 1 - 1 - 0 = 0$$

(note that ∂_3 is the zero map).

These numbers turn out to carry topological information about the space $|\Delta|$. In fact, they depend only on the homotopy type of the space $|\Delta|$. This is a fundamental theorem in topology whose proof is far too

[2]The unreduced homology groups $H_k(\Delta)$ are defined by deleting $C_{-1}(\Delta)$ from the simplicial chain complex. This results in an extra summand of $\mathbb{R}$ in $H_0(\Delta)$ and has no effect elsewhere. Broadly speaking, reduced homology arises more naturally in combinatorics and unreduced homology is more natural in topology, but the information is equivalent.

[3]This would not be true if we replaced $\mathbb{R}$ with a ring that was not a field. Actually, the most information is available over $\mathbb{Z}$. In that case $\beta_k(\Delta)$ can still be obtained as the rank of the free part if $\tilde{H}_k(\Delta)$, but there also may be a torsion part.

elaborate to give here,[4] but provides a crucial tool for studying simplicial complexes: we can now ask how the topology of Δ affects its combinatorics. To begin with, the groups $\tilde{H}_k(\Delta)$ do not depend on the choice of labeling of vertices and are invariant under retriangulation.

A complex all of whose homology groups vanish is called **acyclic**. For example, if $|\Delta|$ is contractible then Δ is acyclic over every ring. If $\Delta \cong \mathbb{S}^d$ (i.e., $|\Delta|$ is a d-dimensional sphere), then

$$\tilde{H}_k(\Delta) \cong \begin{cases} \mathbb{R} & \text{if } k = d, \\ 0 & \text{if } k < d. \end{cases} \tag{6.1}$$

The **(reduced) Euler characteristic** of Δ is

$$\tilde{\chi}(\Delta) = \sum_{k \geq 0} \beta_k(\Delta) = \sum_{i \geq -1} f_i(\Delta). \tag{6.2}$$

The second equality here is called the *Euler-Poincaré theorem*; despite the fancy name, it is easy to prove using little more than the rank-nullity theorem of linear algebra (Problem 6.10). The Euler characteristic is the single most important numerical invariant of Δ. Many combinatorial invariants can be computed by identifying them as the Euler characteristic of a simplicial complex whose topology is known, often one that is acyclic ($\tilde{\chi} = 0$), a sphere of dimension d ($\tilde{\chi} = (-1)^d$), or a wedge of spheres.

Observe that

$$\begin{aligned} \tilde{\chi}(\Delta) &= \sum_{\sigma \in \Delta:\ e \notin \sigma} (-1)^{\dim \sigma} + \sum_{\sigma \in \Delta:\ e \in \sigma} (-1)^{\dim \sigma} \\ &= \sum_{\sigma \in \mathrm{del}_\Delta(e)} (-1)^{\dim \sigma} + \sum_{\tau \in \mathrm{link}_\Delta(e)} (-1)^{1+\dim \tau} \\ &= \tilde{\chi}(\mathrm{del}_\Delta(e)) - \tilde{\chi}(\mathrm{link}_\Delta(e)). \end{aligned} \tag{6.3}$$

which resembles a deletion/contraction recurrence. (This is no accident; see (4.2).)

6.3 Stanley-Reisner theory

The **Stanley-Reisner correspondence** is an extremely important connection between simplicial complexes and commutative algebra. Let $R = \Bbbk[x_1, \ldots, x_n]$ be the ring of polynomials in n variables over your favorite field $\Bbbk$. Define the **support** of a monomial $\mu \in R$ as

$$\operatorname{supp} \mu = \{i \colon x_i \text{ divides } \mu\}.$$

Definition 6.3.1. Let Δ be a simplicial complex on vertex set $[n]$. Its **Stanley-Reisner ideal** in R is

$$I_\Delta = \langle x_{i_1} \cdots x_{i_r} \colon \{i_1, \ldots, i_r\} \notin \Delta \rangle .$$

The **Stanley-Reisner ring** or **face ring** is $\Bbbk[\Delta] := R/I_\Delta$.

[4] Roughly, one defines a much more abstract set of invariants called *singular homology groups*, which are easily seen to be topological invariants but are well-nigh impossible to work with directly; one then shows that repeatedly barycentrically subdividing a space allows us to approximate singular homology by simplicial homology sufficiently accurately — but on the other hand subdivision also preserves simplicial homology, so we can have the best of both worlds. See [Hat02, §2.1] for the full story. Or take my MATH 821 class!

Example 6.3.2. Let Δ_1 and Δ_2 be the complexes of Figure 6.1. Abbreviating $w, x, y, z = x_1, x_2, x_3, x_4$, the Stanley-Reisner ideal of Δ_1 is

$$I_{\Delta_1} = \langle wxyz,\ wxy,\ wyz,\ xyz,\ wy \rangle = \langle xyz, wy \rangle.$$

Note that the minimal generators of I_Δ are the minimal nonfaces of Δ. Similarly,

$$I_{\Delta_2} = \langle wxz,\ xyz,\ wy \rangle.$$

If Δ is the simplex on $[n]$ then it has no nonfaces, so I_Δ is the zero ideal and $\Bbbk[\Delta] = \Bbbk[x_1, \ldots, x_n]$. In general, the more faces Δ has, the bigger its Stanley-Reisner ring is. ◀

Since Δ is a simplicial complex, the monomials in I_Δ are *exactly* those whose support is not a face of Δ. Therefore, the monomials supported on a face of Δ are a natural vector space basis for the graded ring $\Bbbk[\Delta]$. Its Hilbert series can be calculated by counting these monomials:

$$\begin{aligned}
\operatorname{Hilb}(\Bbbk[\Delta^{d-1}], q) &\stackrel{\text{def}}{=} \sum_{i \geq 0} q^i \dim_{\Bbbk}(\Bbbk[\Delta])_i = \sum_{\sigma \in \Delta} \sum_{\substack{\text{monomials } \mu: \\ \operatorname{supp} \mu = \sigma}} q^{\deg(\mu)} \\
&= \sum_{\sigma \in \Delta} \left(\frac{q}{1-q} \right)^{|\sigma|} \\
&= \sum_{i=0}^{d} f_{i-1} \left(\frac{q}{1-q} \right)^i = \frac{\displaystyle\sum_{i=0}^{d} f_{i-1} q^i (1-q)^{d-i}}{(1-q)^d} = \frac{\displaystyle\sum_{i=0}^{d} h_i q^i}{(1-q)^d}
\end{aligned}$$

The numerator of this rational expression is a polynomial in q, called the ***h*-polynomial** of Δ and written $h_\Delta(q)$, and its list of coefficients $(h_0, h_1, \ldots, H_D)$ is called the ***h*-vector** of Δ. Clearing denominators and applying the binomial theorem yields a formula for the h-numbers in terms of the f-numbers:

$$\begin{aligned}
\sum_{i=0}^{d} h_i q^i = \sum_{i=0}^{d} f_{i-1} q^i (1-q)^{d-i} &= \sum_{i=0}^{d} f_{i-1} q^i \sum_{j=0}^{d-i} \binom{d-i}{j} (-1)^j q^j \\
&= \sum_{i=0}^{d} \sum_{j=0}^{d-i} \binom{d-i}{j} (-1)^j q^{i+j} f_{i-1}
\end{aligned}$$

and now extracting the q^k coefficient (i.e., the summand in the second sum with $j = k - i$) yields

$$h_k(\Delta) = \sum_{i=0}^{k} \binom{d-i}{k-i} (-1)^{k-i} f_{i-1}(\Delta). \tag{6.4}$$

where $\dim \Delta = d - 1$. (Note that the upper limit of summation might as well be k instead of d, since the binomial coefficient in the summand vanishes for $i > k$.) These equations can be solved to give the f's in terms of the h's.

$$f_{i-1}(\Delta) = \sum_{k=0}^{i} \binom{d-k}{i-k} h_k(\Delta). \tag{6.5}$$

So the f-vector and h-vector contain the same information about Δ. On the level of generating functions, the conversions look like this [BH93, p. 213]:

$$\sum_i h_i q^i = \sum_i f_{i-1} q^i (1-q)^{d-i}, \tag{6.6}$$

$$\sum_i f_i q^i = \sum_i h_i q^{i-1} (1+q)^{d-i}. \tag{6.7}$$

The equalities (6.4) and (6.5) can be obtained by applying the binomial theorem to the right-hand sides of (6.6) and (6.7) and equating coefficients. Note that it is most convenient simply to sum over all $i \in \mathbb{Z}$.

Two useful special cases are as follows: (6.4) gives

$$h_d = \sum_{i=0}^{d} \binom{d-i}{d-i} (-1)^{d-i} f_{i-1} = (-1)^{d-1} \tilde{\chi}(\Delta),$$

the reduced Euler characteristic. Also, (6.5) gives

$$f_{d-1} = \sum_{k=0}^{d} h_k.$$

Let's go back to the formula for the Hilbert series in terms of the h-vector, namely

$$\mathrm{Hilb}(\Bbbk[\Delta], q) \;=\; \frac{h_\Delta(q)}{(1-q)^d} \;=\; \frac{\displaystyle\sum_{i=0}^{d} h_i q^i}{(1-q)^d}.$$

Note that $1/(1-q)^d$ is just the Hilbert series of the polynomial ring $\Bbbk[x_1, \ldots, x_n]$. More generally, if R is any graded ring and x is an indeterminate of degree 1, then

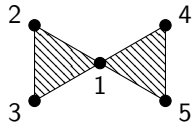
$$\mathrm{Hilb}(R[x], q) = \frac{\mathrm{Hilb}(R, q)}{1-q}.$$

These observations suggest that we should be able to regard the Stanley-Reisner ring $\Bbbk[\Delta]$ as a polynomial ring in d variables over a base ring S whose Hilbert series is the polynomial $h_\Delta(q)$. In particular S would have to be a finite-dimensional vector space and h_i the dimension of its ith graded piece. Also, we should be able to recover S by quotienting out by d linear forms, each of which would remove a factor of $1/(1-q)$ from the Hilbert series. A ring for which this all works is called a **Cohen-Macaulay** (CM) ring, and a **Cohen-Maculay simplicial complex** is one whose Stanley-Reisner ring is CM.

Example 6.3.3. The **bowtie complex** is the pure 2-dimensional complex $\Delta = \langle 123, 145 \rangle$ shown below, with f-vector $(1, 5, 6, 2)$. Therefore, by (6.6), the h-polynomial is

$$\sum_i h_i q^i = 1q^0(1-q)^3 + 5q^1(1-q)^2 + 6q^2(1-q)^1 + 2q^3(1-q)^0 = 1 + 2q - q^2$$

so the h-vector is $(1, 2, -1)$.



This complex cannot possibly be CM, because if $\Bbbk[\Delta]$ were the ring of polynomials in two indeterminates over a subring S, then the Hilbert series of S would have to be $1 + 2q - q^2$ — in particular, the degree-2 graded piece of S would be a vector space of dimension -1, which is absurd. ◀

In particular, the h-numbers of a Cohen-Macaulay simplicial complex are all nonnegative, suggesting that they should count something (that is, something more combinatorial then the dimensions of the subring over which $\Bbbk[\Delta]$ is a polynomial ring).

6.4 Shellable and Cohen-Macaulay simplicial complexes

Here is an important special class of complexes where the h-numbers have a direct combinatorial interpretation.

Definition 6.4.1. A pure simplicial complex Δ^{d-1} is **shellable** if its facets can be ordered $F_1, \ldots, F_n$ such that any of the following conditions are satisfied:

1. For every $i \in [n]$, the set $\Psi_i = \langle F_i \rangle \setminus \langle F_1, \ldots, F_{i-1} \rangle$ has a unique minimal element R_i.
2. For every $i > 1$, the complex $\Phi_i = \langle F_i \rangle \cap \langle F_1, \ldots, F_{i-1} \rangle$ is pure of dimension $d-2$.

The proof of equivalence is left as an exercise (Problem 6.4).

Example 6.4.2. The **bipyramid** is the pure 2-dimensional complex B with 6 facets 124, 134, 234, 125, 135, 235. Vertices 1,2,3 form the "equator"; vertices 4 and 5 are the "poles". The complex B has many shelling orders, one of which is

$$234, \quad 124, \quad 134, \quad 235, \quad 125, \quad 135.$$

The bipyramid and its shelling decomposition is shown in Figure 6.2. The new edges created upon adding each triangle are indicated in bold. The corresponding decomposition of the face poset is

$$[\emptyset, 234] \cup [1, 124] \cup [13, 134] \cup [5, 235] \cup [15, 125] \cup [135, 135]$$

as shown in the figure (each face is color-coded according to the interval $[R_i, F_i]$ that contains it). ◀

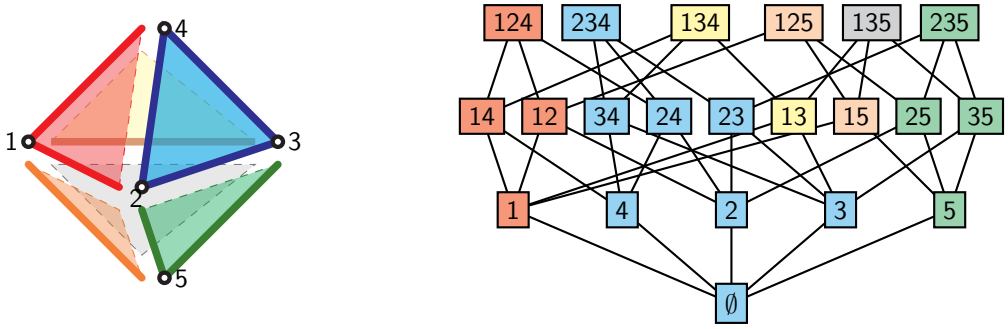


Figure 6.2: A shelling of the bipyramid.

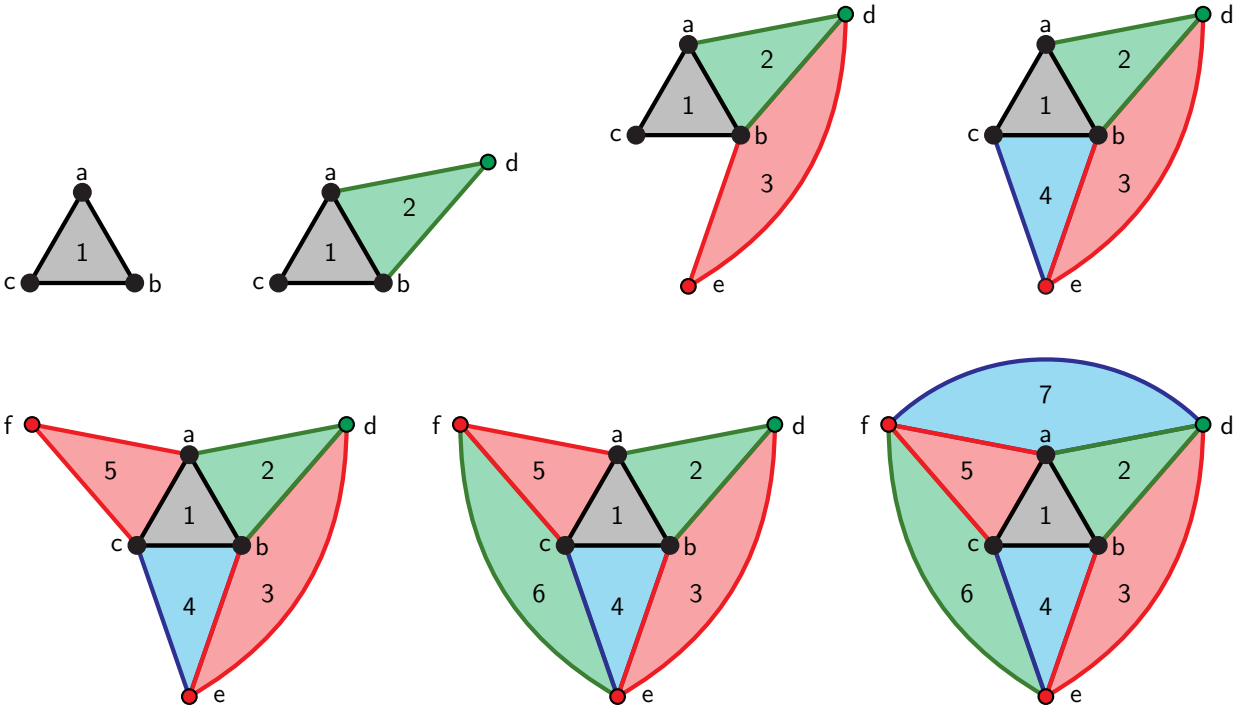
Figure 6.3 shows another example that shows how a shelling builds up a simplicial complex (in this case the boundary of an octahedron) one step at a time. Note that each time a new triangle is attached, there is a unique minimal new face.

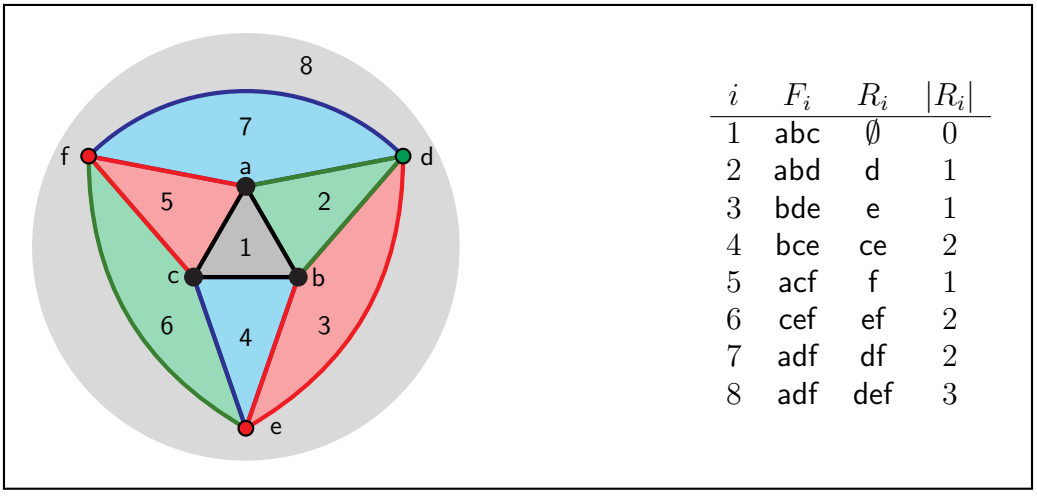
Proposition 6.4.3. *Let Δ^{d-1} be shellable, with h-vector $(h_0, \ldots, h_d)$. Then*

$$\begin{aligned} h_j &= \#\{F_i \colon \#R_i = j\} \\ &= \#\big\{F_i \colon \langle F_i \rangle \cap \langle F_1, \ldots, F_{i-1} \rangle \text{ has } j \text{ faces of dimension } d-2\big\}. \end{aligned}$$

Moreover, if $h_j(\Delta) = 0$ for some j, then $h_k(\Delta) = 0$ for all $k > j$.

The proof is left as an exercise. One consequence is that the h-vector of a shellable complex is strictly non-negative, since its coefficients count something. This statement is emphatically not true about the Hilbert series of arbitrary graded rings, or even arbitrary Stanley-Reisner rings of pure complexes (see Example 6.3.3 above).





i	F_i	R_i	$\lvert R_i\rvert$
1	abc	$\emptyset$	0
2	abd	d	1
3	bde	e	1
4	bce	ce	2
5	acf	f	1
6	cef	ef	2
7	adf	df	2
8	adf	def	3

Figure 6.3: A step-by-step shelling of the octahedron with vertices a,b,c,d,e,f. Facets are labeled $1 \ldots 8$ in shelling order. Enumerating the sets R_i by cardinality gives the h-vector $(1, 3, 3, 1)$.

If a simplicial complex is shellable, then its Stanley-Reisner ring is *Cohen-Macaulay* (CM). This is an important and subtle algebraic condition that can be expressed algebraically in terms of depth or local cohomology (topics beyond the scope of these notes) or in terms of simplicial homology (coming shortly). Shellability is the most common combinatorial technique for proving that a ring is CM. The constraints on the h-vectors of CM complexes are the same as those on shellable complexes, although it is an open problem to give a general combinatorial interpretation of the h-vector of a CM complex.

The Cohen-Macaulay condition can be expressed homologically, using links:

Proposition 6.4.4 (Reisner's theorem)**.** *A simplicial complex Δ is Cohen-Macaulay over R iff (a) Δ is pure (so that $\dim \mathrm{link}_\Delta(\sigma) = \dim \Delta - \dim \sigma - 1$ for all σ) and (b) for every face $\sigma \in \Delta$, one has*

$$\tilde{H}_k(\mathrm{link}_\Delta(\sigma); R) = 0 \qquad \forall k < \dim \Delta - \dim \sigma - 1.$$

Reisner's theorem can be used to prove that shellable complexes are Cohen-Macaulay. The other ingredient of this proof is a Mayer-Vietoris sequence, which is a standard tool in topology that functions sort of like an inclusion/exclusion principle for homology groups, relating the homology groups of X, Y, $X \cup Y$ and $X \cap Y$. Here we can take X to be the subcomplex generated by the first $n-1$ facets in shelling order and Y the nth facet; the shelling condition says that the intersections and their links are extremely well-behaved, so that Reisner's condition can be established by induction on n.

Reisner's theorem often functions as a working definition of the Cohen-Macaulay condition for combinatorialists. The vanishing condition says that every link has the homology type of a **wedge of spheres** of the appropriate dimension. (The wedge sum of a collection of spaces is obtained by identifying a point of each; for example, the wedge of n circles looks like a flower with n petals. Reduced homology is additive on wedge sums, so by (6.1) the wedge sum of n copies of $\mathbb{S}^d$ has reduced homology R^n in dimension d, and 0 in other dimensions.)

A Cohen-Macaulay complex Δ is **Gorenstein** (over R) if in addition $\tilde{H}_{\dim \Delta - \dim \sigma - 1}(\mathrm{link}_\Delta(\sigma); R) \cong R$ for all σ. That is, every link has the homology type of a sphere. This is very close to being a manifold. (I don't know offhand of a Gorenstein complex that is not a manifold, although I'm sure examples exist.)

6.5 Matroid complexes

A simplicial complex Δ on vertex set E is a **matroid complex** if it is the family of independent sets of some matroid M on E (see Defn. 3.4.1); in this case we write $\Delta = \mathscr{I}(M)$. Many of the standard constructions of matroid theory can be translated into simplicial complex language.

- A vertex $e \in E$ is a cone point in Δ if and only if it is a coloop of M.
- $\mathscr{I}(M - e) = \mathrm{del}_\Delta(e)$ and $\mathscr{I}(M/e) = \mathrm{link}_\Delta(e)$. More generally, $\mathscr{I}(M/A) = \mathrm{link}_\Delta(\sigma)$, where σ is any basis of $M|_A$. (It is worth checking that this construction does not depend on the choice of σ.)
- If M, M' have disjoint ground sets, then $\mathscr{I}(M \otimes M')$ is the join $\mathscr{I}(M) * \mathscr{I}(M')$.

Say that a complex Δ **has property P hereditarily** if every induced subcomplex $\Delta|_X$ has property P; for example, we have already seen that matroid complexes are hereditarily pure. (Note that the induced subcomplex of Δ on its entire vertex set is just itself, so if Δ has P hereditarily then in particular it has P.)

Theorem 6.5.1. *Let Δ be an abstract simplicial complex on E. The following are equivalent:*

1. *Δ is a matroid independence complex.*

2. Δ *is hereditarily shellable.*
3. Δ *is hereditarily Cohen-Macaulay.*
4. Δ *is hereditarily pure.*

Proof. Work on this The implications (2) $\implies$ (3) $\implies$ (4) are consequences of the material in Chapter 6 (the first is a homework problem and the second is easy).

(4) $\implies$ (1): Suppose I, J are independent sets with $|I| < |J|$. Then the induced subcomplex $\Delta|_{I\cup J}$ is pure, which means that I is not a maximal face of it. Therefore there is some $x \in (I \cup J) \setminus I = J \setminus I$ such that $I \cup x \in \Delta$, establishing **(I3)**.

(1) $\implies$ (4): Let $F \subseteq E$. If I is a non-maximum face of $\Delta|_F$, then we can pick J to be a maximum face, and then **(I3)** says that there is some $x \in J$ such that $I + x$ is a face of Δ, hence of $\Delta|_F$.

(4) $\implies$ (2): More interesting; left as an exercise. □

6.6 Combinatorial Laplacians

To be written

6.7 Exercises

Problem 6.1. Let Δ be a simplicial complex on vertex set V, and let $v_0 \notin V$. The **cone over Δ** is the simplicial complex $C\Delta$ generated by all faces $\sigma + v_0$ for $\sigma \in \Delta$.

(a) (Warmup) Prove that $f(C\Delta, t) = (t+1)f(\Delta, t)$.
(b) Prove that $h(C\Delta, t) = h(\Delta, t)$.
(c) Prove that Δ is shellable if and only if $C\Delta$ is shellable. More specifically, $F_1, \ldots, F_n$ is a shelling of Δ if and only if $F_1 + v, \ldots, F_n + v$ is a shelling of $C\Delta$.

Problem 6.2. Let Δ be a graph (that is, a 1-dimensional simplicial complex) with c components, v vertices, and e edges. Determine the isomorphism types of the simplicial homology groups $\tilde{H}_0(\Delta; R)$ and $\tilde{H}_1(\Delta; R)$ for any coefficient ring R.

Problem 6.3. Construct two simplicial complexes with the same f-vector such that one is shellable and one isn't.

Problem 6.4. Prove that the two conditions in the definition of shellability (Defn. 6.4.1) are equivalent.

Problem 6.5. Prove Proposition 6.4.3.

Problem 6.6. Prove that the link operation commutes with union and intersection of complexes. That is, if X, Y are simplicial complexes that are subcomplexes of a larger complex $X \cup Y$, and $\sigma \in X \cup Y$, then prove that

$$\operatorname{link}_{X\cup Y}(\sigma) = \operatorname{link}_X(\sigma) \cup \operatorname{link}_Y(\sigma) \qquad \text{and} \qquad \operatorname{link}_{X\cap Y}(\sigma) = \operatorname{link}_X(\sigma) \cap \operatorname{link}_Y(\sigma).$$

Problem 6.7. Let Δ be a pure simplicial complex of dimension $d-1$. Δ is called *shifted* if its vertex set can be labeled $1, \ldots, n$ such that the following property holds: if $\sigma \in \Delta$, $j \in \sigma$, $i \notin \sigma$, and $i < j$, then $\sigma \setminus \{j\} \cup \{i\} \in \Delta$.

Equivalently, define a partial order $\preceq$ (called *Gale order* or *componentwise order*) on d-sets of positive integers as follows: if $\mathbf{a} = (a_1 < \cdots < a_d)$ and $\mathbf{b} = (b_1 < \cdots < b_d)$, then $\mathbf{a} \preceq \mathbf{b}$ if $a_i \leq b_i$ for all $i \in [d]$. Then Δ is shifted if and only if its facets form an order ideal in Gale order.

(a) Prove that every shifted complex is shellable.
(b) Use part (a) to find a combinatorial formula for its h-vector.
(c) [Klivans' Theorem] Prove that a shifted complex is a matroid complex if and only if it has a single maximal element when considered as an order ideal in Gale order.

Problem 6.8. (Requires some experience with homological algebra.) Prove that shellable simplicial complexes are Cohen-Macaulay. (Hint: First do the previous problem. Then use a Mayer-Vietoris sequence.)

Problem 6.9. Complete the proof of Theorem 6.5.1 by showing that hereditarily pure simplicial complexes are shellable. (Hint: Pick a vertex v. Show that the two complexes

$$\Delta_1 = \mathrm{del}_\Delta(v) = \langle \sigma \in \Delta \colon v \notin \sigma \rangle,$$
$$\Delta_2 = \mathrm{link}_\Delta(v) = \langle \sigma - v \in \Delta \colon v \in \sigma \rangle$$

are both shellable. Then concatenate the shelling orders to produce a shelling order on Δ. You will probably need Problem 6.1.) As a consequence of the construction, derive a relationship among the h-polynomials of Δ, Δ_1, and Δ_2.

Problem 6.10. Prove the *Euler-Poincaré formula*:

$$\tilde{\chi}(\Delta) = \sum_{k \geq -1} (-1)^k \dim_{\Bbbk} \tilde{H}_k(\Delta; \Bbbk).$$

(Despite the appearance of homology, all you really need is the rank-nullity theorem from linear algebra. The choice of ground field $\Bbbk$ is immaterial, but you can take it to be $\mathbb{R}$ if you want.)

Problem 6.11. Express the h-vector of a matroid complex in terms of the Tutte polynomial of the underlying matroid. (Hint: First figure out a deletion/contraction recurrence for the h-vector, using Problem 6.9.)

Problem 6.12. Let $V = \{x_{11}, x_{12}, \ldots, x_n 1, x_{n2}\}$. Consider the simplicial complex

$$\Delta_n = \{\sigma \subseteq V : \ \sigma \not\supseteq \{x_i, y_i\} \ \ \forall i \in [n]\}.$$

(In fact, Δ_n is the boundary sphere of the **crosspolytope**, the convex hull of the standard basis vectors and their negatives in $\mathbb{R}^n$.) Determine the f- and h-polynomials of Δ_n.

More generally, let $V_1, \ldots, V_n$ be pairwise-disjoint sets of sizes $c_1, \ldots, c_n$ and let $V = V_1 \cup \cdots \cup V_n$. The corresponding **complete colorful complex** is

$$\Delta(c_1, \ldots, c_n) = \{\sigma \subseteq V : \ |\sigma \cap V_i| \leq 1 \ \ \forall i \in [n]\}.$$

(The previous problem is the case that $c_i = 2$ for all i.) Show that $\Delta(c_1, \ldots, c_n)$ is shellable. Determine its f- and h-polynomials.

Chapter 7

Polytopes and Polyhedra

7.1 The basics

Polytopes include familiar objects such as cubes, pyramids, and Platonic solids. They are central in linear programming and therefore in optimization, and exhibit a wealth of nice combinatorics. The classic book on polytopes is Grünbaum [Grü03]; an equally valuable, more recent reference is Ziegler [Zie95]. A good reference for the basics is chapter 2 of Schrijver's notes [Sch13].

First some key terms. A subset $S \subseteq \mathbb{R}^n$ is **convex** if, for any two points in S, the line segment joining them is also a subset of S. The smallest convex set containing a given set T is called its **convex hull**, denoted $\operatorname{conv}(T)$. Explicitly, one can show (Problem 7.2; not hard) that

$$\operatorname{conv}(\mathbf{x}_1, \dots, \mathbf{x}_r) = \left\{ c_1\mathbf{x}_1 + \cdots + c_r\mathbf{x}_r : 0 \le c_i \le 1 \text{ for all } i \text{ and } \sum_{i=1}^{r} c_i = 1 \right\}. \tag{7.1}$$

These points are called **convex linear combinations** of the $\mathbf{x}_i$. A related definition is the **affine hull** of a point set, which is the smallest affine linear space containing it:

$$\operatorname{aff}(\mathbf{x}_1, \dots, \mathbf{x}_r) = \left\{ c_1\mathbf{x}_1 + \cdots + c_r\mathbf{x}_r : \sum_{i=1}^{r} c_i = 1 \right\}. \tag{7.2}$$

The interior of S as a subspace of its affine span is called the **relative interior** of S, denoted $\operatorname{relint} S$. This concept is necessary to talk about interiors of different-dimensional polyhedra in a sensible way. For example, the closed line segment $S = \{(x, 0) : 0 \le x \le 1\}$ in $\mathbb{R}^2$ has empty interior as a subset of $\mathbb{R}^2$, but its affine span is the x-axis, so $\operatorname{relint} S = \{(x, 0) : 0 < x < 1\}$.

Clearly $\operatorname{conv}(T) \subseteq \operatorname{aff}(T)$ (in fact, the inclusion is strict if $1 < |T| < \infty$). For example, the convex hull of three non-collinear points in $\mathbb{R}^n$ is a triangle, while their affine hull is the unique plane (i.e., affine 2-space) containing that triangle.

Definition 7.1.1. A **polyhedron** P is a nonempty intersection of finitely many closed half-spaces in $\mathbb{R}^n$. Equivalently,

$$P = \{\mathbf{x} \in \mathbb{R}^n : a_{i1}x_1 + \cdots + a_{in}x_n \ge b_i \quad \forall i \in [m]\}$$

where $a_{ij}, b_i \in \mathbb{R}$. These equations are often written as a single matrix equation $A\mathbf{x} \ge \mathbf{b}$, where $A \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$.

Definition 7.1.2. A **polytope** is a bounded polyhedron.

The "Fundamental Theorem of Polytopes" asserts that polytopes are precisely the sets in $\mathbb{R}^n$ that can be expressed as the convex hull of a finite set of points. We will prove this theorem in the next section.

Definition 7.1.3. A point $\mathbf{v}$ in a polyhedron P is a **vertex** of P if $\mathbf{v} \notin \operatorname{conv}(P \setminus \{\mathbf{v}\})$. The set of all vertices of P will be denoted by $V(P)$.

Definition 7.1.4. Let $P \subseteq \mathbb{R}^n$ be a polyhedron. A **face** of P is a subset $F \subseteq P$ that maximizes some linear functional $\ell : \mathbb{R}^n \to \mathbb{R}$, i.e., $\ell(x) \geq \ell(y)$ for all $x \in F$, $y \in P$. In this case, we write $F = \max_P(\ell)$. The face is **proper** if ℓ is not a constant. The **dimension** of a face is the dimension of its affine span.

The only improper face is P itself. Note that the union of all proper faces is the topological boundary ∂P (proof left as an exercise).

To make this a bit more concrete, suppose P is a polytope in $\mathbb{R}^3$. What point or set of points is highest? In other words, what points maximize the linear functional $(x, y, z) \mapsto z$? The answer to this question might be a single vertex, or an edge, or a polygonal face. Of course, there is nothing special about the z-direction. For any direction given by a linear functional ℓ, the extreme points of P in that direction are by definition the maxima of the linear functional $\mathbf{x} \mapsto \ell(\mathbf{x})$, and the set of those points forms a face of P.

For a linear functional chosen "at random", the face it determines will almost surely be a vertex of P. Higher-dimensional faces correspond to more special directions.

Proposition 7.1.5. *Let $P \subseteq \mathbb{R}^n$ be a polyhedron. Then:*

1. *Every face of P is also a polyhedron, and every face of a face of P is a face of P.*
2. *The intersection of any two faces is a face (provided it is nonempty).*
3. *For each $\mathbf{x} \in P$, there is a unique minimal face $F_\mathbf{x} \subseteq P$ containing $\mathbf{x}$.*
4. *$\mathbf{x} \in \operatorname{relint} F_\mathbf{x}$ for every $\mathbf{x} \in P$.*
5. *The vertices of a polytope are exactly its 0-dimensional faces.*
6. *The faces of P form a lattice $\mathscr{F}(P)$ under inclusion, with bottom element $\emptyset$ and top element P, and meet given by intersection. Moreover, the face lattice is ranked, with rank function $r(Q) = 1 + \dim Q$ (in particular, $r(\emptyset) = 0$).*

Proof. (1) Each face F is defined by adding a linear inequality to the list of inequalities defining P. Specifically, if $F = \max_P(\ell)$, and $\ell(\mathbf{x}) = m$ for all $\mathbf{x} \in F$, then $F = \{\mathbf{x} \in P : \ell(\mathbf{x}) \geq m\}$.

(2) Let $F' = \max_P(\ell')$ and $F'' = \max_P(\ell'')$ and suppose that $F' \cap F''$ contains a point $\mathbf{x}$. Let $F = \max_P(\ell)$, where $\ell = \ell' + \ell''$ (in fact any positive linear combination of ℓ', ℓ'' will do). Then $\mathbf{x}$ is a global maximum of ℓ on P, and since $\mathbf{x}$ also maximizes both ℓ' and ℓ'', the face F consists exactly of those points of P maximizing both ℓ' and ℓ''. In other words, $F = F' \cap F''$, as desired.

(3) By (2), the desired face $F_\mathbf{x}$ is the intersection of all faces containing $\mathbf{x}$.

(4) If $\mathbf{x} \in \partial F_\mathbf{x}$ then $F_\mathbf{x}$ has a face G containing $\mathbf{x}$, but G is also a face of P by (1), which contradicts the definition of $F_\mathbf{x}$.

(5) Suppose that $\mathbf{x}$ is a 0-dimensional face, i.e., $\{\mathbf{x}\} = \max_P(\ell)$. If $\mathbf{x}$ is a convex linear combination $\sum c_i \mathbf{y}_i$ of points $\mathbf{y}_i \in P$, then $\ell(\mathbf{x}) \geq \sum c_i \ell(\mathbf{y}_i)$, with equality only if $\ell(\mathbf{y}_i) = \ell(\mathbf{x})$ for all i. But then $\mathbf{y}_i = \mathbf{x}$ for all i by assumption. Therefore $\mathbf{x} \notin \operatorname{conv}(P \setminus \{\mathbf{x}\})$, hence is a vertex.

On the other hand, if $\mathbf{x} \in P$ is not an 0-dimensional face, then by (4) $\mathbf{x} \in \operatorname{relint} F_{\mathbf{x}}$. Then $F_{\mathbf{x}}$ contains a ball centered at $\mathbf{x}$, hence a line segment centered at $\mathbf{x}$, and thus $\mathbf{x}$ is a linear combination of the two endpoints of the segment (namely, their average). Hence $\mathbf{x}$ is not a vertex.

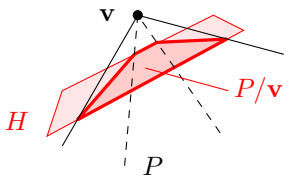
(6) The set $\mathscr{F}(P)$ is certainly a bounded poset under inclusion, and by (2) it is a meet-semilattice, hence a lattice by Prop. 1.2.9.

To prove that it is ranked, we need a new construction. Let $\mathbf{v}$ be a vertex maximized by some linear functional ℓ, so that we can find a constant c such that

$$\ell(\mathbf{v}) > c > \ell(\mathbf{x}) \quad \forall \mathbf{x} \in V(P) \setminus \{\mathbf{x}\}.$$

The **vertex figure** of P with respect to $\mathbf{v}$ is defined by

$$P/\mathbf{v} = P \cap H, \quad \text{where } H = \{\mathbf{x} \colon \ell(\mathbf{x}) = c\}.$$



One can show [Zie95, Prop. 2.4] that $P/\mathbf{v}$ is a polytope of dimension $\dim(P) - 1$, and that there is a bijection

$$\begin{array}{rcl} \{k\text{-dimensional faces of } P/\mathbf{v}\} & \longrightarrow & \{(k+1)\text{-dimensional faces of } P \text{ containing } \mathbf{v}\} \\ F & \longmapsto & \operatorname{conv}(F \cup \{\mathbf{v}\}) \\ G \cap H & \longleftarrow\!\shortmid & G \end{array}$$

regardless of the particular choice of ℓ and H. In particular, the face lattice $\mathscr{F}(P/\mathbf{v})$ is isomorphic to the interval $[\mathbf{v}, P] \subset \mathscr{F}(P)$. □

Here is a fundamental construction.

Definition 7.1.6. Let $P \subset \mathbb{R}^n$ be a polytope. Assume without loss of generality that $\operatorname{aff}(P) = \mathbb{R}^n$ (otherwise, replace $\mathbb{R}^n$ with the affine hull) and that the origin is in the interior of P (translating if necessary). The **dual polytope** (or **polar dual** or **polar**) of P is

$$P^* := \{\mathbf{y} \in \mathbb{R}^n \mid \mathbf{x} \cdot \mathbf{y} \leq 1 \ \ \forall \mathbf{x} \in P\}. \tag{7.3}$$

Observe that P^* is bounded. (P contains a ball of radius ϵ centered at the origin, which implies that P^* is contained in a ball of radius $1/\epsilon$.) Moreover, by definition P^* is the intersection of half-spaces, but it is not clear at this point that it is the intersection of finitely many of them — we will prove that in the next section (and give an example then).

7.2 The Fundamental Theorem of Polytopes

Temporarily, say that a subset of $\mathbb{R}^n$ is a **P-polytope** if it is the convex hull of a finite set of points.

Theorem 7.2.1 (The Fundamental Theorem of Polytopes)**.** *A set $P \subseteq \mathbb{R}^n$ is a polytope (i.e., a bounded polyhedron) if and only if it is a P-polytope.*

The proof occupies this whole section, and will include several proofs of claims along the way.

First, let P be the intersection of finitely many half-spaces, i.e., $P = \{\mathbf{x} \in \mathbb{R}^n \colon A\mathbf{x} \leq b\}$, where $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^{m\times 1}$. By projecting onto the orthogonal complement of the rowspace of A, we can assume WLOG that $\operatorname{rank} A = n$. For each point $\mathbf{x} \in P$, let $A_{\mathbf{x}}$ be the submatrix of A consisting of rows a_i for which $a_i \cdot \mathbf{x} = b_i$. (These rows correspond to linear functionals maximized at $\mathbf{x}$.)

Claim 1: Let $\mathbf{x} \in P$. Then $\{\mathbf{x}\} = \max_P(\ell)$ for some $\ell \in (\mathbb{R}^n)^*$ if and only if $\operatorname{rank} A_{\mathbf{x}} = n$. (More generally, $\operatorname{rank} A_{\mathbf{x}} = n - \dim F_{\mathbf{x}}$.)

Proof of Claim 1. If $\operatorname{rank} A_{\mathbf{x}} = n$ then there is a basis $\{\lambda_1, \dots, \lambda_n\}$ for $(\mathbb{R}^n)^*$ such that $\mathbf{x} \in \max_P(\lambda_i)$ for each i. Let $\lambda = \sum c_i\lambda_i$, where $c_1, \dots, c_n > 0$. Since the λ_i form a basis, if $\mathbf{x}$ is any other point in P, then there is some i such that $\lambda_i(\mathbf{x}) \neq \lambda_i(\mathbf{x})$; by assumption this must mean $\lambda_i(\mathbf{x}) < \lambda_i(\mathbf{x})$, and so $\lambda(\mathbf{x}) < \lambda(\mathbf{x})$. It follows that $\max_P(\lambda) = \{\mathbf{x}\}$.

Now suppose $\operatorname{rank} A_{\mathbf{x}} < n$. For convenience, reorder the rows of A so that $\mathbf{a}_1, \dots, \mathbf{a}_r$ are the rows of $A_{\mathbf{x}}$ and $\mathbf{a}_{r+1}, \dots, \mathbf{a}_m$ are the remaining rows; in particular

$$\mathbf{a}_j \cdot \mathbf{x} < b_j \quad \forall j \in [r+1, m]. \tag{7.4}$$

The system of equations $\{\mathbf{a}_i \cdot \mathbf{x} = b_i \colon i \in [r]\}$ defines an affine space of dimension $n - \operatorname{rank} A_{\mathbf{x}} > 0$. Let $\mathbf{v}$ be any vector parallel to that affine space (equivalently, perpendicular to each of $\mathbf{a}_1, \dots, \mathbf{a}_m$). Then $\mathbf{a}_i \cdot (\mathbf{x} + \epsilon\mathbf{v}) = b_i$ for any $\epsilon \in \mathbb{R}$. By continuity and (7.4), we can choose $\epsilon > 0$ small enough that $\mathbf{a}_j \cdot (\mathbf{x} \pm \epsilon\mathbf{v}) < b_j$ for all $j > r$. Then $\mathbf{x}' = \mathbf{x} + \epsilon\mathbf{v}$ and $\mathbf{x}'' = \mathbf{x} + \epsilon\mathbf{v}$ belong to P, and for any linear functional ℓ, either

$$\ell(\mathbf{x}') \leq \ell(\mathbf{x}) \leq \ell(\mathbf{x}'') \quad \text{or} \quad \ell(\mathbf{x}'') \leq \ell(\mathbf{x}) \leq \ell(\mathbf{x}')$$

(depending on the sign of $\ell(\mathbf{v})$), so that $\mathbf{x}$ cannot be the unique maximum of ℓ on P. □

It follows that every vertex is of the form $A_R^{-1} b_R$, where R is a row basis of A and A_R, b_R denote restrictions. Not every point of this form necessarily lies in P, but this argument does show that the vertex set $V(P)$ of every polyhedron P is finite.

Claim 2: Every polytope is the convex hull of its vertex set.

Proof of Claim 2. Induct on dimension n. If $n = 0$, then $P = V(P)$, while if $n = 1$, then P is a line segment, which is the convex hull of its two endpoints.

In general, let $\mathbf{x}$ be a point that is not a vertex, and let $F_{\mathbf{x}}$ be the unique minimal face of P containing $\mathbf{x}$. By assertion (3) of Prop. 1.26, $\mathbf{x}$ is in the relative interior of $F_{\mathbf{x}}$, so it is a convex combination

$$\mathbf{x} = \sum_{i=1}^{\ell} c_i \mathbf{y}_i, \qquad 0 \leq c_i \leq 1, \ \sum_{i=1}^{\ell} c_i = 1 \tag{7.5}$$

with $\mathbf{y}_1, \dots, \mathbf{y}_\ell \in \partial F_{\mathbf{x}}$. Meanwhile, $\partial F_{\mathbf{x}}$ is a union of faces of $F_{\mathbf{x}}$ (hence of P) of dimension $< n$. By induction, each $\mathbf{y}_i$ is a convex combination of vertices of $F_{\mathbf{x}}$. Assertion (1) of Prop. 1.26 implies in particular that every vertex of a face of P is a vertex of P, so we can write

$$\mathbf{y}_i = \sum_{j=1}^{k} b_{ij} \mathbf{v}_j, \qquad 0 \leq b_{ij} \leq 1, \ \sum_{j=1}^{k} b_{ij} = 1 \tag{7.6}$$

for each $i \in [\ell]$. Plugging (7.6) into (7.5) gives

$$\mathbf{x} \;=\; \sum_{i=1}^{\ell} c_i \sum_{j=1}^{k} b_{ij}\mathbf{v}_j \;=\; \sum_{j=1}^{k} \left(\sum_{i=1}^{\ell} c_i b_{ij} \right) \mathbf{v}_j. \tag{7.7}$$

Note that

$$0 \le \sum_{i=1}^{\ell} c_i b_{ij} \le \sum_{i=1}^{\ell} c_i \le 1$$

for all j, and that

$$\sum_{j=1}^{k}\sum_{i=1}^{\ell} c_i b_{ij} \;=\; \sum_{i=1}^{\ell} c_i \sum_{j=1}^{k} b_{ij} \;=\; \sum_{i=1}^{\ell} c_i \;=\; 1,$$

so formula (7.7) is an expression for $\mathbf{x}$ as a convex combination of the $\mathbf{v}$. (Summary of calculation: A convex combination of convex combinations is a convex combination.) □

At this point, we have shown that every polytope is a P-polytope. Part 2 of the proof is to show the converse.

Let $P \subset \mathbb{R}^n$ be a polytope. By Part 1 of the proof, we know that P has finitely many vertices $\mathbf{v}_1, \ldots, \mathbf{v}_r$ and we can write $P = \mathrm{conv}(\mathbf{v}_1, \ldots, \mathbf{v}_r)$. Assume without loss of generality that $\mathrm{aff}(P) = \mathbb{R}^n$ (otherwise, replace $\mathbb{R}^n$ with the affine hull) and that the origin is in the interior of P (translating if necessary), so that we can consider the polar dual $P^* = \{\mathbf{y} \in \mathbb{R}^n \mid \mathbf{x} \cdot \mathbf{y} \le 1 \;\; \forall \mathbf{x} \in P\}$ (see Definition 7.1.6).

Claim 3:

$$P^* = \{\mathbf{y} \in \mathbb{R}^n \mid \mathbf{v}_i \cdot \mathbf{y} \le 1 \;\; \forall i \in [r]\}.$$

Proof of Claim 3. Let $\widetilde{P^*}$ be the set on the RHS of Claim 3. Clearly $\widetilde{P^*} \supseteq P^*$. On the other hand, suppose $\mathbf{y} \in \widetilde{P^*}$ and $\mathbf{x} \in P$. Write $\mathbf{x}$ as a convex combination of the points $\mathbf{v}_1, \ldots, \mathbf{v}_r$, i.e.,

$$\mathbf{x} = \sum_{i=1}^{r} c_i \mathbf{v}_i, \qquad 0 \le c_i \le 1, \quad \sum_{i=1}^{r} c_i = 1,$$

whence

$$\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^{r} c_i \mathbf{v}_i \cdot \mathbf{y} \le \sum_{i=1}^{r} c_i = 1 \quad \therefore \quad \mathbf{y} \in P^*.$$ □

This claim enables us to draw pictures of duals. For example, consider the polytope

$$P = \left\{ (x,y) \in \mathbb{R}^2 \mid \begin{array}{l} x \ge -2 \\ y \ge -1 \\ 3x + 4y \le 2 \end{array} \right\} = \mathrm{conv}\left\{(-2,2),\ (-2,-1),\ (2,-1)\right\}.$$

From the V-description of P and Claim 3, we can easily read off the H-description of the dual:

$$P^* = \left\{ (x,y) \in \mathbb{R}^2 \mid \begin{array}{l} -2x + 2y \le 1 \\ -2x - y \le 1 \\ 2x - y \le 1 \end{array} \right\}.$$

INSERT FIGURE

In particular, P^* is an intersection of finitely many half-spaces. So, by the first part of the theorem, P^* is a P-polytope, say $P^* = \text{conv}\{\mathbf{y}_1, \ldots, \mathbf{y}_s\}$. Meanwhile, the double dual $P^{**} = (P^*)^*$ is defined by

$$\begin{aligned} P^{**} &= \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \cdot \mathbf{y} \leq 1 \ \ \forall \mathbf{y} \in P^*\} \\ &= \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \cdot \mathbf{y}_j \leq 1 \ \ \forall j \in [s]\} \end{aligned} \tag{7.8}$$

where the second equality comes from Claim 3.

Claim 4: $P = P^{**}$.

Proof of Claim 4. First, we show that $P \subseteq P^{**}$. Let $\mathbf{x} \in P = \text{conv}(\mathbf{z}_1, \ldots, \mathbf{z}_r)$, say $\mathbf{x} = \sum_{i=1}^r c_i \mathbf{z}_i$, and let $j \in [s]$. Then

$$\mathbf{x} \cdot \mathbf{y}_j \ = \ \sum_{i=1}^r c_i \mathbf{z}_i \cdot \mathbf{y}_j \ \leq \ \sum_{i=1}^r c_i \ = \ 1$$

since $\mathbf{z}_i \cdot \mathbf{y}_j \leq 1$ for all i, j by definition of P^*. Therefore $\mathbf{x} \in P^{**}$.

Second, we show that $P^{**} \subseteq P$. Let $\mathbf{x} \in P^{**}$. If $\mathbf{x} \notin P$, then[1] there is a hyperplane H separating $\mathbf{x}$ from P. Let $\mathbf{a}$ be a normal vector to H pointing toward the halfspace containing $\mathbf{x}$; then there is a real number δ such that $\mathbf{a} \cdot \mathbf{x} > \delta > \mathbf{a} \cdot \mathbf{w}$ for all $\mathbf{w} \in P$. Since $\vec{\mathbf{0}} \in P$, $\delta > 0$; scaling $\mathbf{a}$ if necessary, we may assume $\delta = 1$, i.e.,

$$\mathbf{a} \cdot \mathbf{x} > 1 > \mathbf{a} \cdot \mathbf{w} \qquad \forall \mathbf{w} \in P. \tag{7.9}$$

Then $\mathbf{a} \in P^*$ by definition, so we can write $\mathbf{a}$ as a convex combination

$$\mathbf{a} = \sum_{j=1}^s b_j \mathbf{y}_j$$

with $0 \leq b_j \leq 1$ for all j, and $\sum_{j=1}^s b_j = 1$. Therefore

$$1 < \mathbf{a} \cdot \mathbf{x} = \sum_{j=1}^s b_j \mathbf{y}_j \cdot \mathbf{x} \leq \sum_{j=1}^s b_j = 1$$

(where the $<$ comes from (7.9) and the $\leq$ from the hypothesis $\mathbf{x} \in P^{**}$), a contradiction. □

Consequently, (7.8) expresses P as the intersection of finitely many half-spaces, and we have shown that the P-polytope P is in fact a polytope.

7.3 Examples and next steps

Now here comes some more lingo:

Definition 7.3.1. Let P be an n-dimensional polytope in $\mathbb{R}^n$.

- A **facet** of P is a face of codimension 1 (that is, dimension $n-1$). In this case there is a *unique* linear functional (up to scaling) that is maximized on F, given by the outward normal vector from P. Faces of codimension 2 are called **ridges** and faces of codimension 3 are sometimes called **peaks**.

[1] This seemingly obvious assertion is not so easy to prove, although it is true in more generality: if $S \subseteq \mathbb{R}^n$ is a convex set and $\mathbf{y} \notin S$, then there exists a hyperplane separating S from $\mathbf{y}$ — or equivalently a linear functional $\ell : \mathbb{R}^n \to \mathbb{R}$ such that $\ell(\mathbf{y}) > 0$ and $\ell(\mathbf{x}) < 0$ for all $\mathbf{x} \in S$. This is called Minkowski's Hyperplane Separation Theorem. It is equivalent to many other statements, including Farkas' Lemma.

- A **supporting hyperplane** of P is a hyperplane that meets P in a nonempty face.
- P is **simplicial** if every face is a simplex. For example, every 2-dimensional polytope is simplicial, but of the Platonic solids in $\mathbb{R}^3$, only the tetrahedron, octahedron and icosahedron are simplicial — the cube and dodecahedron are not. The boundary of a simplicial polytope is thus a simplicial $(n-1)$-sphere.
- P is **simple** if every vertex belongs to exactly n faces. (In fact no vertex can belong to fewer than n faces.)

Proposition 7.3.2. *1. $\mathscr{F}(P^*) = \mathscr{F}(P)^*$ (i.e., the dual of $\mathscr{F}(P)$ in the sense of Definition 1.1.13).*
2. A polytope P is simple if and only if its dual P^* is simplicial.
3. A polytope is both simple and simplicial if and only if it is a simplex.

Proof. **To be written.** □

7.4 Shelling simplicial polytopes

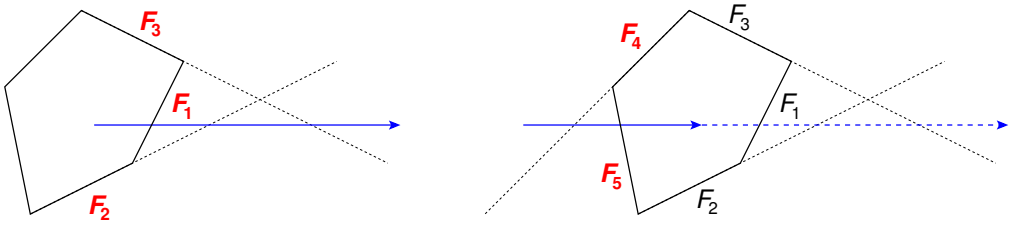
One of the big questions about polytopes is to classify their possible f-vectors and, more generally, the structure of their face posets. Here is a result of paramount importance.

Theorem 7.4.1. *Let Δ be the boundary sphere of a convex simplicial polytope $P \subseteq \mathbb{R}^d$. Then Δ is shellable, and its h-vector is a palindrome, i.e., $h_i = h_{d-i}$ for all i.*

These equations are the **Dehn-Sommerville relations**. They were first proved early in the 20th century, but the following proof, due to Bruggesser and Mani [BM71], is undoubtedly the one in the Book.

Sketch of proof. Let $F_1, \ldots, F_n$ be the facets of P and let $A_i = \operatorname{aff}(F_i)$. Let ℓ be a line that passes through the interior of P and meets the hyperplanes A_i in n distinct points. (Note that almost any line will do.) Imagine walking along ℓ, starting inside P. When you get to infinity, Stage 1 ends and Stage 2 starts by "hopping" to the other side of the line and come back the other way until you get back to inside P. Relabel the facets in the order that you encounter their affine spans. Let A_m be the last affine span you cross before "hopping".

As you keep walking, keep your eyes on P. In Stage 1, after crossing A_1, all you can see is F_1, but for each $i \in \{2, \ldots, m\}$, the facet F_i pops into view as soon as you cross A_i. (You have probably already seen some of its boundary, but not the entire facet.) In Stage 2, facets $A_{m+1}, \ldots, A_n$ are visible, but after you cross A_i the facet F_i disappears from view (although some of its boundary may still be visible). Finally, just before you cross A_n and enter P again, all you can see is F_n.



Stage 1 Stage 2

In fact $F_1, \ldots, F_n$ is a shelling order (called a **line shelling**), because

$$\langle F_j \rangle \cap \langle F_1, \ldots, F_{j-1} \rangle = \begin{cases} \langle \text{ridges of } F_j \text{ that are visible before crossing } A_j \rangle & \text{for } 2 \leq j \leq m, \\ \langle \text{ridges of } F_j \text{ that are invisible after crossing } A_j \rangle & \text{for } m+1 \leq j \leq n \end{cases}$$

Alternatively, $\langle F_j \rangle \setminus \langle F_1, \ldots, F_{j-1} \rangle = [R_i, F_i]$. where

$$R_i = \begin{cases} \text{smallest face of } F_j \text{ that is invisible before crossing } A_j & \text{for } 1 \leq j \leq m, \\ \text{smallest face of } F_j \text{ that remains visible after crossing } A_j & \text{for } m+1 \leq j \leq n. \end{cases}$$

(This assertion does need to be checked.) A shelling of P coming from a line in this way is called a **line shelling**. Moreover, since every ridge belongs to exactly two facets, we observe that each facet F_i contributes to $h_k(P)$, where

$$k = k(F_i) = \#\{j < i \colon F_i, F_j \text{ have a common ridge}\}.$$

On the other hand, the **reversal** of this shelling order is also a line shelling (by traversing ℓ in the opposite direction). Since each facet shares a ridge with exactly d other facets (because P is simplicial!), the previous formula says that if a facet contributes to h_i with respect to the original shelling order, then it contributes to h_{d-i} in the reverse shelling order. The h-vector is an invariant of P, so it follows that $h_i = h_{d-i}$ for all i. □

The Dehn-Sommerville relations are a basic tool in classifying h-vectors, and therefore f-vectors, of simplicial polytopes. Since $h_0 = 1$ for shellable complexes, it follows immediately that the only possible h-vectors for simplicial polytopes in $\mathbb{R}^2$ and $\mathbb{R}^3$ are $(1, k, 1)$ and $(1, k, k, 1)$, respectively (where k is a positive integer), and in particular the number of facets determines the h-vector (which is not the case in higher dimensions).

7.5 The normal fan and generalized permutahedra

Recall from Definition 7.1.4 that a face of a polyhedron $P \subset \mathbb{R}^n$ is defined as the subset of P that maximizes a linear functional. We can get a lot of mileage out of classifying linear functionals by *which* face of P they maximize. The resulting structure $\mathcal{N}(P)$ is called the **normal fan** of P. (Technical note: officially $\mathcal{N}(P)$ is a structure on the dual space $(\mathbb{R}^n)^*$, but we typically identify $(\mathbb{R}^n)^*$ with $\mathbb{R}^n$ by declaring the standard basis to be orthonormal — equivalently, letting each vector in $\mathbb{R}^n$ act by the standard dot product.)

Given a face $F \subset P$, let σ_F be the collection of linear functionals maximized on F. As we will see, the sets σ_F are in fact the interiors of **cones** (convex unions of rays from the origin).

Example 7.5.1. Let $P = \text{conv}\{(1,1), (1,-1), (-1,1)\} \subset \mathbb{R}^2$. The polytope and its normal fan are shown below.

The word "fan" means "collection of cones". Multiplying a linear functional by a positive scalar does not change the face on which it is maximized, and that if ℓ and ℓ' are linear functionals maximized on the same face, then so is every functional $a\ell + b\ell'$, where a, b are positive scalars. Therefore, each σ_F is a cone. The vertices x, y, z correspond to the 2-dimensional cones, the edges Q, R, S to 1-dimensional cones (a.k.a. rays) and the polytope P itself to the trivial cone consisting of the origin alone. In general, if F is a face of a polytope $P \subseteq \mathbb{R}^n$, then

$$\dim \sigma_F = n - \dim F. \tag{7.10}$$

◀

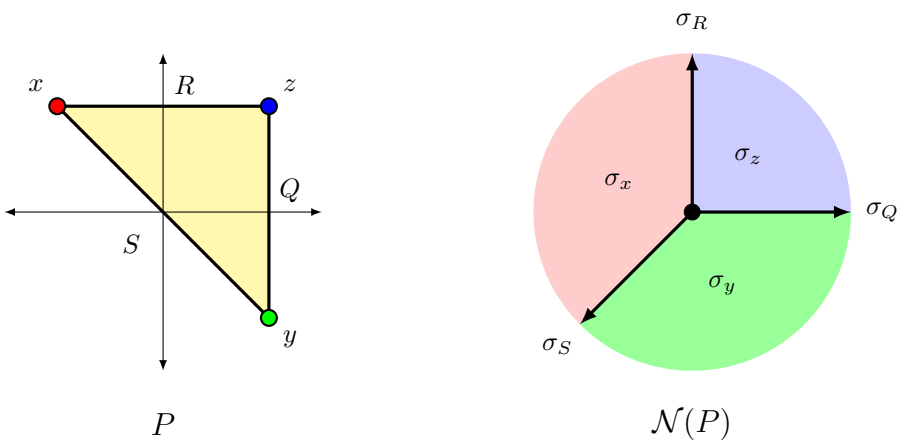
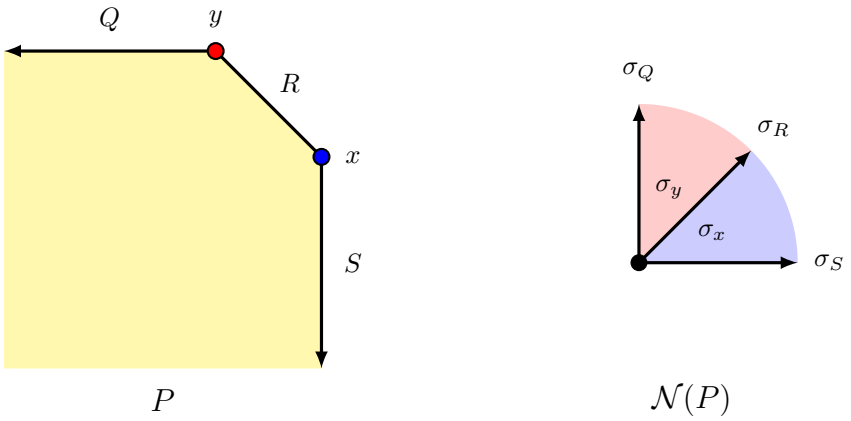


Figure 7.1: A polytope P and its normal fan $\mathcal{N}(P)$.

Example 7.5.2 (The normal fan of an unbounded polyhedron)**.** Let P be the unbounded polyhedron defined by the inequalities $x \leq 1$, $y \leq 1$, $x + y \leq 1$ (so its vertices are $x = (0, 1)$ and $y = (1, 0)$). The polytope and its normal fan are shown below.



This normal fan is *incomplete*: it does not cover every linear functional in $(\mathbb{R}^2)^*$, only the ones that have a well-defined maximum on P (in this case, those in the first quadrant). It is not hard to see that the normal fan of a polyhedron is complete if and only if the polyhedron is bounded, i.e., a polytope. The dimension formula for normal cones (7.10) is still valid in the unbounded case. ◀

In general the normal fan of a polytope can be quite complicated, and there exist fans in $\mathbb{R}^n$ that are not the normal fans of any polytope, even for $n = 3$; see, e.g., [Zie95, Example 7.5]. However, for some polytopes, we can describe the normal fan using other combinatorics, such as the following important class.

Definition 7.5.3. A polytope $P \subseteq \mathbb{R}^n$ is a **generalized permutahedron** if its normal fan is a coarsening of the braid fan (i.e., the fan of faces of the braid arrangement). Equivalently, for every linear functional $\ell(\mathbf{x}) = a_1x_1 + \cdots + a_nx_n$ the face of P maximized by ℓ is determined solely by the equalities and inequalities among the coefficients a_i.

The theory of generalized permutahedra is usually considered to have started with Postnikov's paper [Pos09]; other important sources include [PRW08] and [AA17]. Edmonds [Edm70] considered equivalent objects earlier under the name "polymatroids" **(add details)**.

Theorem 7.5.4. *A polytope $P \subseteq \mathbb{R}^n$ is a generalized permutahedron if and only if every edge of P is parallel to $\mathbf{e}_i - \mathbf{e}_j$ for some i, j, where $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is the standard basis.*

Generalized permutahedra can also be described as certain degenerations of the **standard permutahedron**, which is the convex hull of the vectors $(w_1, \ldots, w_n)$, where w ranges over all permutations of $[n]$. The normal fan of the standard permutahedron is precisely the braid fan.

One important family of generalized permutahedra are **matroid base polytopes**. Given a matroid M on ground set $[n]$, let P be the convex hull of all characteristic vectors of bases of M. It turns out that P is a generalized permutahedron; in fact, the matroid base polytopes are *exactly* the generalized permutahedra whose vertices are $0/1$ vectors [GGMS87, Thm. 4.1]. Describing the faces of matroid polytopes in terms of the combinatorics of the matroid is an interesting and difficult problem; see [FS05].

7.6 Ehrhart theory (contributed by Margaret Bayer)

The central problem considered in this section is the following: **How many integer or rational points are in a convex polytope?** An excellent and comprehensive source is [BR15]. There is some material on the case of a rational polytope in [Sta12, §4.6.2].

Definition 7.6.1. A polytope $P \subseteq \mathbb{R}^N$ is **integral** (resp. **rational**) if and only if all vertices of P have integer (resp. rational) coordinates.

For a set $P \subseteq \mathbb{R}^N$ and a positive integer n, let $nP = \{nx : x \in P\}$. (nP is called a **dilation** of P.)

The **(relative) boundary** of P, written ∂P, is the union of proper faces of P, that is, the set of points $x \in P$ such that for every $\varepsilon > 0$, the ball of radius ε (its intersection with $\operatorname{aff}(P)$) contains both points of P and points not in P.

For a polytope $P \subseteq \mathbb{R}^N$ define sequences

$$i(P, n) = |nP \cap \mathbb{Z}^N|$$
$$i^*(P, n) = |n(\operatorname{relint} P) \cap \mathbb{Z}^N|$$

$i(P, n)$ is the number of integer points in nP or, equivalently, the number of rational points in P of the form $\left(\frac{a_0}{n}, \frac{a_1}{n}, \ldots, \frac{a_N}{n}\right)$. Our goal is to understand the functions $i(P, n)$ and $i^*(P, n)$.

We start with P a simplex, and with an easy example. Let

$$P = \operatorname{conv}\{(0,0,0), (1,1,0), (1,0,1), (0,1,1)\} \in \mathbb{R}^3.$$

Then

$$nP = \operatorname{conv}\{(0,0,0), (n,n,0), (n,0,n), (0,n,n)\}.$$

Each point in nP can be written as $\beta_1(n,n,0) + \beta_2(n,0,n) + \beta_3(0,n,n) + \beta_4(0,0,0)$, with $0 \leq \beta_i \leq 1$ and $\sum \beta_i = 1$; or, alternatively, as $\alpha_1(1,1,0) + \alpha_2(1,0,1) + \alpha_3(0,1,1)$, with $0 \leq \alpha_i \leq n$ and $\sum \alpha_i \leq n$.

Case 1. If the α_i are all integers, the resulting points are integer points and the sum of the coordinates is even. How many such points are there? The answer is the number of monomials in four variables of degree n, that is, $\binom{n+3}{3}$. However, there are other integer points in nP.

Case 2. We can allow the fractional part of α_i to be $1/2$. If any one of the α_i has fractional part $1/2$, the others must be also. Writing $\gamma_i = \alpha_i - 1/2$, we get points of the form

$$\begin{aligned}&(\gamma_1 + 1/2)(1,1,0) + (\gamma_2 + 1/2)(1,0,1) + (\gamma_3 + 1/2)(0,1,1)\\ &= \gamma_1(1,1,0) + \gamma_2(1,0,1) + \gamma_3(0,1,1) + (1,1,1).\end{aligned}$$

Note here that $\sum \gamma_i = (\sum \alpha_i) - 3/2 \leq n - 3/2$. Since the γ_i are integers, $\sum \gamma_i \leq n - 2$. So the number of these points equals the number of monomials in four variables of degree $n-2$, that is, $\binom{n+1}{3}$.

Adding these we get

$$i(P,n) = \binom{n+3}{3} + \binom{n+1}{3} = \frac{1}{3}n^3 + n^2 + \frac{5}{3}n + 1.$$

Note, in particular, that this is a polynomial in n.

And what about the number of integer points in the interior of P?

Note that all the points in Case 2 are interior points because each $\alpha_i = \gamma_i + 1/2 > 0$ and their sum is at most $n - 2 + 3/2$ (less than n). A point in Case 1 is an interior point if and only if all the $\alpha_i > 0$ and $\sum \alpha_i < n$. The four-tuples $(\alpha_1 - 1, \alpha_2 - 1, \alpha_3 - 1, n - 1 - \sum \alpha_i)$ correspond to monomials in four variables of degre $n-4$; there are $\binom{n-1}{3}$ of them. Thus we get

$$i^*(P,n) = \binom{n+1}{3} + \binom{n-1}{3} = \frac{1}{3}n^3 - n^2 + \frac{5}{3}n - 1,$$

another polynomial. (Anything else you notice? Is it a coincidence?)

It is convenient to visualize the dilations nP of P in a cone. For $P \subseteq \mathbb{R}^N$ an integral N-simplex, let $\tilde{P} = \{(x,1) \in \mathbb{R}^{N+1} : x \in P\}$, and let C be the simplicial cone generated by $\tilde{P}$:

$$C = C(\tilde{P}) = \{ry : y \in \tilde{P}, r \in \mathbb{R}, r \geq 0\}.$$

The boundary and interior of C are $\partial C = \{ry : y \in \partial\tilde{P}\}$ and $\operatorname{relint} C = C \setminus \partial C$. Then the polytope nP can be identified with a cross-section of C:

$$C \cap \{(z,n) \in \mathbb{R}^{N+1} : z \in \mathbb{R}^N\} = \{(z,n) \in \mathbb{R}^{N+1} : z \in nP\}.$$

The integer point functions are then

$$\begin{aligned}i(P,n) &= |C \cap \{(z,n) \in \mathbb{R}^{N+1} : z \in \mathbb{Z}^N\}|\\ i^*(P,n) &= |\operatorname{relint} C \cap \{(z,n) \in \mathbb{R}^{N+1} : z \in \mathbb{Z}^N\}|.\end{aligned}$$

We can represent all points in the cone in terms of the vertices of P.

Proposition 7.6.2. *Let P be a rational N-simplex in $\mathbb{R}^N$, with vertices v_0, v_1, ..., v_N, and let $C = C(\tilde{P})$. A point $z \in \mathbb{R}^{N+1}$ is a rational point in C if and only if $z = \sum_{i=0}^N c_i(v_i, 1)$ for some nonnegative rational numbers c_i. Furthermore, this representation of z is unique.*

A slightly different representation is more useful. Let

$$Q = \left\{ \sum_{i=0}^N r_i(v_i,1) \colon 0 \leq r_i < 1 \quad \forall i \right\}.$$

Thus Q is a half-open parallelepiped containing 0 and $\tilde{P}$.

Proposition 7.6.3. *Let P be an integral N-simplex in $\mathbb{R}^N$, with vertices v_0, v_1, …, v_N, and let $C = C(\tilde{P})$. A point $z \in \mathbb{Z}^{N+1}$ is an integer point in C if any only if $z = y + \sum_{i=0}^{N} r_i(v_1, 1)$ for some $y \in Q \cap \mathbb{Z}^{N+1}$ and some nonnegative integers r_i. Furthermore, this representation of z is unique.*

So to count integer points in C (and hence to determine $i(P, n)$), we only need to know how many integer points are in Q with each fixed (integer) last coordinate. We call the last coordinate of $z \in Q$ the **degree** of z. Note that for $z \in Q$, $\deg z = \sum_{i=0}^{N} r_i$ for some r_i, $0 \le r_i < 1$, so if $\deg z$ is an integer, $0 \le \deg z \le N$.

Theorem 7.6.4. *Let P be an integral N-simplex in $\mathbb{R}^N$, with vertices v_0, v_1, …, v_N, let $C = C(\tilde{P})$, and let $Q = \{\sum_{i=0}^{N} r_i(v_i, 1) : \text{for each } i,\ 0 \le r_i < 1\}$. Let δ_j be the number of points of degree j in $Q \cap \mathbb{Z}^{N+1}$. Then*

$$\sum_{n=0}^{\infty} i(P, n)\lambda^n = \frac{\delta_0 + \delta_1\lambda + \cdots + \delta_N\lambda^N}{(1-\lambda)^{N+1}}.$$

Corollary 7.6.5. *For P an integral N-simplex, $i(P, n)$ is a polynomial in n.*

Proof.

$$\begin{aligned}\sum_{n=0}^{\infty} i(P, n)\lambda^n &= (\delta_0 + \delta_1\lambda + \cdots + \delta_N\lambda^N)(1 + \lambda + \lambda^2 + \cdots)^{N+1} \\ &= (\delta_0 + \delta_1\lambda + \cdots + \delta_N\lambda^N)\left(\sum_{k-0}^{\infty} \binom{k+N}{N}\lambda^k\right).\end{aligned}$$

The coefficient of λ^n on the right hand side is $\sum_{j=0}^{N} \delta_j \binom{n-j+N}{N}$. □

For the interior of P (and of C) we use an analogous construction, but with the opposite half-open parallelipiped. Let

$$Q^* = \left\{\sum_{i=0}^{N} r_i(v_i, 1) \colon 0 < r_i \le 1 \forall i\right\}.$$

Proposition 7.6.6. *Let P be an integral N-simplex in $\mathbb{R}^N$, with vertices v_0, v_1, …, v_N, and let $C = C(\tilde{P})$. A point $z \in \mathbb{Z}^{N+1}$ is an integer point in $\operatorname{relint} C$ if and only if $z = y + \sum_{i=0}^{N} c_i(v_1, 1)$ for some $y \in Q^* \cap \mathbb{Z}^{N+1}$ and some nonnegative integers c_i. Furthermore, this representation of z is unique.*

So to count integer points in $\operatorname{relint} C$ (and hence to determine $i^*(P, n)$), we only need to know how many integer points are in Q^* with each fixed (integer) last coordinate. Note that for $z \in Q^*$, $0 < \deg z \le N + 1$.

Theorem 7.6.7. *Let P be an integral N-simplex in $\mathbb{R}^N$, with vertices v_0, v_1, …, v_N, let $C = C(\tilde{P})$, and let $Q^* = \{\sum_{i=0}^{N} r_i(v_i, 1) : \text{for each } i,\ 0 < r_i \le 1\}$. Let δ_j^* be the number of points of degree j in $Q^* \cap \mathbb{Z}^{N+1}$. Then*

$$\sum_{n=0}^{\infty} i^*(P, n)\lambda^n = \frac{\delta_1^*\lambda + \delta_2^*\lambda^2 + \cdots + \delta_{N+1}^*\lambda^{N+1}}{(1-\lambda)^{N+1}}.$$

Corollary 7.6.8. *For P an integral N-simplex, $i^*(P, n)$ is a polynomial in n.*

Now the punchline is that there is an easy relationship between the δ_i and the δ_i^*. Note that

$$\begin{aligned} Q^* &= \left\{ \sum_{i=0}^{N} r_i(v_i, 1) : \text{for each } i,\, 0 < r_i \le 1 \right\} \\ &= \left\{ \sum_{i=0}^{N} (1 - t_i)(v_i, 1) : \text{for each } i,\, 0 \le t_i < 1 \right\} \\ &= \left\{ \sum_{i=0}^{N} (v_i, 1) - \sum_{i=0}^{N} t_i(v_i, 1) : \text{for each } i,\, 0 \le t_i < 1 \right\} \\ &= \sum_{i=0}^{N} (v_i, 1) - Q = \left(\sum_{i=0}^{N} v_i, N + 1 \right) - Q \end{aligned}$$

An element of $Q^* \cap \mathbb{Z}^{N+1}$ of degree k corresponds to an element of $Q \cap \mathbb{Z}^{N+1}$ of degree $N + 1 - k$. Thus $\delta_k^* = \delta_{N+1-k}$.

Theorem 7.6.9. *If P is an integral N-simplex in $\mathbb{R}^N$, then*

$$F(P, \lambda) := \sum_{n=0}^{\infty} i(P, n)\lambda^n = \frac{\delta_0 + \delta_1\lambda + \cdots + \delta_N\lambda^N}{(1 - \lambda)^{N+1}}$$

$$F^*(P, \lambda) := \sum_{n=0}^{\infty} i^*(P, n)\lambda^n = \frac{\delta_N\lambda + \delta_{N-1}\lambda^2 + \cdots + \delta_0\lambda^{N+1}}{(1 - \lambda)^{N+1}}.$$

Thus

$$F^*(P, \lambda) = (-1)^{N+1} F(P, 1/\lambda).$$

This relationship is known as **Ehrhart reciprocity**.

So far I have considered only integral simplices. To extend the result to integral polytopes requires triangulation of the polytope, that is, subdivision of the polytope into simplices. The extension is nontrivial. We cannot just add up the functions i and i^* for the simplices in the triangulation, since interior points of the polytope can be contained in the boundary of a simplex of the triangulation, and in fact in the boundary of more than one simplex of the triangulation. But it works in the end.

Theorem 7.6.10. *Let $P \subseteq \mathbb{R}^N$ be an integral polytope of dimension N. Then*

$$(1 - \lambda)^{N+1} \sum_{i=0}^{\infty} i(P, n)\lambda^n$$

is a polynomial in λ of degree at most N.

As before, write this polynomial as $\sum_{j=0}^{N} \delta_j \lambda^j$. What can we say about the coefficients δ_j?

$\delta_0 = i(P, 0) = 1$, since this is the number of integer points in the polytope $0P = \{0\}$.

$\delta_1 + (N + 1)\delta_0 = i(P, 1)$, so $\delta_1 = |P \cap \mathbb{Z}^N| - (N + 1)$.

Also, recall that $i(P, n) = \sum_{j=0}^{N} \delta_j \binom{n-j+N}{N}$. Let C be the leading coefficient of $i(P, n)$ as a polynomial in n, i.e.,

$$C = \frac{1}{N!} \sum_{j=0}^{N} \delta_j = \lim_{n \to \infty} \frac{i(P, n)}{n^N}.$$

I claim C is the volume of P. To see this, note that $\mathrm{vol}(nP) = n^N \mathrm{vol}(P)$ (if P is of full dimension N). Now the volume of nP can be estimated by the number of lattice points in nP, that is, by $i(P,n)$. In fact,

$$0 = \lim_{n\to\infty} \frac{i(P,n) - \mathrm{vol}(nP)}{n^N} = \lim_{n\to\infty} \frac{i(P,n)}{n^N} - \mathrm{vol}(P).$$

So $C = \lim_{n\to\infty} \dfrac{i(P,n)}{n^N} = \mathrm{vol}(P)$.

One last comment. The Ehrhart theory can be generalized to rational polytopes. In the more general case, the functions $i(P,n)$ and $i^*(P,n)$ need not be polynomials, but are **quasipolynomials**—restricted to a congruence class in some modulus (depending on the denominators occurring in the coordinates of the vertices) they are polynomials. An equivalent description is that the function $i(P,n)$ is a polynomial in n and expressions of the form $\gcd(n,k)$, e.g.,

$$i(P,n) = \begin{cases} (n+1)^2 & n \text{ even} \\ n^2 & n \text{ odd} \end{cases} = (n + \gcd(n,2) - 1)^2.$$

7.7 Exercises

Problem 7.1. Prove that the topological boundary of a polyhedron is the union of its proper faces.

Problem 7.2. Prove that the convex hull of a finite point set $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \subseteq \mathbb{R}^d$ is the set of convex linear combinations of it.

Problem 7.3. Prove Theorem 7.5.4.

Problem 7.4. Let M be a matroid and let P be its base polytope. Prove that P is a generalized permutahedron in two different ways:

(a) Show that the normal fan $\mathcal{N}_P$ coarsens the braid cone, using the properties of greedy algorithms.
(b) Show that every edge of P is parallel to some $\mathbf{e}_i - \mathbf{e}_j$, using the properties of basis exchange.

Problem 7.5. Let Δ^n be the n-dimensional simplex whose vertices are the n standard basis vectors in $\mathbb{R}^n$, together with the origin. That is,

$$\Delta^n = \{\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n : \ 0 \le x_i \le 1, \ 0 \le x_1 + \cdots + x_n \le 1\}.$$

Calculate the Ehrhart polynomials $i(\Delta^n, k)$ and $i^*(\Delta^n, k)$.

Problem 7.6. The **crosspolytope** is defined as

$$X_n = \mathrm{conv}\{\pm\mathbf{e}_1, \ldots, \pm\mathbf{e}_n\} \subseteq \mathbb{R}^n\} = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : \ |x_i| \le 1, \ 0 \le |x_1| + \cdots + |x_n| \le 1\}.$$

Calculate $i(X_n, k)$ and $i^*(X_n, k)$.

Chapter 8

Group Representations

8.1 Basic definitions

Definition 8.1.1. Let G be a group (typically finite) and let $V \cong \Bbbk^n$ be a finite-dimensional vector space over a field $\Bbbk$. A **representation of G on V** is a group homomorphism $\rho : G \to GL(V)$, where $GL(V) = GL_n(\Bbbk)$ is the group of linear automorphisms of V, or equivalently the group of invertible $n \times n$ matrices over $\Bbbk$. That is, for each $g \in G$ there is an invertible $n \times n$ matrix $\rho(g)$, satisfying

$$\rho(g)\rho(h) = \rho(gh) \qquad \forall g, h \in G.$$

(That's matrix multiplication on the left side of the equation, and group multiplication in G on the right.) The number n is called the **dimension** (sometimes **degree**) of the representation, written $\dim \rho$.

Some remarks:

- ρ specifies an action of G on V that respects its vector space structure. So we have all the accoutrements of group actions, such as orbits and stabilizers. If there is only one representation under consideration, it is often convenient to use group-action notation and write gv (or $g \cdot v$, $g(v)$, etc.) instead of the bulkier $\rho(g)v$.
- It is common to say that ρ is a representation, or that V is a representation, or that the pair (ρ, V) is a representation.
- ρ is **faithful** if it is injective as a group homomorphism.

Example 8.1.2. Let G be any group. The **trivial representation** is the map

$$\rho_{\text{triv}} : G \to GL_1(\Bbbk) \cong \Bbbk^\times$$

sending $g \mapsto 1$ for all $g \in G$. (This is as non-faithful as you can get.) ◀

Example 8.1.3. Let $\Bbbk G$ be the vector space of formal $\Bbbk$-linear combinations of elements of G: that is, $\Bbbk G = \left\{\sum_{h \in G} a_h h \colon a_h \in \Bbbk\right\}$. The **regular representation** of G is the map $\rho_{\text{reg}} : G \to GL(\Bbbk G)$ defined by

$$g\left(\sum_{h \in G} a_h h\right) = \sum_{h \in G} a_h (gh).$$

That is, g permutes the standard basis vectors of $\Bbbk G$ according to the group multiplication law. Thus $\dim \rho_{\text{reg}} = |G|$. This representation is faithful. ◀

The vector space $\Bbbk G$ is a ring, with multiplication given by multiplication in G and extended $\Bbbk$-linearly. In this context it is called the **group algebra** of G over $\Bbbk$.

Remark 8.1.4. A representation of G is equivalent to a (left) **module** over the group algebra $\Bbbk G$. Technically "representation" refers to the way G acts and "module" refers to the space on which it acts, but the two terms really carry the same information.

Example 8.1.5. Let $G = \mathfrak{S}_n$, the symmetric group on n elements. The **defining representation** ρ_{def} of G on $\Bbbk^n$ maps each permutation $\sigma \in G$ to the $n \times n$ permutation matrix with 1's in the positions $(i, \sigma(i))$ for every $i \in [n]$, and 0's elsewhere. This representation is faithful. ◀

Example 8.1.6. More generally, let G act on a finite set X. Then there is an associated **permutation representation** on $\Bbbk X$, the vector space with basis X, given by

$$g\left(\sum_{x\in X} a_x x\right) = \sum_{x\in X} a_x (gx).$$

For short, we might specify the action of G on X and say that it "extends linearly" to $\Bbbk X$. For instance, the action of G on itself by left multiplication gives rise in this way to the regular representation, and the usual action of $\mathfrak{S}_n$ on an n-element set gives rise to the defining representation. ◀

Example 8.1.7. Let $G = \mathbb{Z}/n\mathbb{Z}$ be the cyclic group of order n, and let $\zeta \in \Bbbk$ be a nth root of unity (not necessarily primitive). Then G has a 1-dimensional representation given by $\rho(x) = \zeta^x$. This representation is faithful if and only ζ is a primitive root of unity. In a sense, every representation of G is built from these 1-dimensional representations, as we will see later (§8.8). ◀

Example 8.1.8. Consider the dihedral group D_n of order $2n$, i.e., the group of symmetries of a regular n-gon, given in terms of generators and relations by

$$\langle s, r \ : \ s^2 = r^n = 1, \ srs = r^{-1} \rangle.$$

There are several natural representations of D_n. Here are a few:

1. Regarding s as a reflection and r as a rotation in $\mathbb{R}^2$ gives a faithful 2-dimensional representation, the **geometric representation**:

$$\rho_{\text{geo}}(s) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \qquad \rho_{\text{geo}}(r) = \begin{bmatrix} \cos(2\pi/n) & \sin(2\pi/n) \\ -\sin(2\pi/n) & \cos(2\pi/n) \end{bmatrix}$$

   extended by group multiplication (e.g., $\rho_{\text{geo}}(sr^2) = \rho_{\text{geo}}(s)\rho_{\text{geo}}(r)^2$, etc.).
2. We can regard D_n as the symmetries of a regular n-gon with vertices labeled $1, 2, \ldots, n$ in cyclic order. We can take r to be the rotation by $2\pi/n$, We can take s to be the reflection across the line L through the center and vertex n. Note that if n is even then L also passes through vertex $n/2$, while if n is odd then L passes through the midpoint of the side with vertices $\lfloor n/2 \rfloor, \lceil n/2 \rceil$. Thus the action is given by the homomorphism $D_n \to \mathfrak{S}_n$ defined by

$$r \mapsto (1\ 2\ \cdots\ n), \qquad s \mapsto (1\ n-1)(2\ n-2)\cdots(\lfloor (n-1)/2 \rfloor\ \lceil (n-1)/2 \rceil)$$

   which gives rise, as in Example 8.1.6, to a faithful n-dimensional permutation representation ρ_V of D_n. (Questions to consider: What about the action of D_n on edges? Is it isomorphic to the action on vertices? What does "isomorphic" mean?)
3. The n-gon has n diameters (lines of reflection symmetry). The dihedral group acts on diameters and thus gives rise to another n-dimensional permutation representation. This representation is faithful if and only if n is odd. (If n is even, then the element $r^{n/2}$ acts by rotation by 180° and fixes all diameters.) ◀

Example 8.1.9. The symmetric group $\mathfrak{S}_n$ has another 1-dimensional representation, the **sign representation**, given by

$$\rho_{\text{sign}}(\sigma) = \begin{cases} 1 & \text{if } \sigma \text{ is even,} \\ -1 & \text{if } \sigma \text{ is odd.} \end{cases}$$

The sign representation is nontrivial provided $\operatorname{char} \Bbbk \neq 2$. Note that $\rho_{\text{sign}}(g) = \det \rho_{\text{def}}(g)$ (see Example 8.1.5). (More generally, if ρ is any representation, then $\det \rho$ is a 1-dimensional representation.) ◀

Example 8.1.10. Let (ρ, V) and (ρ', V') be representations of G. The **direct sum** $\rho \oplus \rho' : G \to GL(V \oplus V')$ is defined by

$$(\rho \oplus \rho')(g)(v + v') = \rho(g)(v) + \rho'(g)(v')$$

for $v \in V$, $v' \in V'$. In terms of matrices, $(\rho \oplus \rho')(g)$ is a block-diagonal matrix:

$$\left[\begin{array}{c|c} \rho(g) & 0 \\ \hline 0 & \rho'(g) \end{array}\right].$$

In particular, $\dim(\rho \oplus \rho') = \dim \rho + \dim \rho'$. ◀

Example 8.1.11. Similarly, if (ρ, V) and (ρ', V') are representations of *different* groups G, G', then there is a representation $\rho \times \rho'$ of the direct product $G \times G'$ on $V \times V'$, given by

$$(\rho \times \rho')(g, g')(v + v') = (\rho(g)(v), \rho'(g')(v')) \qquad \text{or} \qquad (\rho \times \rho')(g, g') = \left[\begin{array}{c|c} \rho(g) & 0 \\ \hline 0 & \rho'(g') \end{array}\right].$$

This construction looks superficially similar to Example 8.1.10 but really is different, hence the different notation. For the most part, we will be focusing on representations of a single group at a time. ◀

Having defined objects, we should now define subobjects.

Definition 8.1.12. Let (ρ, V) be a G-representation. A vector subspace $W \subseteq V$ is ***G*-invariant** if $gW \subseteq W$ for every $g \in G$. This condition is equivalent to $gW = W$ (because $\rho(g)$ is invertible, hence has no kernel) for every $g \in G$. Thus $(\rho|_W, W)$ is a representation of G (the **restriction** of ρ to W). In module language, W is a G-**submodule** of V. The subspaces 0 and V are trivially G-invariant; any other G-invariant subspace is **nontrivial**.

For example, both V and V' are G-invariant subspaces of $V \oplus V'$. For a more subtle example, let $G = \mathfrak{S}_n$ and let $(\rho_{\text{def}}, \Bbbk^n)$ be the defining representation. The one-dimensional subspace spanned by the vector $(1, 1, \ldots, 1)$ is fixed point wise by $\mathfrak{S}_n$, so it is a G-invariant subspace that carries the trivial representation.

8.2 Homomorphisms and isomorphisms of representations

When are two representations the same? More generally, what is a map between representations?

Definition 8.2.1. Let (ρ, V) and (ρ', V') be representations of G. A linear transformation $\phi : V \to V'$ is ***G*-equivariant**, or for short a ***G*-map**, if $\rho'(g) \cdot \phi(v) = \phi(\rho(g) \cdot v)$ for all $g \in G$ and $v \in V$. (more concisely, $g\phi = \phi g$ for all $g \in G$). An equivalent condition is that the following diagram commutes:

$$\begin{array}{ccc} V & \xrightarrow{\phi} & V' \\ {\scriptstyle \rho(g)}\downarrow & & \downarrow{\scriptstyle \rho'(g)} \\ V & \xrightarrow{\phi} & V' \end{array} \tag{8.1}$$

We sometimes use the notation $\phi\colon \rho \to \rho'$ for a G-equivariant map. In the language of modules, a G-equivariant transformation is the same thing as a homomorphism of G-modules.[1]

An **isomorphism** of G-representations is a G-equivariant map that is a vector space isomorphism. Clearly, being a vector space isomorphism is necessary for a G-equivariant map to be considered an isomorphism of G-modules. On the other hand, it is also sufficient, because the inverse of an invertible G-equivariant map is G-equivariant (proof left to the reader).

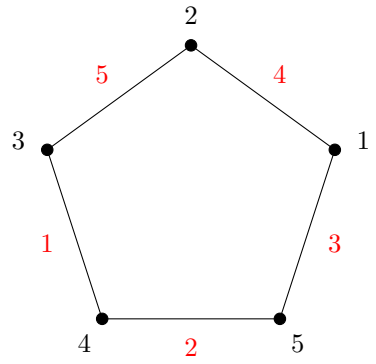
Example 8.2.2. Trivial(-ish) examples: The identity map $V \to V$ is an automorphism for any representation of G on V, and the zero map $V \to W$ is a homomorphism for any representations on V, W. ◀

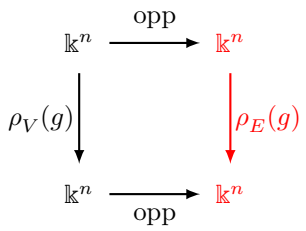
Example 8.2.3. Let $G = \mathfrak{S}_n$ act on $\Bbbk^n$ by the defining representation, and on $\Bbbk$ by the trivial representation. The map $\Bbbk^n \to \Bbbk$ given by

$$\phi\left(\sum_{i=1}^{n} a_i \mathbf{e}_i\right) = \sum_{i=1}^{n} a_i$$

is G-equivariant because permuting the coordinates of a vector does not change their sum. ◀

Example 8.2.4. Let n be odd, and consider the dihedral group D_n acting on a regular n-gon (see Example 8.1.8). Label the vertices $1, \dots, n$ in cyclic order. Label each edge the same as its opposite vertex, as in the figure on the left. Then the permutation action ρ_V on vertices is isomorphic to the action ρ_E on edges. In other words, the diagram on the right commutes for all $g \in D_n$, where "opp" is the map that sends the basis vector for a vertex to the basis vector for its opposite edge.





The case that n is even is trickier, because then each reflection either fixes two vertices or two edges, but not both. ◀

Example 8.2.5. Let $\mathbf{v}_1, \dots, \mathbf{v}_n$ be the points of a regular n-gon in $\mathbb{R}^2$ centered at the origin, e.g., $\mathbf{v}_j = \left(\cos\frac{2\pi j}{n}, \sin\frac{2\pi j}{n}\right)$. Then the map $\mathbb{R}^n \to \mathbb{R}^2$ sending the jth standard basis vector to $\mathbf{v}_j$ is D_n-equivariant, where D_n acts on $\mathbb{R}^n$ by permutation and on $\mathbb{R}^2$ via the geometric representation. ◀

Example 8.2.6. One way in which G-equivariant transformations occur is when one group action naturally gives rise to another action. For instance, consider the permutation action of $\mathfrak{S}_4$ on the vertices of K_4, which induces a representation ρ_V on the space $V = \Bbbk\langle \mathbf{v}_1, \dots, \mathbf{v}_4 \rangle \cong \Bbbk^4$. This action naturally determines an action on the six edges of K_4, which in turn induces a permutation representation ρ_E on $E = \Bbbk\langle \mathbf{e}_{12}, \dots, \mathbf{e}_{34} \rangle \cong \Bbbk^6$. The relation between the two actions can be described by a G-equivariant map — but be careful: it is not a map $V \to E$ but a map $E \to V$, namely

$$\phi(\mathbf{e}_{ij}) = \mathbf{v}_i + \mathbf{v}_j$$

[1] Technically this should be "$\Bbbk G$-modules", but this is a common shorthand.

so that $\rho_V \circ \phi(g) = \phi \circ \rho_E(g)$ for all g, i.e., the following diagram commutes:

$$\begin{array}{ccc} E & \xrightarrow{\phi} & V \\ {\scriptstyle \rho_E(g)}\downarrow & & \downarrow{\scriptstyle \rho_V(g)} \\ E & \xrightarrow{\phi} & V \end{array}$$

◀

Kernels and images of G-equivariant maps are well-behaved. Those familiar with modules will not be surprised: every kernel or image of a R-module homomorphism is also a R-module. In category-theoretic terms, the family of R-modules forms an abelian category.

Proposition 8.2.7. *Any G-equivariant map $\phi : (\rho, V) \to (\rho', V')$ has G-invariant kernel and G-invariant image.*

Proof. First, let $v \in \ker \phi$. Then $\phi(g \cdot v) = g \cdot \phi(v) = g \cdot 0 = 0$. So $g \cdot v \in \ker \phi$.

Second, let $w \in \operatorname{im} \phi$, say $w = \phi(v)$. Then $g \cdot w = g \cdot \phi(v) = \phi(g \cdot v) \in \operatorname{im} \phi$. □

Example 8.2.8. Let $\Bbbk$ be a field of characteristic $\neq 2$, and let $V = \Bbbk^2$, with standard basis $\{e_1, e_2\}$. Let $G = \mathfrak{S}_2 = \{\mathrm{id}, \mathrm{flip}\}$, where $\mathrm{flip} = (1\ 2)$ in cycle notation. The defining representation $\rho = \rho_{\mathrm{def}}$ of G on V is given by

$$\rho(\mathrm{id}) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad \rho(\mathrm{flip}) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

On the other hand, the representation $\sigma = \rho_{\mathrm{triv}} \oplus \rho_{\mathrm{sign}}$ on V (see Examples 8.1.2 and 8.1.9) is given by

$$\sigma(\mathrm{id}) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad \sigma(\mathrm{flip}) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

These two representations ρ and σ are in fact isomorphic. Indeed, ρ acts trivially on $\Bbbk\langle e_1 + e_2 \rangle$ and acts by the sign representation on $\Bbbk\langle e_1 - e_2 \rangle$. These two vectors form a basis of V (here is where we use the assumption $\operatorname{char} \Bbbk \neq 2$), and one can check that the change-of-basis map

$$\phi = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix}$$

is G-equivariant, hence an isomorphism $\rho \to \sigma$. ◀

Example 8.2.9. Consider the representations ρ_{triv} and ρ_{sign} of $\mathfrak{S}_2 = \{12, 21\}$. What are the G-equivariant maps between them? In other words, which linear maps $\phi : \Bbbk \to \Bbbk$ satisfy the diagram (8.1), where $\rho, \rho' \in \{\rho_{\mathrm{triv}}, \rho_{\mathrm{sign}}\}$? Each such map is of the form $\phi(v) = cv$ for some $c \in \Bbbk$.

If $\rho = \rho'$, then any linear transformation $\phi : \Bbbk \to \Bbbk$ (i.e., any map $\phi(v) = cv$ for some $c \in \Bbbk$) will do. Thus the set of G-equivariant homomorphisms is actually isomorphic to $\Bbbk$.

Assume $\operatorname{char} \Bbbk \neq 2$ (otherwise $\rho_{\mathrm{triv}} = \rho_{\mathrm{sign}}$ and we are done at this point). If $\phi : \rho_{\mathrm{triv}} \to \rho_{\mathrm{sign}}$ is G-equivariant, then we have diagrams

$$\begin{array}{ccc} V & \xrightarrow{\phi} & V' \\ {\scriptstyle \rho_{\mathrm{triv}}(12)}\downarrow & & \downarrow{\scriptstyle \rho_{\mathrm{sign}}(12)} \\ V & \xrightarrow{\phi} & V' \end{array} \qquad \begin{array}{ccc} V & \xrightarrow{\phi} & V' \\ {\scriptstyle \rho_{\mathrm{triv}}(21)}\downarrow & & \downarrow{\scriptstyle \rho_{\mathrm{sign}}(21)} \\ V & \xrightarrow{\phi} & V' \end{array}$$

The first diagram always commutes because $\rho_{\text{triv}}(12) = \rho_{\text{sign}}(12)$ is the identity map, but the second diagram says that for every $v \in \Bbbk$

$$\rho_{\text{sign}}(\phi(v)) = -\phi(v) = -cv, \qquad \phi(\rho_{\text{triv}}(v)) = \phi(v) = cv$$

and since $\operatorname{char} \Bbbk \neq 2$ we are forced to conclude that $c = 0$. Therefore, there is no nontrivial G-homomorphism $\rho_{\text{triv}} \to \rho_{\text{sign}}$. ◀

Example 8.2.9 is the tip of an iceberg: we can use the vector space $\operatorname{Hom}_G(\rho, \rho')$ of G-homomorphisms $\phi : \rho \to \rho'$ to measure how similar ρ and ρ' are.

8.3 Irreducibility, indecomposability, and Maschke's theorem

Definition 8.3.1. Let (ρ, V) be a representation of G.

- V is **decomposable** if there are nontrivial G-invariant subspaces $W, W^\perp \subsetneq V$ with $W \cap W^\perp = 0$ and $W + W^\perp = V$. Here $W^\perp$ is called the **(invariant) complement** of W in V. The notation does not presuppose the existence of a scalar product.
- V is **irreducible** (or **simple**, or colloquially an **irrep**) if it has no proper G-invariant subspace.
- A representation that can be decomposed into a direct sum of irreps is called **semisimple** or **completely reducible**. A semisimple representation is determined up to isomorphism by the multiplicity with which each isomorphism type of irrep appears.

Clearly, every representation can be written as the direct sum of indecomposable representations, and every irreducible representation is indecomposable. On the other hand, there exist indecomposable representations that are not irreducible.

Example 8.3.2. As in Example 8.2.8, let $V = \{e_1, e_2\}$ be the standard basis for $\Bbbk^2$, where $\operatorname{char} \Bbbk \neq 2$. Recall that the defining representation of $\mathfrak{S}_2 = \{\text{id}, \text{flip}\}$ is given by

$$\rho_{\text{def}}(\text{id}) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad \rho_{\text{def}}(\text{flip}) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

and the change-of-basis map

$$\phi = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix}^{-1}$$

is a G-equivariant isomorphism $\rho_{\text{def}} \to \rho_{\text{triv}} \oplus \rho_{\text{sign}}$. On the other hand, if $\operatorname{char} \Bbbk = 2$, then the matrix ϕ is not invertible and this argument breaks down. In fact, in this case $\Bbbk\langle e_1 + e_2 \rangle$ is the *only* proper G-invariant subspace of V, and consequently ρ_{def} is not semisimple. ◀

Fortunately, we can rule out this kind of pathology most of the time.

Theorem 8.3.3 (Maschke's Theorem)**.** *Let G be a finite group, let $\Bbbk$ be a field whose characteristic does not divide $|G|$, and let (ρ, V) be a representation of G over $\Bbbk$. Then every G-invariant subspace has a G-invariant complement. In particular, (ρ, V) is semisimple.*

Proof. If ρ is irreducible, then there is nothing to prove. Otherwise, let W be a nontrivial G-invariant subspace, and let $\pi : V \to W$ be a **projection**, i.e., a linear surjection that fixes the elements of W pointwise.

(Such a map π can be constructed as follows: choose a basis for W, extend it to a basis for V, and let π fix all the basis elements in W and kill all the ones in $V \setminus W$.)

The map π is $\Bbbk$-linear, but not necessarily G-equivariant. However, we can turn π into a G-equivariant projection by "averaging over G". (This trick will come up again and again.) Define $\tilde{\pi} : V \to W$ by

$$\tilde{\pi}(v) = \frac{1}{|G|} \sum_{g \in G} g\pi(g^{-1}v). \tag{8.2}$$

Claim 1: $\tilde{\pi}$ is a projection $V \to W$. For any $v \in V$ and $g \in G$, we have $\pi(g^{-1}v) \in W$ since π is a projection, and then $g\pi(g^{-1}) \in W$ because W is G-invariant. So $\operatorname{im}(\tilde{\pi}) \subseteq W$. Also, if $w \in W$, then $g^{-1}w \in W$, so $\pi(g^{-1}w) = g^{-1}w$ and

$$\tilde{\pi}(w) = \frac{1}{|G|} \sum_{g \in G} gg^{-1}w = w.$$

Claim 2: $\tilde{\pi}$ is G-equivariant. Indeed,

$$\begin{aligned}
\tilde{\pi}(hv) &= \frac{1}{|G|} \sum_{g \in G} g\pi(g^{-1}hv) \\
&= \frac{1}{|G|} \sum_{k \in G:\ hk = g} (hk)\pi((hk)^{-1}hv) \\
&= \frac{1}{|G|} h \sum_{k \in G} k\pi(k^{-1}v) \quad = \quad h\tilde{\pi}(v).
\end{aligned}$$

Define $W^{\perp} = \ker \tilde{\phi}$. By Prop. 8.2.7, $\ker \tilde{\pi}$ is G-invariant, and $V \cong \operatorname{im} \tilde{\pi} \oplus \ker \tilde{\pi} = W \oplus W^{\perp}$ as vector spaces, so we have found our desired G-invariant complement. □

Note that if $\operatorname{char} \Bbbk$ does divide $|G|$, then the proof breaks down because the definition of $\tilde{\pi}$ is invalid. (Removing the $1/|G|$ factor doesn't work, because $\sum_{g \in G} g\pi(g^{-1}v) = |G|v = 0$.)

Maschke's Theorem implies that, when the conditions hold, a representation ρ is determined up to isomorphism by the multiplicity of each irreducible representation in ρ (i.e., the number of isomorphic copies appearing as direct summands of ρ). Accordingly, to understand representations of G, we should first study irreps.

Example 8.3.4. Let $\Bbbk$ have characteristic 0 (for simplicity), and $G = \mathfrak{S}_n$. The defining representation of G on $\Bbbk^n$ (Example 8.1.5) is *not* simple. The one-dimensional subspace $L = \langle (1, 1, \ldots, 1) \rangle$ is fixed pointwise by every permutation $\sigma \in \mathfrak{S}_n$, and is therefore an invariant subspace, carrying the trivial representation.[2] By Maschke's theorem, L has a G-invariant complement. In fact, $L^{\perp}$ is the orthogonal complement of L under the standard inner product on $\Bbbk^n$, namely the space of all vectors whose coordinates sum to 0. This is called (a little confusingly) the **standard representation** of $\mathfrak{S}_n$, denoted ρ_{std}. That is,

$$\rho_{\text{def}} = \rho_{\text{triv}} \oplus \rho_{\text{std}}.$$

Thus $\dim \rho_{\text{std}} = n - 1$. We will soon be able to prove that ρ_{std} is irreducible (Problem 8.2). ◀

[2] For the same reason, every permutation representation of every group has a trivial summand.

8.4 Characters

The first miracle of representation theory is that we can detect the isomorphism type of a representation ρ without knowing every matrix $\rho(g)$: it turns out that we just need to know their *traces*.

Definition 8.4.1. Let (ρ, V) be a representation of G over $\Bbbk$. Its **character** is the function $\chi_\rho : G \to \Bbbk$ given by

$$\chi_\rho(g) = \operatorname{tr} \rho(g).$$

Note that characters are in general *not* group homomorphisms.

Example 8.4.2. Some simple facts and some characters we've seen before:

- A one-dimensional representation is its own character. (In fact these are exactly the characters that are homomorphisms.)
- For any representation ρ, we have $\chi_\rho(\mathrm{Id}_G) = \dim \rho$, because $\rho(\mathrm{Id}_G)$ is the $n \times n$ identity matrix.
- The defining representation ρ_{def} of $\mathfrak{S}_n$ has character

  $$\chi_{\mathrm{def}}(\sigma) = \text{number of fixed points of } \sigma.$$

  Indeed, this is true for any permutation representation of any group.
- The regular representation ρ_{reg} has character

  $$\chi_{\mathrm{reg}}(\sigma) = \begin{cases} |G| & \text{if } \sigma = \mathrm{Id}_G \\ 0 & \text{otherwise.} \end{cases}$$ ◀

Example 8.4.3. Consider the geometric representation ρ_{geo} of the dihedral group $D_n = \langle r, s \colon r^n = s^2 = e,\ srs = r^{-1} \rangle$ by rotations and reflections:

$$\rho_{\mathrm{geo}}(s) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \qquad \rho_{\mathrm{geo}}(r) = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}.$$

The character of ρ_{geo} is

$$\chi_{\mathrm{geo}}(r^j) = 2\cos(j\theta) \quad (0 \le j < n), \qquad \chi_{\mathrm{geo}}(sr^j) = 0 \quad (0 \le j < n).$$

On the other hand, if ρ' is the n-dimensional permutation representation on the vertices, then

$$\chi_{\rho'}(g) = \begin{cases} n & \text{if } g = 1, \\ 0 & \text{if } g \text{ is a nontrivial rotation,} \\ 1 & \text{if } n \text{ is odd and } g \text{ is a reflection,} \\ 0 & \text{if } n \text{ is even and } g \text{ is a reflection through two edges,} \\ 2 & \text{if } n \text{ is even and } g \text{ is a reflection through two vertices.} \end{cases}$$ ◀

Proposition 8.4.4. *Characters are **class functions**; that is, they are constant on conjugacy classes of G. Moreover, if $\rho \cong \rho'$, then $\chi_\rho = \chi_{\rho'}$.*

Proof. For the first assertion, observe that

$$\operatorname{tr}\left(\rho(hgh^{-1})\right) \;=\; \operatorname{tr}\left(\rho(h)\rho(g)\rho(h^{-1})\right) \;=\; \operatorname{tr}\left(\rho(h)\rho(g)\rho(h)^{-1}\right) \;=\; \operatorname{tr}\rho(g).$$

because $\operatorname{tr}(ABA^{-1}) = \operatorname{tr}(B)$ for all matrices A, B with A invertible.

Now, let $\phi : \rho \to \rho'$ be an isomorphism represented by an invertible matrix Φ; then $\Phi\rho(g) = \rho'(g)\Phi$ for all $g \in G$ and so $\Phi\rho(g)\Phi^{-1} = \rho'(g)$. Taking traces gives $\chi_\rho = \chi_{\rho'}$. □

Surprisingly, the converse of the second assertion is also true: a representation is determined up to isomorphism by its character!

8.5 New representations and characters from old

The basic vector space functors[3] of direct sum, duality, tensor product and Hom carry over naturally to representations, and behave well on their characters. Throughout this section, let G be a finite group and let (ρ, V) and (ρ', W) be finite-dimensional representations of G over $\mathbb{C}$, with $V \cap W = 0$.

1. **Direct sum**. To construct a basis for $V \oplus W$, we can take the union of a basis for V and a basis for W. Equivalently, we can write the vectors in $V \oplus W$ as column block vectors:

$$V \oplus W = \left\{ \begin{bmatrix} v \\ w \end{bmatrix} : v \in V,\ w \in W \right\}.$$

Accordingly, we can define the direct sum $(\rho \oplus \rho', V \oplus W)$ by

$$(\rho \oplus \rho')(h) = \left[\begin{array}{c|c} \rho(h) & 0 \\ \hline 0 & \rho'(h) \end{array}\right].$$

From this it is clear that $\chi_{\rho\oplus\rho'}(h) = \chi_\rho(h) + \chi_{\rho'}(h)$.

2. **Duality**. The *dual space* V^* of V consists of all $\Bbbk$-linear transformations $\phi : V \to \Bbbk$. A G-representation (ρ, V) gives rise to a representation (ρ^*, V^*) given by

$$(h\phi)(v) = \phi(h^{-1}v)$$

for $h \in G$, $\phi \in V^*$, $v \in V$. Alternatively, $h\phi = \phi \circ h^{-1}$. This is a little counterintuitive (one might expect $\phi(hv)$ on the right-hand side) but it needs to be defined this way in order for ρ^* to be a homomorphism, i.e., for $\rho^*(gh)$ to equal $\rho^*(g)\rho^*(h)$ rather than $\rho^*(h)\rho^*(g)$. (Try it.) The representation ρ^* is called the **dual representation** (or **contragredient**) of ρ.

Proposition 8.5.1. *For every $h \in G$,*

$$\chi_{\rho^*}(h) = \overline{\chi_\rho(h)}. \tag{8.3}$$

where the bar denotes complex conjugate.

Proof. Let J be the Jordan canonical form of $\rho(h)$ (which exists since we are working over $\mathbb{C}$), so that $\chi_{\rho(h)} = \operatorname{tr} J$. The diagonal entries J_{ii} are its eigenvalues, which must be roots of unity since h has finite order, so their inverses are their complex conjugates. Meanwhile, J^{-1} is an upper-triangular matrix with $(J^{-1})_{ii} = (J_{ii})^{-1} = \overline{J_{ii}}$, and $\operatorname{tr} J^{-1} = \chi_{\rho^*}(h) = \overline{\chi_\rho(h)}$. □

3. **Tensor product.** Fix bases $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ and $\{\mathbf{f}_1, \ldots, \mathbf{f}_m\}$ for V and W respectively. As a vector space, define[4]

$$V \otimes W = \Bbbk \left\langle \mathbf{e}_i \otimes \mathbf{f}_j \mid 1 \le i \le n,\ 1 \le j \le m \right\rangle,$$

[3]Technically, a functor is an operation that not only takes objects to objects, but also maps to maps. E.g., applying duality to vector spaces not only takes V and W to V^* and W^*, but also takes every linear transformation $\phi : V \to W$ to a linear transformation $\phi^* : W^* \to V^*$. We will not be too concerned with functors as such in this section, but will mostly focus on their effect on characters/

[4]The "official" definition of the tensor product is much more functorial and can be made basis-free, but this concrete definition is more convenient for our present purposes.

equipped with a multilinear action of $\Bbbk$ (that is, $c(x \otimes y) = cx \otimes y = x \otimes cy$ for $c \in \Bbbk$). In particular, $\dim(V \otimes W) = (\dim V)(\dim W)$. We can accordingly define a representation $(\rho \otimes \rho', V \otimes W)$ by

$$(\rho \otimes \rho')(h)(v \otimes w) = \rho(h)(v) \otimes \rho'(h)(w)$$

or more concisely

$$h \cdot (v \otimes w) = hv \otimes hw$$

extended bilinearly to all of $V \otimes W$.

In terms of matrices, $(\rho \otimes \rho')(h)$ is represented by the $nm \times nm$ matrix $\rho(h) \otimes \rho'(h)$, defined by

$$(\rho(h) \otimes \rho'(h))_{(i,j),(k,\ell)} \;=\; (\rho(h))_{i,k}\,(\rho'(h))_{j,\ell}.$$

That is, the left-hand side is the entry in the row corresponding to $\mathbf{e}_i \otimes \mathbf{f}_j$ and column corresponding to $\mathbf{e}_k \otimes \mathbf{f}_\ell$. In particular,

$$\chi_{\rho\otimes\rho'}(h) = \sum_{(i,j)\in[n]\times[m]} (\rho(h))_{i,i}\,(\rho'(h))_{j,j} = \left(\sum_{i=1}^{n}(\rho(h))_{i,i}\right)\left(\sum_{j=1}^{m}(\rho'(h))_{j,j}\right) = \chi_\rho(h)\chi_{\rho'}(h). \tag{8.4}$$

4. **Hom.** There are two kinds of Hom:

$$\begin{aligned}\mathrm{Hom}_{\mathbb{C}}(V,W) &= \{\mathbb{C}\text{-linear transformations } \phi : V \to W\},\\ \mathrm{Hom}_{G}(V,W) &= \{G\text{-equivariant maps } \phi : V \to W\}.\end{aligned}$$

First, the vector space $\mathrm{Hom}_{\mathbb{C}}(V,W)$ admits a representation of G, in which $h \in G$ acts on a linear transformation $\phi : V \to W$ by sending it to the map $h \cdot \phi$ defined by

$$(h \cdot \phi)(v) = h(\phi(h^{-1}v)) = \rho'(h)\Big(\phi(\rho(h^{-1})(v))\Big). \tag{8.5}$$

for $h \in G$, $\phi \in \mathrm{Hom}_{\mathbb{C}}(V,W)$, $v \in V$. It is straightforward to verify that this is a genuine group action, i.e., that $(hh') \cdot \phi = h \cdot (h' \cdot \phi)$.

Moreover, $\mathrm{Hom}_{\mathbb{C}}(V,W) \cong V^* \otimes W$ as vector spaces and G-modules. To see this, suppose that $\dim V = n$ and $\dim W = m$; then the elements of V^* and W can be regarded as $1 \times n$ and $m \times 1$ matrices respectively (the former acting on V, which consists of $n \times 1$ matrices, by matrix multiplication). Then the previous description of tensor product implies that $V^* \otimes W$ consists of $m \times n$ matrices, which correspond to elements of $\mathrm{Hom}_{\mathbb{C}}(V,W)$. This isomorphism is G-equivariant by (8.5), so

$$\chi_{\mathrm{Hom}(\rho,\rho')}(h) = \chi_{\rho^*\otimes\rho'}(h) = \overline{\chi_\rho(h)}\;\chi_{\rho'}(h). \tag{8.6}$$

To summarize so far, we have seen that

$$\chi_{\rho\oplus\rho'} = \chi_\rho + \chi_{\rho'}, \qquad \chi_{\rho^*} = \overline{\chi_\rho}, \qquad \chi_{\rho\otimes\rho'} = \chi_\rho \cdot \chi_{\rho'}, \qquad \chi_{\mathrm{Hom}_{\mathbb{C}}(\rho,\rho')} = \overline{\chi_\rho} \cdot \chi_{\rho'},$$

so we may consider these functors as operations on characters: e.g., $(\chi \otimes \psi)(g) = \chi(g)\psi(g)$, etc.

What about $\mathrm{Hom}_G(V,W)$? Evidently $\mathrm{Hom}_G(V,W) \subseteq \mathrm{Hom}_{\mathbb{C}}(V,W)$, but equality need not hold. For example, if V and W are the trivial and sign representations of $\mathfrak{S}_n$ (for $n \geq 2$), then $\mathrm{Hom}_{\mathbb{C}}(V,W) \cong \mathbb{C}$ but $\mathrm{Hom}_G(V,W) = 0$. (See Example 8.2.9.)

The two Homs are related as follows. In general, when a group G acts on a vector space V, the **subspace of G-invariants** is defined as

$$V^G = \{v \in V \mid hv = v \quad \forall h \in G\}.$$

This is the largest subspace of V that carries the trivial action.

Observe that a linear map $\phi : V \to W$ is G-equivariant if and only if $h\phi = \phi$ for all $h \in G$, where G acts on $\mathrm{Hom}_{\mathbb{C}}(V, W)$ as in (8.5). (The proof of this fact is left to the reader.)

$$\mathrm{Hom}_G(V, W) = \mathrm{Hom}_{\mathbb{C}}(V, W)^G. \tag{8.7}$$

Moreover, G acts by the identity on $\mathrm{Hom}_G(\rho, \rho')$, so its character is a constant function whose value is $\dim_{\mathbb{C}} \mathrm{Hom}_G(\rho, \rho')$. To put it another way, the action of G is a sum of copies of the trivial representation. We want to understand how many.

8.6 The fundamental theorem of character theory

From now on, we assume that $\Bbbk = \mathbb{C}$ (though everything would be true over an algebraically closed field of characteristic 0), unless otherwise specified.

Recall that a **class function** is a function $\chi : G \to \mathbb{C}$ that is constant on conjugacy classes of G. The set $C\ell(G)$ of class functions forms a vector space. Define an inner product on $C\ell(G)$ by

$$\langle \chi,\, \psi \rangle_G = \frac{1}{|G|} \sum_{h \in G} \overline{\chi(h)}\, \psi(h) = \frac{1}{|G|} \sum_{C} |C|\, \overline{\chi(C)}\, \psi(C) \tag{8.8}$$

where C runs over all conjugacy classes. Observe that $\langle \cdot,\, \cdot \rangle_G$ is a sesquilinear form (i.e., $\mathbb{C}$-linear in the second term and conjugate linear in the first). It is also non-degenerate, because the indicator functions of conjugacy classes form an orthogonal basis for $C\ell(G)$. Analysts might want to regard the inner product as a convolution (with summation over G as a discrete analogue of integration).

Proposition 8.6.1. *Let (ρ, V) be a representation of G. Then*

$$\dim_{\mathbb{C}} V^G = \frac{1}{|G|} \sum_{h \in G} \chi_\rho(h) = \langle \chi_{\mathrm{triv}},\, \chi_\rho \rangle_G .$$

Proof. The second equality follows from the definition of the inner product. For the first equality, define a linear map $\pi : V \to V$ by

$$\pi = \frac{1}{|G|} \sum_{h \in G} \rho(h).$$

Note that $\pi(v) \in V^G$ for all $v \in V$, because

$$g\pi(v) = \frac{1}{|G|} \sum_{h \in G} ghv = \frac{1}{|G|} \sum_{gh \in G} ghv = \pi(v)$$

and moreover if $v \in V^G$ then

$$\pi(v) = \frac{1}{|G|} \sum_{h \in G} hv = \frac{1}{|G|} \sum_{h \in G} v = v.$$

That is, π is a projection from $V \to V^G$. Choose a basis for V consisting of a basis for V^G and extend it to a basis for V. With respect to that basis, π can be represented by the block matrix

$$\left[\begin{array}{c|c} I & * \\ \hline 0 & 0 \end{array}\right]$$

(where $*$ is some matrix irrelevant for our present purpose), so that

$$\dim_{\mathbb{C}} V^G = \operatorname{tr}(\pi) = \frac{1}{|G|} \sum_{h \in G} \chi_\rho(h). \qquad \square$$

By the way, we know by Maschke's Theorem that V is semisimple, so we can decompose it as a direct sum of irreps. Then V^G is precisely the direct sum of the irreducible summands on which G acts trivially.

Example 8.6.2. Let G act on a set X, and let ρ be the corresponding permutation representation on the space $\mathbb{C}X$. For each orbit $\mathcal{O} \subseteq X$, the vector $v_{\mathcal{O}} = \sum_{x \in \mathcal{O}} x$ is fixed by G. On the other hand, any vector $\sum_{x \in X} a_x x$ fixed by G must have a_x constant on each orbits. Therefore the vectors $v_{\mathcal{O}}$ are a basis for V^G, and $\dim V^G$ is the number of orbits. So Proposition 8.6.1 becomes

$$\text{\# orbits } = \frac{1}{|G|} \sum_{h \in G} \text{\# fixed points of } h$$

which is Burnside's Lemma from abstract algebra. ◀

Proposition 8.6.3. *For any two representations ρ, ρ' of G, we have $\left\langle \chi_\rho,\ \chi_{\rho'} \right\rangle_G = \dim_{\mathbb{C}} \operatorname{Hom}_G(\rho, \rho')$.*

Proof. Calculate $\left\langle \chi_\rho,\ \chi_{\rho'} \right\rangle_G$ as

$$\begin{aligned} \frac{1}{|G|} \sum_{h \in G} \overline{\chi_\rho(h)} \chi_{\rho'}(h) &= \frac{1}{|G|} \sum_{h \in G} \chi_{\operatorname{Hom}(\rho, \rho')}(h) && \text{by (8.6)} \\ &= \dim_{\mathbb{C}} \operatorname{Hom}(\rho, \rho')^G && \text{by Proposition 8.6.1, with } V = \operatorname{Hom}(\rho, \rho') \\ &= \dim_{\mathbb{C}} \operatorname{Hom}_G(\rho, \rho') && \text{by (8.7).} \end{aligned}$$

□

One intriguing observation is that this expression is symmetric in ρ and ρ', since $\langle \alpha,\ \beta \rangle_G = \overline{\langle \beta,\ \alpha \rangle_G}$ in general, but $\dim_{\mathbb{C}} \operatorname{Hom}_G(\rho, \rho')$ is real. (It is not algebraically obvious that $\operatorname{Hom}_G(\rho, \rho')$ and $\operatorname{Hom}_G(\rho', \rho)$ should have equal dimension.)

Proposition 8.6.4 (Schur's Lemma)**.** *Let G be a group, and let (ρ, V) and (ρ', V') be finite-dimensional irreps of G over a field $\Bbbk$ (not necessarily of characteristic 0).*

1. *Every G-equivariant map $\phi : V \to V'$ is either zero or an isomorphism.*
2. *If in addition $\Bbbk$ is algebraically closed, then*

$$\operatorname{Hom}_G(V, V') \cong \begin{cases} \Bbbk & \text{if } \rho \cong \rho' \\ 0 & \text{otherwise.} \end{cases}$$

That is, the only G-equivariant maps from an G-irrep to itself are multiplication by a scalar.

Proof. For (1), recall from Proposition 8.2.7 that $\ker\phi$ and $\operatorname{im}\phi$ are G-invariant subspaces. But since ρ, ρ' are simple, there are not many possibilities. Either $\ker\phi = 0$ and $\operatorname{im}\phi = W$, when ϕ is an isomorphism. Otherwise, $\ker\phi = V$ or $\operatorname{im}\phi = 0$, either of which implies that $\phi = 0$.

For (2), the "otherwise" case follows from (1), so suppose $V \cong V'$. Actually, for simplicity, assume $V = V'$. Since $\Bbbk$ is algebraically closed, every G-equivariant map $\phi : V \to V$ has an eigenvalue λ. Then $\phi - \lambda I$ is G-equivariant and singular, hence zero by (1). So $\phi = \lambda I$ is multiplication by λ. □

We can now prove the following omnibus theorem, which essentially reduces the study of representations of finite groups to the study of characters.

Theorem 8.6.5 (Fundamental Theorem of Character Theory for Finite Groups). *Let (ρ, V) and (ρ', V') be finite-dimensional representations of G over $\mathbb{C}$.*

1. *If ρ and ρ' are irreducible, then*
$$\langle \chi_\rho, \chi_{\rho'} \rangle_G = \begin{cases} 1 & \text{if } \rho \cong \rho', \\ 0 & \text{otherwise.} \end{cases} \tag{8.9}$$
In particular, there are only finitely many isomorphism classes of irreps of G.
2. *If $\rho_1, \dots, \rho_n$ are distinct irreducible representations and*
$$\rho = \bigoplus_{i=1}^{n} (\underbrace{\rho_i \oplus \cdots \oplus \rho_i}_{m_i}) = \bigoplus_{i=1}^{n} \rho_i^{\oplus m_i}$$
then
$$\langle \chi_\rho, \chi_{\rho_i} \rangle_G = m_i \qquad \text{and} \qquad \langle \chi_\rho, \chi_\rho \rangle_G = \sum_{i=1}^{n} m_i^2.$$
In particular, $\langle \chi_\rho, \chi_\rho \rangle_G = 1$ if and only if ρ is irreducible.
3. *Characters are complete invariants for their representations. That is, if $\chi_\rho = \chi_{\rho'}$ then $\rho \cong \rho'$.*
4. *Let $\rho_1, \dots, \rho_n$ be a complete list of irreps of G up to isomorphism. Then*
$$\rho_{\text{reg}} \cong \bigoplus_{i=1}^{n} \rho_i^{\oplus \dim \rho_i}$$
and consequently
$$\sum_{i=1}^{n} (\dim \rho_i)^2 = |G|. \tag{8.10}$$
5. *The irreducible characters (i.e., characters of irreps) form an orthonormal basis for $C\ell(G)$. In particular, the number of irreducible characters equals the number of conjugacy classes of G.*

Proof. For assertion (1), the equation (8.9) follows from part (2) of Schur's Lemma together with Proposition 8.6.3. It follows that the characters of isomorphism classes of irreps are an orthonormal basis for some subspace of the finite-dimensional space $C\ell(G)$, so there can be only finitely many of them. (This result continues to amaze me every time I think about it.)

Assertion (2) follows because the inner product is additive on direct sums. That is, every class function ψ satisfies
$$\langle \chi_{\rho\oplus\rho'}, \psi \rangle_G = \langle \chi_\rho + \chi_{\rho'}, \psi \rangle_G = \langle \chi_\rho, \psi \rangle_G + \langle \chi_{\rho'}, \psi \rangle_G .$$

For (3), Maschke's Theorem says that every complex representation ρ can be written as a direct sum of irreducibles. Their multiplicities determine ρ up to isomorphism, and can be recovered from χ_ρ by (2). (Again, amazing.)

For (4), recall that $\chi_{\text{reg}}(\text{Id}_G) = |G|$ and $\chi_{\text{reg}}(g) = 0$ for $g \neq \text{Id}_G$. Therefore

$$\langle \chi_{\text{reg}},\, \rho_i \rangle_G = \frac{1}{|G|} \sum_{g \in G} \overline{\chi_{\text{reg}}(g)} \rho_i(g) = \frac{1}{|G|} |G| \rho_i(\text{Id}_G) = \dim \rho_i$$

so ρ_i appears in ρ_{reg} with multiplicity equal to its dimension.

For (5) Schur's Lemma together with assertion (3) imply that the irreducible characters are orthonormal in $C\ell(G)$, hence linearly independent. The trickier part is to show that they in fact span $C\ell(G)$. Let Y be the subspace of $C\ell(G)$ spanned by the irreducible characters, and let

$$Z = Y^\perp = \left\{ \phi \in C\ell(G) :\ \langle \phi,\, \chi_\rho \rangle_G = 0 \text{ for every irreducible character } \rho \right\}.$$

We will show that $Z = 0$.

Let $\phi \in Z$. For any representation (ρ, V), define a map $T_\rho = T_{\rho,\phi} : V \to V$ by

$$T_\rho = \frac{1}{|G|} \sum_{g \in G} \overline{\phi(g)}\, \rho(g)$$

or equivalently

$$T_\rho(v) = \frac{1}{|G|} \sum_{g \in G} \overline{\phi(g)} gv$$

(to parse this, note that $\phi(g)$ is a number). Our plan is to show that T_ρ is the zero map (in disguise), then deduce that $\phi = 0$.

First, we show that T_ρ is G-equivariant. Let $h \in G$; then

$$\begin{aligned}
T_\rho(hv) &= \frac{1}{|G|} \sum_{g \in G} \overline{\phi(g)}\, ghv \\
&= \frac{1}{|G|} h \sum_{g \in G} \overline{\phi(g)}\, h^{-1} ghv \\
&= h \frac{1}{|G|} \sum_{k \in G} \overline{\phi(hkh^{-1})}\, kv && \text{(setting } k = h^{-1}gh,\ hkh^{-1} = g\text{)} \\
&= h \frac{1}{|G|} \sum_{k \in G} \overline{\phi(k)}\, kv && \text{(because } \phi \text{ is a class function)} \\
&= hT_\rho(v).
\end{aligned}$$

Second, we show that $T_\rho = 0$. We start with the case that ρ is irreducible. Since T_ρ is G-equivariant, Schur's Lemma implies that it is multiplication by a scalar. On the other hand,

$$\text{tr}(T_\rho) \;=\; \frac{1}{|G|} \sum_{g \in G} \overline{\phi(g)} \chi_\rho(g) \;=\; \langle \phi,\, \chi_\rho \rangle_G \;=\; 0$$

because $\phi \in Z = Y^\perp$. Since T_ρ has trace zero and is multiplication by a scalar, it is the zero map. (Again, here we need $\mathbb{k}$ to have characteristic 0.)

Now we consider an arbitrary representation ρ. By Maschke's Theorem, every complex representation ρ is semisimple, and the definition of T_ρ implies that it is additive on direct sums (that is, $T_{\rho\oplus\rho'} = T_\rho + T_{\rho'}$), proving that $T_\rho = 0$ for all ρ.

In particular, if $\rho = \rho_{\text{reg}}$ is the regular representation, then

$$0 \;=\; T_{\rho_{\text{reg}}}(\text{Id}_G) \;=\; \frac{1}{|G|}\sum_{g\in G}\overline{\phi(g)}\rho_{\text{reg}}(g).$$

This is an equation in the vector space of $|G| \times |G|$ matrices. Observe that the permutation matrices $\rho_{\text{reg}}(g)$ have disjoint supports (because the only group element that maps h to k is kh^{-1}), hence are linearly independent. Therefore $\phi(g) = 0$ for all g, so ϕ is the zero map.

We have now shown that Y has trivial orthogonal complement as a subspace of $C\ell(G)$, so $Y = C\ell(G)$, completing the proof. □

Corollary 8.6.6. *Let $\rho_1, \ldots, \rho_n$ be the irreps of G, and let $\alpha = \bigoplus_{i=1}^n \rho_i^{\oplus a_i}$ and $\beta = \bigoplus_{i=1}^n \rho_i^{\oplus b_i}$ be two representations. Then $\dim \text{Hom}_G(\alpha, \beta) = \sum_{i=1}^n a_i b_i$, and in particular $\text{Hom}_G(\alpha, \beta)$ is a direct sum of constant maps between irreducible summands of α and β.*

The irreducible characters of G can thus be written as a square matrix X with columns corresponding to conjugacy classes. This is called the **character table** of G. (It is helpful to include the size of the conjugacy class in the table, for ease of computing scalar products.) By orthonormality of characters, the matrix X is close to being unitary: $XDX^* = I$, where the star denotes conjugate transpose and D is the diagonal matrix with entries $|C|/|G|$. It follows that $X^*X = D^{-1}$, which is equivalent to the following statement:

Proposition 8.6.7. *Let $\chi_1, \ldots, \chi_n$ be the irreducible characters of G and let C, C' be any two conjugacy classes. Then*

$$\sum_{i=1}^n \overline{\chi_i(C)}\chi_i(C') = \frac{|G|}{|C|}\,\delta_{C,C'}.$$

This is a convenient tool because it says that different columns of the character table are orthogonal under the usual scalar product (without having to correct for the size of conjugacy classes).

At this point, we can also take care of some unfinished business: we claimed in Example 8.3.4 that the standard representation ρ_{std} of $\mathfrak{S}_n$ (defined by $\rho_{\text{def}} = \rho_{\text{std}} \oplus \rho_{\text{triv}}$) is irreducible for all $n \geq 2$. This claim can now be proved using characters. We leave the proof as an exercise (Problem 8.2), but will use the fact freely in what follows. Note that the standard character maps each permutation to its number of fixed points minus one.

8.7 Computing character tables

Theorem 8.6.5 provides the basic tools to calculate character tables. In general, the character table of a finite group G with k conjugacy classes is a $k \times k$ table in which rows correspond to irreducible characters $\chi_1 \ldots, \chi_k$ and columns to conjugacy classes. Part (1) of the Theorem says that the rows form an orthonormal basis under the inner product on class functions, so computing a character table resembles a Gram-Schmidt process. The hard part is coming up with enough representations whose characters span $C\ell(G)$. Here are some ways of generating them:

- Every group carries the trivial and regular characters, which are easy to write down. The regular character contains at least one copy of every irreducible character.
- The symmetric group also has the sign and standard characters, both of which are irreducible.
- Many groups come with natural permutation actions whose characters can be added to the mix.
- The operations of duality and tensor product can be used to come up with new characters. Duality preserves irreducibility, but tensor product typically does not.

Remark 8.7.1. While we have defined $C\ell(G)$ as a vector space, we are really concerned with the $\mathbb{Z}$-module $C\ell_{\mathbb{Z}}(G)$ generated by the irreducible characters. Its elements are called **virtual characters** (so called because a virtual character is a sum of irreducible characters with some negative coefficients, then it does not come from a genuine representation). Being a basis for $C\ell_{\mathbb{Z}}(G)$ as a $\mathbb{Z}$-module (an **integral basis**) is a much stronger condition than being a basis for $C\ell(G)$ as a vector space.

When we say "character", we always mean "character of an honest-to-goodness representation," or equivalently (by Maschke's theorem) a linear combination of irreducible characters *with nonnegative integer coefficients*. When computing character tables, the goal is to find the characters of irreps, not any old orthonormal basis of $C\ell(G)$. Thus, while linear-algebraic tools like the Gram-Schmidt process can be useful, there is more to it than that!

In the following examples, we will notate a character χ by a bracketed list of its values on conjugacy classes, in the same order that they are listed in the table. Numerical subscripts will always be reserved for irreducible characters.

Example 8.7.2. The group $G = \mathfrak{S}_3$ has three conjugacy classes, indexed by cycle-types:

$$C_{111} = \{\mathrm{Id}_G\}, \qquad C_{21} = \{(12),\ (13),\ (23)\}, \qquad C_3 = \{(123),\ (132)\}.$$

We know three irreducible characters of $\mathfrak{S}_3$: the trivial and sign characters, which are both 1-dimensional, and the standard character, which is 2-dimensional. So this must be the complete character table:

	1 C_{111}	3 C_{21}	2 C_3
χ_{triv}	1	1	1
χ_{sign}	1	-1	1
χ_{std}	2	0	-1

Note that the regular character $\chi_{\mathrm{reg}} = [6, 0, 0]$ equals $\chi_{\mathrm{triv}} + \chi_{\mathrm{sign}} + 2\chi_{\mathrm{std}}$, as predicted by Theorem 8.6.5(e). It is also worth noting that $\chi_{\mathrm{sign}} \otimes \chi_{\mathrm{std}} = \chi_{\mathrm{std}}$. ◀

Example 8.7.3. We calculate all the irreducible characters of $\mathfrak{S}_4$. There are five conjugacy classes, corresponding to the cycle-types 1111, 211, 22, 31, and 4. The squares of their dimensions must add up to $|\mathfrak{S}_4| = 24$; the only list of five positive integers with that property is 1, 1, 2, 3, 3.

We do know three irreducible characters, namely χ_{triv}, χ_{sign}, and χ_{std}, of dimensions 1,1,3 respectively. So we need to come up with two more. The regular character is always useful, and we can also try out the character $\chi_{\mathrm{std}} \otimes \chi_{\mathrm{sign}}$:

	1 C_{1111}	6 C_{211}	3 C_{22}	8 C_{31}	6 C_4
$\chi_1 = \chi_{\text{triv}}$	1	1	1	1	1
$\chi_2 = \chi_{\text{sign}}$	1	-1	1	1	-1
$\chi_3 = \chi_{\text{std}}$	3	1	-1	0	-1
$\chi_4 = \chi_{\text{std}} \otimes \chi_{\text{sign}}$	3	-1	-1	0	1
χ_{reg}	24	0	0	0	0

It is easy to check that $\langle \chi_4,\ \chi_4 \rangle_G = 1$, so χ_4 is irreducible. A general fact illustrated here is that tensoring with the sign character preserves irreducibility — for an even more general statement, see Problem 8.1.

The other irreducible character χ_5 has dimension 2. We can calculate it from the regular character and the other four irreducibles, because

$$\chi_{\text{reg}} \ = \ (\chi_1 + \chi_2) + 3(\chi_3 + \chi_4) + 2\chi_5$$

and so

$$\chi_5 = \frac{\chi_{\text{reg}} - \chi_1 - \chi_2 - 3\chi_3 - 3\chi_4}{2}.$$

Therefore the complete character table of $\mathfrak{S}_4$ is as follows.

	1 C_{1111}	6 C_{211}	3 C_{22}	8 C_{31}	6 C_4
χ_1	1	1	1	1	1
χ_2	1	-1	1	1	-1
χ_3	3	1	-1	0	-1
χ_4	3	-1	-1	0	1
χ_5	2	0	2	-1	0

◀

Example 8.7.4. For a non-symmetric example, let G be the dihedral group $D_4 = \langle r, s \colon r^4 = s^2 = e,\ srs = r^{-1} \rangle$ of order 8. The elements e, r, r^2, r^3 are *rotations* (of orders 1,4,2,4 respectively) and the elements s, sr, sr^2, sr^3 are *reflections* (all of order 2).

First we determine the conjugacy classes. The identity is of course its own conjugacy class, and r^2 is central (it commutes with everything) so it is also its own conjugacy class. The two order-4 rotations are conjugate by the defining relation of D_4. That leaves the reflections. On the one hand, $rsr^{-1} = rsr^3 = (rsr)r^2 = sr^2$, which says that s is conjugate to sr^2, and a similar calculation shows that sr and sr^3 are conjugate. On the other hand, conjugation by s preserves the parity of the number of r's in a word, so these are the only conjugacies. In summary, there are five conjugacy classes:

$$C_1 = \{e\}, \quad C_2 = \{r^2\}, \quad C_3 = \{r, r^3\}, \quad C_4 = \{s, sr^2\}, \quad C_5 = \{sr, sr^3\}.$$

Therefore there are 5 irreps, and the squares of their dimensions must sum to 8; the only possibility is 1,1,1,1,2.

Now let's collect some characters. There's always the trivial character and the regular character. By the parity argument, sending $r \mapsto 1$ and $s \mapsto -1$ is a homomorphism $G \to \mathbb{C}^\times$; we'll call this the **orientation character**. From Example 8.1.8, we also have the characters of the geometric representation ρ_{geo} and the

permutation representations ρ_{vert} and ρ_{diag} on vertices and diagonals respectively. Here's what we have so far:

		rotations			reflections		
		1	1	2	2	2	
		C_1	C_2	C_3	C_4	C_5	Scalar product with self
Trivial	χ_{triv}	1	1	1	1	1	1
Orientation	χ_{ori}	1	1	1	-1	-1	1
Geometric	χ_{geo}	2	-2	0	0	0	1
Diagonal	χ_{diag}	2	2	0	2	0	2
Vertex	χ_{vert}	4	0	0	2	0	3
Regular	χ_{reg}	8	0	0	0	0	

Evidently χ_{triv} and χ_{ori} are irreducible, and $\langle \chi_{\text{geo}}, \chi_{\text{geo}} \rangle = 1$ so it must be the irreducible character of dimension 2. Meanwhile, χ_{diag} and χ_{vert} come from permutation representations, so they each must include at least one copy of the trivial character; in fact $\left\langle \chi_{\text{diag}}, \chi_{\text{triv}} \right\rangle_G = \langle \chi_{\text{vert}}, \chi_{\text{triv}} \rangle_G = 1$. This observation says that we should look at the characters

$$\chi_{\text{diag}} - \chi_{\text{triv}} = [1, 1, -1, 1, -1], \qquad \chi_{\text{vert}} - \chi_{\text{triv}} = [3, -1, -1, 1, -1].$$

The first of these is clearly irreducible; call it χ_1. The second one equals $\chi_1 + \chi_{\text{geo}}$, so this character does not give us anything new.

To find the fifth and final irreducible character, one strategy is to use the regular character:

$$\chi_{\text{reg}} = \sum_{\text{irr. chars. } \chi} (\dim \chi)\chi = \chi_{\text{triv}} + \chi_{\text{ori}} + 2\chi_{\text{geo}} + \chi_1 + \chi_2.$$

where χ_2 is the irreducible character we are looking for. Solving this equation gives $\chi_2 = [1, 1, -1, -1, 1]$. Alternatively, we could have observed that $\chi_2 = \chi_1 \otimes \chi_{\text{ori}}$. The final character table of D_4 is as follows:

	1	1	2	2	2
	C_1	C_2	C_3	C_4	C_5
χ_{triv}	1	1	1	1	1
χ_{ori}	1	1	1	-1	-1
χ_1	1	1	-1	1	-1
χ_2	1	1	-1	-1	1
χ_{geo}	2	-2	0	0	0

In particular, each of the one-dimensional characters is its own dual, and in fact they form a Klein four-group under tensor product, with duality as the inverse. In the next section, we take this observation and run with it. ◀

8.8 One-dimensional characters

A one-dimensional character of G is identical with the representation it comes from: a group homomorphism $G \to \mathbb{C}^\times$. Since tensor product is multiplicative on dimension, it follows that the tensor product of two one-dimensional characters is also one-dimensional. In fact $(\chi \otimes \chi')(g) = \chi(g)\chi'(g)$ (this is immediate from the definition of tensor product) and $\chi \otimes \chi^* = \chi_{\text{triv}}$. So the set $\text{Ch}(G)$ of one-dimensional characters, i.e., $\text{Ch}(G) = \text{Hom}(G, \mathbb{C}^\times)$, is an abelian group under tensor product (equivalently, pointwise multiplication), with identity χ_{triv}.

Definition 8.8.1. The **commutator** of two elements $a, b \in G$ is the element $[a, b] = aba^{-1}b^{-1}$. The (normal) subgroup of G generated by all commutators is called the **commutator subgroup**, denoted $[G, G]$. The quotient $G^{ab} = G/[G, G]$ is the **abelianization** of G.

The abelianization can be regarded as the group obtained by forcing all elements G to commute, in addition to whatever relations already exist in G; in other words, it is the largest abelian quotient of G. It is routine to check that $[G, G]$ is indeed normal in G, and also that $\chi([a, b]) = 1$ for all $\chi \in \mathrm{Ch}(G)$ and $a, b \in G$. (In fact, this condition characterizes the elements of the commutator subgroup, as will be shown soon.) Therefore $\mathrm{Ch}(G) \cong \mathrm{Ch}(G^{ab})$ and it suffices to understand the character groups of abelian groups.

Accordingly, let G be an abelian group of finite order n. The conjugacy classes of G are all singleton sets (since $ghg^{-1} = h$ for all $g, h \in G$), so there are n distinct irreducible representations of G. By (8.10) of Theorem 8.6.5, so in fact every irreducible character is 1-dimensional (and every representation of G is a direct sum of 1-dimensional representations). We have now reduced the problem to describing the group homomorphisms $G \to \mathbb{C}^\times$.

The simplest case is that $G = \mathbb{Z}/n\mathbb{Z}$ is cyclic. Write G multiplicatively, and let g be a generator. Then each $\chi \in \mathrm{Ch}(G)$ is determined by its value on g, which must be some nth root of unity. There are n possibilities for $\chi(g)$, so all the irreducible characters of G arise in this way, and in fact form a group isomorphic to $\mathbb{Z}/n\mathbb{Z}$, generated by any character that maps g to a *primitive* nth root of unity. So $\mathrm{Hom}(G, \mathbb{C}^\times) \cong G$ (although this isomorphism is not canonical).

Now we consider the general case. Every abelian group G can be written as

$$G \cong \prod_{i=1}^{r} \mathbb{Z}/n_i\mathbb{Z}.$$

Let g_i be a generator of the ith factor, and let ζ_i be a primitive (n_i)th root of unity. Then each character χ is determined by the numbers $j_1, \ldots, j_r$, where $j_i \in \mathbb{Z}/n_i\mathbb{Z}$ and $\chi(g_i) = \zeta_i^{j_i}$ for all i. Thus $\mathrm{Hom}(G, \mathbb{C}^\times) \cong G$, an isomorphism known as **Pontryagin duality**. More generally, for any finite group G (not necessarily abelian), there is an isomorphism

$$\mathrm{Hom}(G, \mathbb{C}^\times) \cong G^{ab}. \tag{8.11}$$

This is quite useful when computing the character table of a group: if you can figure out the commutator subgroup and/or the abelianization, then you can immediately write down the one-dimensional characters. Sometimes the size of the abelianization can be determined from the size of the group and the number of conjugacy classes. (The commutator subgroup is normal, so it itself is a union of conjugacy classes.)

The description of characters of abelian groups implies that if G is abelian and $g \neq \mathrm{Id}_G$, then $\chi(g) \neq 1$ for at least one character χ. Therefore, for *every* group G, we have

$$[G, G] = \{g \in G \colon \chi(g) = 1 \text{ for all 1-dimensional characters } \chi\}$$

since $[G, G]$ is the identity element of G^{ab}. This

Example 8.8.2. Suppose that G is a group of order 24 with 8 conjugacy classes.[5] There is only one possibility for the dimensions of the irreps (i.e., only one solution to the equation $\sum_{i=1}^{8} d_i^2 = 24$ in positive integers), namely 1,1,1,1,1,1,3,3. In particular the abelianization must have size 6 and the commutator subgroup must have size $24/6 = 4$. There is only one abelian group of order 6, so we know the 1-dimensional characters of G^{ab}, and it should not be hard to pull back to the 1-dimensional characters of G^{ab}, since the quotient map $G \to G^{ab}$ is constant on conjugacy classes.

[5] According to `Group Properties Wiki` (11/9/22), there happens to be exactly one such group, namely $\mathfrak{A}_4 \times \mathbb{Z}_2$.

If instead the group were known to have 6 conjugacy classes, then the equation has two solutions, namely 1,1,1,1,2,4 and 2,2,2,2,2,2, but the latter is impossible since every group has at least one 1-dimensional irrep, namely the trivial representation. ◀

Example 8.8.3. Consider the case $G = \mathfrak{S}_n$. Certainly $[\mathfrak{S}_n, \mathfrak{S}_n] \subseteq \mathfrak{A}_n$, and in fact equality holds. This is trivial for $n \leq 2$. If $n \leq 3$, then the equation $(a\ b)(b\ c)(a\ b)(b\ c) = (a\ b\ c)$ in $\mathfrak{S}_n$ (multiplying left to right) shows that $[\mathfrak{S}_n, \mathfrak{S}_n]$ contains every 3-cycle, and it is not hard to show that the 3-cycles generate the full alternating group. Therefore (8.11) gives

$$\operatorname{Hom}(\mathfrak{S}_n, \mathbb{C}^\times) \ \cong\ \mathfrak{S}_n/\mathfrak{A}_n \ \cong\ \mathbb{Z}/2\mathbb{Z}.$$

It follows that χ_{triv} and χ_{sign} are the only one-dimensional characters of $\mathfrak{S}_n$. A more elementary way of seeing this is that a one-dimensional character must map the conjugacy class of 2-cycles to either 1 or -1, and the 2-cycles generate all of $\mathfrak{S}_n$, hence determine the character completely.

For instance, suppose we want to compute the character table of $\mathfrak{S}_5$ (Problem 8.6), which has seven conjugacy classes. There are 21 lists of seven positive integers whose squares add up to $|\mathfrak{S}_5| = 5! = 120$, but only four of them that contain exactly two 1's:

$$1,1,2,2,2,5,9, \quad 1,1,2,2,5,6,7, \quad 1,1,2,3,4,5,8, \quad 1,1,4,4,5,5,6.$$

By examining the defining representation and using the tensor product, you should be able to figure out which one of these is the actual list of dimensions of irreps. ◀

Example 8.8.4. The **dicyclic group** $G = \text{Dic}_3$ can be presented as

$$\langle a, x \mid a^6 = 1,\ x^2 = a^3,\ x^{-1}ax = a^{-1} \rangle.$$

It has 12 elements and 6 conjugacy classes:

$$C_1 = \{1\},\ C_2 = \{a^3\},\ C_3 = \{a, a^5\},\ C_4 = \{a^2, a^4\},\ C_5 = \{x, a^2x, a^4x\},\ C_6 = \{ax, a^3x, a^5x\}.$$

In particular it must have six irreps, four of dimension 1 and two of dimension 2. That's a lot of 1's, so it is worth computing the commutator subgroup to get at the one-dimensional characters. It turns out that $[G, G] = \{1, a^2, a^4\}$, and the quotient G^{ab} is cyclic of order 4, generated by x. So the one-dimensional characters are as follows:

	1 C_1	1 C_2	2 C_3	2 C_4	3 C_5	3 C_6
χ_1	1	1	1	1	1	1
χ_2	1	1	1	1	-1	-1
χ_3	1	-1	-1	1	i	$-i$
χ_4	1	-1	-1	1	$-i$	i

The remaining two irreducible characters χ_5, χ_6 evidently satisfy

$$\chi_5 + \chi_6 = \frac{\chi_{\text{reg}} - \chi_1 - \chi_2 - \chi_3 - \chi_4}{2} = \frac{[12,0,0,0,0,0] - [4,0,0,4,0,0]}{2} = [4,0,0,-2,0,0].$$

Write them as

$$\chi_5 = [2, a, b, -1+c, d, e], \quad \chi_6 = [2, -a, -b, -1-c, -d, -e].$$

Tensoring with any one-dimensional character has to either preserve or swap both χ_5 and χ_6, and in particular tensoring with $\chi_2 = \chi_3 \otimes \chi_3$ must preserve it. But then $d = -d$ and $e = -e$, and our lives have just gotten somewhat easier, as we can write

$$\chi_5 = [2, a, b, -1+c, 0, 0], \quad \chi_6 = [2, -a, -b, -1-c, 0, 0].$$

Now let's take some scalar products:

$$\langle \chi_1,\, \chi_5 \rangle_G = 2 + a + 2b + 2(-1+c) = 0,$$
$$\langle \chi_3,\, \chi_5 \rangle_G = 2 - a - 2b + 2(-1+c) = 0.$$

Adding these equations gives $4 + 4(-1+c) = 0$, or $c = 0$; subtracting them gives $a = -2b$. At this point we know that the C_3 and C_4 columns of the character table are $(1, 1, -1, -1, b, -b)$ and $(1, 1, 1, 1, -1, 1)$ respectively. By Proposition 8.6.7 they are orthogonal, i.e., $2 - 2b = 0$, or $b = 1$. So the final character table is as follows:

	1 C_1	1 C_2	2 C_3	2 C_4	3 C_5	3 C_6
χ_1	1	1	1	1	1	1
χ_2	1	1	1	1	-1	-1
χ_3	1	-1	-1	1	i	$-i$
χ_4	1	-1	-1	1	$-i$	i
χ_5	2	-2	1	-1	0	0
χ_6	2	2	-1	1	0	0

◀

8.9 Restriction, induction, and Frobenius reciprocity

Let $H \subseteq G$ be finite groups. Representations of G give rise to representations of H via an (easy) process called **restriction**, and representations of H give rise to representations of G via a (somewhat more involved) process called **induction**. These processes are sources of more characters to put in character tables, and the two are related by an equation called **Frobenius reciprocity.**

1. Restriction. Let $\rho : G \to GL(V)$ be a representation of G. Then the restriction of ρ to H is a representation of G, denoted $\mathrm{Res}^G_H(\rho)$. Alternative notations include $\mathrm{Res}(\rho)$ (if the groups G and H are clear from context) and $\rho{\downarrow}^G_H$. On the level of characters, we define

$$\mathrm{Res}(\chi_\rho) = \chi_\rho|_H = \chi(\mathrm{Res}(\rho)).$$

That is, restricting a representation does not change its character on the level of group elements. On the other hand, the restriction of an irreducible representation is not always irreducible. Also, two elements conjugate in G are not necessarily conjugate in a subgroup H, so restriction is not as trivial an operation as it first might appear.

Example 8.9.1. Let C_λ denote the conjugacy class in $\mathfrak{S}_n$ of permutations of cycle-type λ. Recall that the standard representation ρ_{std} of $G = \mathfrak{S}_3$ has character $[2, 0, -1]$ on the conjugacy classes C_{111}, C_{21}, C_3. Moreover, $\langle \chi_{\mathrm{std}},\, \chi_{\mathrm{std}} \rangle_G = 1$ because ρ_{std} is irreducible.

Now let $H = \mathfrak{A}_3 < \mathfrak{S}_3$. This is an abelian group isomorphic to $\mathbb{Z}/3\mathbb{Z}$, so the two-dimensional representation $\mathrm{Res}(\rho_{\mathrm{std}})$ cannot possibly be irreducible. In fact $H = C_{111} \cup C_3$, so $\langle \chi_{\mathrm{std}},\, \chi_{\mathrm{std}} \rangle_G = (1 \cdot 2^2 + 2 \cdot (-1)^2)/3 = 2$. (We knew this already, since if a 2-dimensional representation is not irreducible then it must be the direct sum of two one-dimensional irreps.) The group $\mathfrak{A}_3$ is cyclic, so its character table is

	Id_G	$(1\,2\,3)$	$(1\,3\,2)$
χ_{triv}	1	1	1
χ_1	1	ω	ω^2
χ_2	1	ω^2	ω

(8.12)

where $\omega = e^{2\pi i/3}$. (Note also that the conjugacy class $C_3 \subseteq \mathfrak{S}_3$ splits into two singleton conjugacy classes in $\mathfrak{A}_3$.) Now it is evident that $\mathrm{Res}(\chi_{\mathrm{std}}) = [2, -1, -1] = \chi_1 + \chi_2$. ◀

2. Induction. Let $H \subseteq G$ be groups and let (ρ, W) be a representation of H. We want to "lift" ρ to a representation of G. The ability to do so is very powerful, because it is typically easy to get our hands on subgroups of G (e.g., every element of G generates a cyclic subgroup) and so this lifting operation can potentially be a rich source of representations of G.

Let $n = [G : H] = |G|/|H|$. Choose coset representatives $B = \{b_1, \ldots, b_n\} \subseteq G$, so that $G = b_1 H \cup \cdots \cup b_n H$. Let V be the vector space defined as a direct sum of copies of W, one for each (left) coset of H in G. We can think of V as a tensor product:

$$V = \mathbb{C}B \otimes W = (b_1 \otimes W) \oplus \cdots \oplus (b_n \otimes W).$$

To say how a group element $g \in G$ acts on the summand $b_i \otimes W$, we need to write gb_i in the form $b_j h$, where $j \in [n]$ and $h \in H$. Note that for each b_i, there is a unique pair b_j, h that satisfies these conditions. We then make g act by

$$g(b_i \otimes w) = b_j \otimes (h \cdot w) = b_j \otimes \rho(h)(w), \tag{8.13}$$

extended linearly to all of V. Heuristically, this formula is justified by the equation

$$g(b_i \otimes w) = gb_i \otimes w = b_j h \otimes w = b_j \otimes hw = b_j \otimes \rho(h)(w).$$

In other words, g sends $b_i \otimes W$ to $b_j \otimes W$, acting by h along the way. We have a map $\mathrm{Ind}_H^G(\rho)$ that sends each $g \in G$ to the linear transformation $V \to V$ just defined. Alternative notations for $\mathrm{Ind}_H^G(\rho)$ include $\mathrm{Ind}(\rho)$ (if G and H are clear from context) and $\rho \uparrow_H^G$.

On the level of matrices, $\mathrm{Ind}_H^G(\rho)(g)$ is an $n \times n$ block matrix with blocks B_{ij} of size $\dim \rho$, given by

$$B_{ij} = \begin{cases} \rho(h) & \text{if } gb_i = b_j h \text{ for some } h \in H \\ 0 & \text{otherwise} \end{cases} = \begin{cases} \rho(b_j^{-1} g b_i) & \text{if } b_j^{-1} g b_i \in H \\ 0 & \text{otherwise.} \end{cases} \tag{8.14}$$

Example 8.9.2. Let $G = \mathfrak{S}_3$ and $H = \mathfrak{A}_3 = \{\mathrm{Id},\ (1\,2\,3),\ (1\,3\,2)\}$, and let (ρ, W) be a representation of H, where $W = \mathbb{C}\langle \mathbf{e}_1, \ldots, \mathbf{e}_n \rangle$. Let $B = \{b_1 = \mathrm{Id},\ b_2 = (1\,2)\}$, so that $V = b_1 \otimes W \oplus b_2 \otimes W$. To define $\mathrm{Ind}_G^H(\rho)$, we need to solve the equations $gb_i = b_j h$. That is, for each $g \in G$ and each $b_i \in B$, we need to determine the unique pair b_j, h that satisfy the equation.

	$i=1$				$i=2$			
g	gb_i	$=$	b_j	h	gb_i	$=$	b_j	h
Id	Id	$=$	b_1	Id	$(1\,2)$	$=$	b_2	Id
$(1\,2\,3)$	$(1\,2\,3)$	$=$	b_1	$(1\,2\,3)$	$(1\,3)$	$=$	b_2	$(1\,3\,2)$
$(1\,3\,2)$	$(1\,3\,2)$	$=$	b_1	$(1\,3\,2)$	$(2\,3)$	$=$	b_2	$(1\,2\,3)$
$(1\,2)$	$(1\,2)$	$=$	b_2	Id	Id	$=$	b_1	Id
$(1\,3)$	$(1\,3)$	$=$	b_2	$(1\,2\,3)$	$(1\,2\,3)$	$=$	b_1	$(1\,2\,3)$
$(2\,3)$	$(2\,3)$	$=$	b_2	$(1\,3\,2)$	$(1\,3\,2)$	$=$	b_1	$(1\,3\,2)$

Therefore, the representation $\mathrm{Ind}_H^G(\rho)$ sends the elements of $\mathfrak{S}_3$ to the following block matrices. Each block is of size $n \times n$; the first block corresponds to $b_1 \otimes W$ and the second block to $b_2 \otimes W$.

$$\mathrm{Id} \mapsto \begin{bmatrix} \rho(\mathrm{Id}) & 0 \\ 0 & \rho(\mathrm{Id}) \end{bmatrix} \qquad (1\,2\,3) \mapsto \begin{bmatrix} \rho(1\,2\,3) & 0 \\ 0 & \rho(1\,3\,2) \end{bmatrix} \qquad (1\,3\,2) \mapsto \begin{bmatrix} \rho(1\,3\,2) & 0 \\ 0 & \rho(1\,2\,3) \end{bmatrix}$$

$$(1\,2) \mapsto \begin{bmatrix} 0 & \rho(\mathrm{Id}) \\ \rho(\mathrm{Id}) & 0 \end{bmatrix} \qquad (1\,3) \mapsto \begin{bmatrix} 0 & \rho(1\,2\,3) \\ \rho(1\,2\,3) & 0 \end{bmatrix} \qquad (2\,3) \mapsto \begin{bmatrix} 0 & \rho(1\,3\,2) \\ \rho(1\,3\,2) & 0 \end{bmatrix}$$

For instance, if ρ is the 1-dimensional representation (= character) χ_1 of (8.12), then the character of $\mathrm{Ind}(\rho)$ is given on conjugacy classes in $\mathfrak{S}_3$ by

$$\chi_{\mathrm{Ind}(\rho)}(C_{111}) = 2, \qquad \chi_{\mathrm{Ind}(\rho)}(C_{21}) = 0, \qquad \chi_{\mathrm{Ind}(\rho)}(C_3) = \omega + \omega^2 = -1,$$

which implies that $\mathrm{Ind}_H^G(\rho)$ is the nontrivial summand of the defining representation of $\mathfrak{S}_3$. ◀

Example 8.9.3. Let $G = \mathfrak{S}_3$ and $H = \{\mathrm{Id},\ (1\,2)\}$. Let (ρ, W) be a representation of H, where $W = \mathbb{C}\langle \mathbf{e}_1, \ldots, \mathbf{e}_n \rangle$. Let $B = \{b_1 = \mathrm{Id},\ b_2 = (1\,3),\ b_3 = (2\,3)\}$ and $V = \mathbb{C}B \otimes W$. To define the induced representation, we take the equations $gb_i = b_j h$, for each g and i, and solve them for b_j and h:

	$i=1$				$i=2$				$i=3$			
g	gb_i	$=$	b_j	h	gb_i	$=$	b_j	h	gb_i	$=$	b_j	h
Id	Id	$=$	b_1	Id	$(1\,3)$	$=$	b_2	Id	$(2\,3)$	$=$	b_3	Id
$(1\,2)$	$(1\,2)$	$=$	b_1	$(1\,2)$	$(1\,3\,2)$	$=$	b_3	$(1\,2)$	$(1\,2\,3)$	$=$	b_2	$(1\,2)$
$(1\,3)$	$(1\,3)$	$=$	b_2	Id	Id	$=$	b_1	Id	$(1\,3\,2)$	$=$	b_3	$(1\,2)$
$(1\,2\,3)$	$(1\,2\,3)$	$=$	b_2	$(1\,2)$	$(2\,3)$	$=$	b_3	Id	$(1\,2)$	$=$	b_1	$(1\,2)$
$(2\,3)$	$(2\,3)$	$=$	b_3	Id	$(1\,2\,3)$	$=$	b_2	$(1\,2)$	Id	$=$	b_1	Id
$(1\,3\,2)$	$(1\,3\,2)$	$=$	b_3	$(1\,2)$	$(1\,2)$	$=$	b_1	$(1\,2)$	$(1\,3)$	$=$	b_2	Id

Thus $\mathrm{Ind}_H^G(\rho)$ is as follows:

$$\mathrm{Id} \mapsto \begin{bmatrix} \rho(\mathrm{Id}) & 0 & 0 \\ 0 & \rho(\mathrm{Id}) & 0 \\ 0 & 0 & \rho(\mathrm{Id}) \end{bmatrix} \qquad (1\,3) \mapsto \begin{bmatrix} 0 & \rho(\mathrm{Id}) & 0 \\ \rho(\mathrm{Id}) & 0 & 0 \\ 0 & 0 & \rho(1\,2) \end{bmatrix} \qquad (2\,3) \mapsto \begin{bmatrix} 0 & 0 & \rho(\mathrm{Id}) \\ 0 & \rho(1\,2) & 0 \\ \rho(\mathrm{Id}) & 0 & 0 \end{bmatrix}$$

$$(1\,2) \mapsto \begin{bmatrix} \rho(1\,2) & 0 & 0 \\ 0 & 0 & \rho(1\,2) \\ 0 & \rho(1\,2) & 0 \end{bmatrix} \quad (1\,2\,3) \mapsto \begin{bmatrix} 0 & \rho(1\,2) & 0 \\ 0 & 0 & \rho(\mathrm{Id}) \\ \rho(1\,2) & 0 & 0 \end{bmatrix} \quad (1\,3\,2) \mapsto \begin{bmatrix} 0 & 0 & \rho(1\,2) \\ \rho(1\,2) & 0 & 0 \\ 0 & \rho(\mathrm{Id}) & 0 \end{bmatrix}$$

◀

In fact, $\mathrm{Ind}(\rho)$ is a representation, and there is a general formula for its character. (That is a good thing, because as you see computing the induced representation itself is a lot of work.)

Proposition 8.9.4. *Let H be a subgroup of G and let (ρ, W) be a representation of H with character χ. Then $\mathrm{Ind}_H^G(\rho)$ is a representation of G, with character defined on $g \in G$ by*

$$\mathrm{Ind}_H^G(\chi)(g) \ = \ \frac{1}{|H|} \sum_{k \in G:\ k^{-1}gk \in H} \chi(k^{-1}gk).$$

In particular, $\mathrm{Ind}(\chi_\rho)$, hence $\mathrm{Ind}(\rho)$, is independent (up to isomorphism) of the choice of B.

(We know that characters are class functions, so why not write $\chi(g)$ instead of $\chi(k^{-1}gk)$? Because χ is a function on H, so the former expression is not well-defined in general.)

Proof. First, we verify that $\mathrm{Ind}(\rho)$ is a representation. Let $g, g' \in G$ and $b_i \otimes w \in V$. Then there is a unique $b_k \in B$ and $h \in H$ such that

$$gb_i = b_k h \tag{8.15}$$

and in turn there is a unique $b_\ell \in B$ and $h' \in H$ such that

$$g'b_k = b_\ell h'. \tag{8.16}$$

We need to verify that $g' \cdot (g \cdot (b_i \otimes w)) \;=\; (g'g) \cdot (b_i \otimes w)$. Indeed,

$$\Big(g' \cdot (g \cdot (b_i \otimes w)) \;=\; g' \cdot (b_k \otimes hw)\Big) \;=\; b_\ell \otimes h'hw.$$

On the other hand, by (8.15) and (8.16), $gb_i = b_k h b_i^{-1}$ and $g' = b_\ell h' b_k^{-1}$, so

$$(g'g) \cdot (b_i \otimes w) \;=\; (b_\ell h' h b_i^{-1}) \cdot (b_i \otimes w) \;=\; b_\ell \otimes h'hw$$

as desired. Note by the way that

$$\dim \mathrm{Ind}(\rho) = \frac{|G|}{|H|} \dim \rho.$$

Now that we know that $\mathrm{Ind}(\rho)$ is a representation of G on V, we calculate its character using (8.14):

$$\begin{aligned}\mathrm{Ind}(\chi)(g) = \sum_{i=1}^{n} \mathrm{tr}(B_{i,i}) &= \sum_{i \in [n]:\ b_i^{-1} g b_i \in H} \chi(b_i^{-1} g b_i) \\ &= \sum_{i \in [n]:\ b_i^{-1} g b_i \in H} \frac{1}{|H|} \sum_{h \in H} \chi(h^{-1} b_i^{-1} g b_i h)\end{aligned}$$

(because χ_ρ is constant on conjugacy classes of H — the averaging trick again!)

$$= \frac{1}{|H|} \sum_{k \in G:\ k^{-1} g k \in H} \chi(k^{-1} g k) \tag{8.17}$$

as desired. Here $k = b_i h$ runs over all elements of G as the indices of summation i, h on the previous sum run over $[r]$ and H respectively. (Also, $k^{-1}gk = h^{-1}b_i^{-1} g b_i h \in H$ if and only if $b_i^{-1} g b_i \in H$, simply because H is a group.) Since $\mathrm{Ind}(\chi)$ is independent of the choice of B, so is the isomorphism type of $\mathrm{Ind}(\rho)$. □

Corollary 8.9.5. *Let $H \subseteq G$ and let ρ be the trivial representation of H. Then*

$$\mathrm{Ind}_H^G(\chi_{\mathrm{triv}})(g) = \frac{\#\{k \in G :\ k^{-1} g k \in H\}}{|H|}.$$

Corollary 8.9.6. *Suppose H is a normal subgroup of G of index n. Then*

$$\mathrm{Ind}_H^G(\chi)(g) = \begin{cases} n\chi(g) & \text{if } g \in H, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. Normality implies that $k^{-1}gk \in H$ if and only if $g \in H$, independently of k. If $g \in H$ then the sum in (8.17) has $|G|$ terms, all equal to $\chi(g)$; otherwise, the sum is empty. (Alternative proof: normality implies that left cosets and right cosets coincide, so the blocks in the block matrix $\mathrm{Ind}(\rho)(g)$ will all be on the main diagonal (and equal to $\rho(g)$) when $g \in H$, and off the main diagonal otherwise.) □

Example 8.9.7. Let $n \geq 2$, so that $\mathfrak{A}_n$ is a normal subgroup of $\mathfrak{S}_n$ of index 2. By Corollary 8.9.5,

$$\mathrm{Ind}_{\mathfrak{A}_n}^{\mathfrak{S}_n}(\chi_{\mathrm{triv}})(g) = \begin{cases} 2 & \text{for } g \in \mathfrak{A}_3, \\ 0 & \text{for } g \notin \mathfrak{A}_3, \end{cases}$$

so $\mathrm{Ind}(\chi_{\mathrm{triv}})$ is the sum of the trivial and sign characters on $\mathfrak{S}_n$. ◀

Example 8.9.8. Let $G = \mathfrak{S}_n$ and $H = \{g \in \mathfrak{S}_n \mid g(n) = n\}$, which we identify with $\mathfrak{S}_{n-1}$. (Note that H is not normal.) Now Corollary 8.9.5 gives

$$\begin{aligned} \mathrm{Ind}_{\mathfrak{S}_{n-1}}^{\mathfrak{S}_n}(\chi_{\mathrm{triv}})(g) &= \frac{\#\{k \in \mathfrak{S}_n \mid (k^{-1}gk)(n) = n\}}{(n-1)!} \\ &= \frac{\#\{k \in \mathfrak{S}_n \mid k(n) \text{ is a fixed point of } g\}}{(n-1)!} \\ &= \#\{\text{fixed points of } g\} \\ &= \chi_{\mathrm{def}}(g) \end{aligned}$$

which implies that $\mathrm{Ind}_{\mathfrak{S}_{n-1}}^{\mathfrak{S}_n}(\rho_{\mathrm{triv}}) = \rho_{\mathrm{def}}$. ◀

Example 8.9.9. Let $G = \mathfrak{S}_4$ and let H be the non-normal subgroup $\{\mathrm{id},\ (1\ 2),\ (3\ 4),\ (1\ 2)(3\ 4)\}$. Let ρ be the trivial representation of G and χ its character. We can calculate $\psi = \mathrm{Ind}_H^G(\chi)$ using Corollary 8.9.5:

$$\psi(C_{1111}) = 6, \qquad \psi(C_{211}) = 2, \qquad \psi(C_{22}) = 2, \qquad \psi(C_{31}) = 0, \qquad \psi(C_4) = 0.$$

where, as usual, C_λ denotes the conjugacy class in $\mathfrak{S}_4$ of permutations with cycle-type λ. In the notation of Example 8.7.3, the decomposition into irreducible characters is $\chi_1 + \chi_2 + 2\chi_5$. ◀

Restriction and induction are related by the following useful formula.

Theorem 8.9.10 (Frobenius Reciprocity)**.** *Let $H \subseteq G$ be groups, let χ be a character of H, and let ψ be a character of G. Then*

$$\left\langle \mathrm{Ind}_H^G(\chi),\ \psi \right\rangle_G = \left\langle \chi,\ \mathrm{Res}_H^G(\psi) \right\rangle_H.$$

Proof.

$$\begin{aligned} \langle \mathrm{Ind}(\chi),\ \psi\rangle_G &= \frac{1}{|G|} \sum_{g \in G} \overline{\mathrm{Ind}(\chi)(g)} \cdot \psi(g) \\ &= \frac{1}{|G|} \sum_{g \in G} \frac{1}{|H|} \sum_{k \in G:\ k^{-1}gk \in H} \overline{\chi(k^{-1}gk)} \cdot \psi(g) && \text{(by Prop. 8.9.4)} \\ &= \frac{1}{|G||H|} \sum_{h \in H} \sum_{k \in G} \sum_{g \in G:\ k^{-1}gk = h} \overline{\chi(h)} \cdot \psi(k^{-1}gk) \\ &= \frac{1}{|G||H|} \sum_{h \in H} \sum_{k \in G} \overline{\chi(h)} \cdot \psi(h) && \text{(i.e., } g = khk^{-1}\text{)} \\ &= \frac{1}{|H|} \sum_{h \in H} \overline{\chi(h)} \cdot \psi(h) \ = \ \langle \chi,\ \mathrm{Res}(\psi)\rangle_H. \end{aligned}$$

□

This will be useful later; for now, here is a quick application.

Example 8.9.11. Frobenius reciprocity sometimes suffices to calculate the isomorphism type of an induced representation. Let $G = \mathfrak{S}_3$ and $H = \mathfrak{A}_3$, and let χ_{std}, χ_1 and χ_2 be as in Example 8.9.1. We would like to compute $\text{Ind}(\chi_1)$. By Frobenius reciprocity

$$\langle \text{Ind}(\chi_1),\ \chi_{\text{std}} \rangle_G = \langle \chi_1,\ \text{Res}(\chi_{\text{std}}) \rangle_H = 1.$$

But χ_{std} is irreducible and $\dim \chi_{\text{std}} = \dim \text{Ind}(\chi_1) = 2$. Therefore, it must be the case that $\text{Ind}(\chi_1) = \chi_{\text{std}}$, and the corresponding representations are isomorphic. The same is true if we replace χ_1 with χ_2. ◀

8.10 Characters of the symmetric group

We have worked out the irreducible characters of $\mathfrak{S}_3$, $\mathfrak{S}_4$ and $\mathfrak{S}_5$ *ad hoc* (the last as an exercise). In fact, we can do this for all n, exploiting a vast connection to the combinatorics of partitions and tableaux.

Recall (Defn. 1.2.4) that a **partition** of n is a sequence $\lambda = (\lambda_1, \ldots, \lambda_\ell)$ of weakly decreasing positive integers whose sum is n. We will sometimes drop the parentheses and commas. We write $\lambda \vdash n$ or $|\lambda| = n$ to indicate that λ is a partition of n. The number $\ell = \ell(\lambda)$ is the **length** of λ. The set of all partitions of n is $\mathsf{Par}(n)$, and the number of partitions of n is $p(n) = |\mathsf{Par}(n)|$. For example,

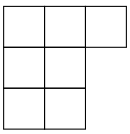
$$p(5) = |\mathsf{Par}(5)| = |\{5,\ 41,\ 32,\ 311,\ 221,\ 2111,\ 11111\}| = 7.$$

We will write Par for the set of all partitions. (As a set this is the same as Young's lattice.)

For each $\lambda \vdash n$, let C_λ be the conjugacy class in $\mathfrak{S}_n$ consisting of all permutations with cycle-type λ. Since the conjugacy classes are naturally indexed by $\mathsf{Par}(n)$, it makes sense to look for a set of representations indexed by partitions.

Definition 8.10.1. Let $\mu = (\mu_1, \ldots, \mu_\ell) \vdash n$.

- The **Ferrers diagram** of shape μ is the top- and left-justified array of boxes with μ_i boxes in the ith row.
- A **(Young) tableau**[6] **of shape μ** is a Ferrers diagram with the numbers $1, 2, \ldots, n$ placed in the boxes, one number to a box.
- Two tableaux T, T' of shape μ are **row-equivalent**, written $T \sim T'$, if the numbers in each row of T are the same as the numbers in the corresponding row of T'.
- A **(Young) tabloid** of shape μ is an equivalence class of tableaux under row-equivalence. A tabloid can be represented as a tableau without vertical lines.
- We write $\text{sh}(T) = \mu$ to indicate that a tableau or tabloid T is of shape μ.



Ferrers diagram

1	3	6
2	7	
4	5	

Young tableau

1	3	6
2	7	
4	5	

Young tabloid

[6] Terminology of tableaux is not consistent: some authors reserve the term "Young tableau" for a tableau in which the numbers increase downward and rightward. In these notes, I will call such a tableau a "standard tableau". For the moment, I am not placing any restrictions on which numbers can go in which boxes: thus there are $n!$ Young tableaux of shape μ for any $\mu \vdash n$.

A Young tabloid can be regarded as an ordered set partition $(T_1, \ldots, T_m)$ of $[n]$ in which $|T_i| = \mu_i$. The order of the blocks T_i matters, but not the order of entries within each block. Thus the number of tabloids of shape μ is

$$\binom{n}{\mu} = \frac{n!}{\mu_1! \cdots \mu_m!}.$$

The symmetric group $\mathfrak{S}_n$ acts on tabloids by permuting the numbers. This action gives rise to a permutation representation (ρ_μ, V_μ) of $\mathfrak{S}_n$, the ***μ*-tabloid representation** of $\mathfrak{S}_n$. Here V_μ is the vector space of all formal $\mathbb{C}$-linear combinations of tabloids of shape μ. The character of ρ_μ will be denoted τ_μ.

Example 8.10.2. For $n = 3$, the characters of the tabloid representations ρ_μ are as follows.

		Conjugacy classes		
		C_{111}	C_{21}	C_3
Characters	τ_3	1	1	1
	τ_{21}	3	1	0
	τ_{111}	6	0	0
	$\lvert C_\mu \rvert$	1	3	2

(8.18)

◀

Many familiar representations of $\mathfrak{S}_n$ can be expressed in this form.

- There is a unique tabloid of shape $\mu = (n)$. Every permutation fixes it, so
$$\rho_{(n)} \cong \rho_{\mathrm{triv}}.$$
- The tabloids of shape $\mu = (1, 1, \ldots, 1)$ are just the permutations of $[n]$. Therefore
$$\rho_{(1,1,\ldots,1)} \cong \rho_{\mathrm{reg}}.$$
- A tabloid of shape $\mu = (n-1, 1)$ is determined by its singleton part. So the representation ρ_μ is isomorphic to the action on the singleton by permutation, i.e.,
$$\rho_{(n-1,1)} \cong \rho_{\mathrm{def}}.$$

In fact, all tabloid representations can be obtained as inductions of an appropriate trivial representation. Some notation first. For a partition $\lambda = (\lambda_1, \ldots, \lambda_\ell) \vdash n$, define

$$\begin{aligned} \lambda_{[i]} &= \lambda_1 + \cdots + \lambda_i, & L_i &= [\lambda_{[i-1]} + 1, \lambda_{[i]}], \\ \lambda! &= \lambda_1! \cdots \lambda_\ell!, & \mathfrak{S}_\lambda &= \{\sigma \in \mathfrak{S}_n \mid \sigma(L_i) = L_i \ \forall i\}, \end{aligned} \tag{8.19}$$

so that $\mathfrak{S}_\lambda \cong \mathfrak{S}_{\lambda_1} \times \cdots \times \mathfrak{S}_{\lambda_\ell}$. This is called a **Young subgroup**; its cardinality is $\lambda!$. For example, if $\lambda = (4, 4, 2, 1) \vdash 11$, then $\mathfrak{S}_\lambda$ is the subgroup of $\mathfrak{S}_{11}$ mapping each of the sets $\{1, 2, 3, 4\}$, $\{5, 6, 7, 8\}$, $\{9, 10\}$, $\{11\}$ to itself. Note that $\mathfrak{S}_\lambda$ is not a normal subgroup of $\mathfrak{S}_n$ unless $\lambda = (n)$ or $\lambda = (1^n)$, because replacing $L_1, \ldots, L_\ell$ with any partition of $[n]$ into blocks of the same sizes gives a conjugate subgroup.

Proposition 8.10.3. *Let $\lambda = (\lambda_1, \ldots, \lambda_\ell) \vdash n$. Then* $\mathrm{Ind}_{\mathfrak{S}_\lambda}^{\mathfrak{S}_n}(\rho_{\mathrm{triv}}) \cong \rho_\lambda$.

Proof. We will show that the characters of these representations are equal, i.e., that $\mathrm{Ind}_{\mathfrak{S}_\lambda}^{\mathfrak{S}_n}(\chi_{\mathrm{triv}}) = \tau_\lambda$.

Assign labels $1, \ldots, n$ to the cells of the Ferrers diagram of λ reading from left to right and top to bottom, so that the cells in the ith row are labeled by L_i. For every $w \in \mathfrak{S}_n$, let $T_{\lambda,w}$ be the tableau of shape λ in which cell k is filled with the number $w(k)$.

Let g be a permutation of cycle-type $\mu = (\mu_1, \ldots, \mu_k)$, say

$$g = (1 \ \cdots \ \mu_{[1]})(\mu_{[1]}+1 \ \cdots \ \mu_{[2]}) \ \cdots \ (\mu_{[k-1]}+1 \ \cdots \ \mu_{[k]}), \tag{8.20}$$

and let $M_j = [\mu_{[j-1]}+1, \mu_{[j]}]$ be the support of the jth cycle of g. Then, by Corollary 8.9.5 (replacing w with w^{-1} for brevity),

$$\begin{aligned}
\mathrm{Ind}_{\mathfrak{S}_\lambda}^{\mathfrak{S}_n}(\chi_{\mathrm{triv}})(g) &= \frac{1}{\lambda!}\#\{w \in \mathfrak{S}_n : \ wgw^{-1} \in \mathfrak{S}_\lambda\} \\
&= \frac{1}{\lambda!}\#\left\{w \in \mathfrak{S}_n : \ \Big(w(1) \ \cdots \ w(\mu_1)\Big) \ \cdots \ \Big(w(n-\mu_k+1) \ \cdots \ w(n)\Big) \in \mathfrak{S}_\lambda\right\} \\
&= \frac{1}{\lambda!}\#\left\{w \in \mathfrak{S}_n : \ \forall j \ \exists i : \ w(M_j) \subseteq L_i\right\} \\
&= \frac{1}{\lambda!}\#\left\{\text{tableaux } T_{\lambda,w} \text{ with all elements of } M_j \text{ in the same row}\right\} \\
&= \#\left\{\text{tabloids of shape } \lambda \text{ with all elements of } M_j \text{ in the same row}\right\} \\
&= \#\left\{\text{tabloids of shape } \lambda \text{ fixed by } g\right\} \\
&= \tau_\lambda(g) \\
&= \tau_\lambda(C_\mu).
\end{aligned}$$

□

For $n = 3$, the table in (8.18) is a triangular matrix. In particular, the characters τ_μ are linearly independent, hence a basis, in the vector space $C\ell(\mathfrak{S}_3)$. We will prove that this is the case for all n. We will need two orders on the set $\mathsf{Par}(n)$.

Definition 8.10.4. Let $\lambda, \mu \in \mathsf{Par}(n)$.

1. **Lexicographic ("lex") order:** $\lambda < \mu$ if for some $k > 0$ we have
$$\lambda_1 = \mu_1, \quad \lambda_2 = \mu_2, \quad \ldots, \quad \lambda_{k-1} = \mu_{k-1}, \quad \lambda_k < \mu_k.$$
2. **Dominance order:** $\lambda \lhd \mu$ ("λ is dominated by μ") if $\lambda \neq \mu$ and $\lambda_{[k]} \leq \mu_{[k]}$ for all k.

Lex order is a total order on $\mathsf{Par}(n)$. For instance, if $n = 5$, lex order is

$$(5) \ > \ (4,1) \ > \ (3,2) \ > \ (3,1,1) \ > \ (2,2,1) \ > \ (2,1,1,1) \ > \ (1,1,1,1,1).$$

("Lex-greater partitions are short and wide; lex-smaller ones are tall and skinny.")

Dominance is a partial order on $\mathsf{Par}(n)$. It first fails to be a total order for $n = 6$ (neither of 33 and 411 dominates the other). Lex order is a linear extension of dominance order: if $\lambda \lhd \mu$ then $\lambda < \mu$.

Since the tabloid representations ρ_μ are permutation representations, we can calculate τ_μ by counting fixed points. That is, for any permutation $w \in C_\lambda$, we have

$$\begin{aligned}
\tau_\mu(C_\lambda) = \tau_\mu(w) &= \#\{\text{tabloids } T\colon \ \mathrm{sh}(T) = \mu, \ w(T) = T\} \\
&= \#\{\text{tabloids } T \text{ of shape } \mu \text{ such that every cycle of } w \text{ is contained in some row of } T\}.
\end{aligned} \tag{8.21}$$

Proposition 8.10.5. *Let $\lambda, \mu \vdash n$. Then:*

1. *$\tau_\mu(C_\mu) \neq 0$.*
2. *$\tau_\mu(C_\lambda) \neq 0$ only if $\lambda \trianglelefteq \mu$ (thus, only if $\lambda \leq \mu$ in lex order).*

Proof. First, let $w \in C_\mu$. Take T to be any tabloid whose blocks are the cycles of w; then $wT = T$. For example, if $w = (1\ 3\ 6)(2\ 7)(4\ 5) \in \mathfrak{S}_7$, then T can be either of the following two tabloids:

1	3	6
2	7	
4	5	

1	3	6
4	5	
2	7	

It follows from (8.21) that $\tau_\mu(C_\mu) \neq 0$. In fact, $\tau_\mu(C_\mu) = \prod_j r_j!$, where r_j is the number of occurrences of j in μ.

For the second assertion, observe that $w \in \mathfrak{S}_n$ fixes a tabloid T of shape μ if and only if every cycle of w is contained in a row of T. This is possible only if, for every k, the largest k rows of T are collectively big enough to hold the k largest cycles of w. This is precisely the condition $\lambda \trianglelefteq \mu$. □

Dominance is not enough for $\tau_\mu(C_\lambda)$ to be nonzero: for instance, take $\lambda = (2,2)$ and $\mu = (3,1)$.

Corollary 8.10.6. *The characters $\{\tau_\mu : \mu \vdash n\}$ form a basis for $C\ell(\mathfrak{S}_n)$.*

Proof. Make the characters into a $p(n) \times p(n)$ matrix $X = [\tau_\mu(C_\lambda)]_{\mu,\lambda\vdash n}$ with rows and columns ordered by lex order on $\mathsf{Par}(n)$. By Proposition 8.10.5, X is a triangular matrix with nonzero entries on the diagonal, so it is nonsingular. □

We can transform the characters τ_μ into a list of irreducible characters χ_μ of $\mathfrak{S}_n$ by applying the Gram-Schmidt process with respect to the inner product $\langle \cdot, \cdot \rangle_{\mathfrak{S}_n}$. We will start with $\mu = (n)$ and work our way up in lex order. In fact, the τ_μ are not only a vector space basis for $C\ell(\mathfrak{S}_n)$, but an integral basis for the $\mathbb{Z}$-module $Cl_{\mathbb{Z}}(\mathfrak{S}_n)$ of virtual characters (although we cannot prove that statement at this point). Thus no division will be required in the Gram-Schmidt process. Each tabloid character τ_μ will decompose as

$$\tau_\mu = \chi_\mu + \sum_{\lambda<\mu} K_{\lambda,\mu}\chi_\lambda \tag{8.22}$$

for certain integers $K_{\lambda,\mu}$ called **Kostka numbers**. We will have more to say about them in the next chapter. For the time being, we will only be able to observe (8.22) for particular examples, including $\mathfrak{S}_3$ and $\mathfrak{S}_4$, but we will eventually be able to prove it in general (Corollary 9.11.3). The irrep with character χ_μ is called a **Specht module**. There is an independent construction of Specht modules, which I have not written up yet.

Example 8.10.7. We will use tabloid representations to derive the character tables of $\mathfrak{S}_3$ and $\mathfrak{S}_4$.

Recall the table of characters (8.18) of the tabloid representations for $G = \mathfrak{S}_3$. Here is how the Gram-Schmidt process goes.

First, $\tau_3 = [1,1,1] = \chi_{\text{triv}}$ is irreducible, so we label it as χ_3. (This is χ_{triv}.)

Second, $\langle \tau_{21},\ \chi_3 \rangle_G = 1$. Thus $\tau_{21} - \chi_3 = [2,0,-1]$ is orthogonal to χ_3, and in fact it is irreducible, so we label it as χ_{21}. (This is χ_{std}.)

Third, feed τ_{111} into the Gram-Schmidt machine:

$$\begin{aligned}\chi_{111} &= \tau_{111} - \langle \tau_{111},\ \chi_3 \rangle_G\, \chi_3 - \langle \tau_{111},\ \chi_{21} \rangle_G\, \chi_{21} \\ &= \tau_{111} - \chi_3 - 2\chi_{21} \\ &= [1,-1,1]\end{aligned}$$

which is 1-dimensional, hence irreducible (in fact it is χ_{sign}). To summarize,

$$\begin{bmatrix} \tau_3 \\ \tau_{21} \\ \tau_{111} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 3 & 1 & 0 \\ 6 & 0 & 0 \end{bmatrix} = \overbrace{\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 2 & 1 \end{bmatrix}}^{K} \overbrace{\begin{array}{c} \begin{array}{ccc} C_{111} & C_{21} & C_3 \end{array} \\ \begin{bmatrix} 1 & 1 & 1 \\ 2 & 0 & -1 \\ 1 & -1 & 1 \end{bmatrix} \end{array}}^{X} \begin{array}{c} \chi_3 \\ \chi_2 \\ \chi_{111} \end{array} \tag{8.23}$$

where X is the character table. For $\mathfrak{S}_4$, the same procedure produces

$$\begin{bmatrix} \tau_4 \\ \tau_{31} \\ \tau_{22} \\ \tau_{211} \\ \tau_{1111} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 4 & 2 & 0 & 1 & 0 \\ 6 & 2 & 2 & 0 & 0 \\ 12 & 2 & 0 & 0 & 0 \\ 24 & 0 & 0 & 0 & 0 \end{bmatrix} = \overbrace{\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 1 & 0 \\ 1 & 3 & 2 & 3 & 1 \end{bmatrix}}^{K} \overbrace{\begin{array}{c} \begin{array}{ccccc} C_{1111} & C_{211} & C_{22} & C_{31} & C_4 \end{array} \\ \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 3 & 1 & -1 & 0 & -1 \\ 2 & 0 & 2 & -1 & 0 \\ 3 & -1 & -1 & 0 & 1 \\ 1 & -1 & 1 & 1 & -1 \end{bmatrix} \end{array}}^{X} \begin{array}{c} \chi_4 \\ \chi_3 \\ \chi_{22} \\ \chi_{211} \\ \chi_{1111} \end{array}.$$

The matrix K has 1's on the main diagonal, which witnesses the earlier claim that the tabloid characters form an integral basis for $C\ell_{\mathbb{Z}}(\mathfrak{S}_n)$. ◀

At this point you should feel a bit dissatisfied, since I have not told you what the values of the irreducible characters actually are in general, just that you can obtain them from the tabloid characters plus Gram-Schmidt. That is a harder problem; the answer is given by the *Murnaghan–Nakayama Rule* (see §9.14), which expresses the values of the irreducible characters as signed counts of certain tableaux.

I have also not told you the multiplicities of the irreps in the tabloid representations, i.e., the numbers in the matrix K. For $\mathfrak{S}_3$ and $\mathfrak{S}_4$ these matrices are unitriangular (1's on the main diagonal and 0's above); in fact this property holds for all $\mathfrak{S}_n$. Thus the tabloid characters are not just a vector basis for $C\ell(\mathfrak{S}_n)$, but, more strongly, a basis for the free abelian group generated by irreducible characters. We will prove this eventually (Corollary 9.11.3), by which point we will have a combinatorial description of the entries of K.

8.11 Exercises

In all exercises, unless otherwise specified, G is a finite group and (ρ, V) and (ρ', V') are finite-dimensional representations of G over $\mathbb{C}$.

Problem 8.1. Let χ be an irreducible character of G and let ψ be a one-dimensional character. Prove that $\omega := \chi \otimes \psi$ is an irreducible character.

Problem 8.2. Let $n \geq 2$. Prove that the standard representation ρ_{std} of $\mathfrak{S}_n$ (see Example 8.3.4) is irreducible. (Hint: Compute $\langle \chi_{\text{def}}, \chi_{\text{def}} \rangle$ and $\langle \chi_{\text{def}}, \chi_{\text{triv}} \rangle$. The latter boils down to finding the expected number of fixed points in a permutation selected uniformly at random; this is an old classic that uses what is essentially linearity of expectation.)

Problem 8.3. Let G be a group of order 63. Prove that G cannot have exactly 5 conjugacy classes. (You are encouraged to use a computer for part of this problem.)

Problem 8.4. Let $X = \{12|34, 13|24, 14|23\}$ be the set of set partitions of $[4]$ into two doubletons, and let $V = \mathbb{C}X$. The standard permutation action of $\mathfrak{S}_4$ on $\{1, 2, 3, 4\}$ induces an action on X. On the level of representations, the defining representation ρ_{def} induces a 3-dimensional representation (ρ, V).

(a) Compute the character of ρ.
(b) Show that ρ is the direct sum of the trivial representation and one other irrep.
(c) Explicitly describe all G-equivariant linear transformations $\phi : \rho \to \rho_{\text{def}}$. (Hint: Schur's lemma should be useful.)

Problem 8.5. Prove that the irreps of a direct product $G \times G'$ are exactly the direct products (see Example 8.1.11) of the irreps of G with irreps of G'. (This was mentioned in §8.8 for abelian groups but in fact is true in general.)

Problem 8.6. Work out the character table of $\mathfrak{S}_5$ without using any of the material in Section 8.10. (Hint: To construct another irreducible character, start by considering the action of $\mathfrak{S}_5$ on the edges of the complete graph K_5 induced by the usual permutation action on the vertices.)

Problem 8.7. Work out the character table of the **quaternion group** Q; this is the group of order 8 whose elements are $\{\pm 1, \pm i, \pm j, \pm k\}$ with relations $i^2 = j^2 = k^2 = -1$, $ij = k$, $jk = i$, $ki = j$.

Problem 8.8. Work out the irreducible characters of $\mathfrak{S}_5$ using tabloid characters. Feel free to use a computer algebra system to automate the tedious parts. Compare your result to the character table of $\mathfrak{S}_5$ calculated ad hoc in Problem 8.6. Make as many observations or conjectures as you can about how the partition λ is related to the values of the character χ_λ, and about the Kostka numbers.

Problem 8.9. Recall that the *alternating group* $\mathfrak{A}_n$ consists of the $n!/2$ even permutations in $\mathfrak{S}_n$, that is, those with an even number of even-length cycles.

(a) Show that the conjugacy classes in $\mathfrak{A}_4$ are not simply the conjugacy classes in $\mathfrak{S}_4$. (Hint: Consider the possibilities for the dimensions of the irreducible characters of $\mathfrak{A}_4$.)
(b) Determine the conjugacy classes in $\mathfrak{A}_4$
(c) Use this information to determine $[\mathfrak{A}_4, \mathfrak{A}_4]$ without computing any more commutators.
(d) Now compute the character table of $\mathfrak{A}_4$.

Chapter 9

Symmetric Functions

9.1 Prelude: Symmetric polynomials

Definition 9.1.1. Let R be an integral domain, typically $\mathbb{Q}$ or $\mathbb{Z}$. A **symmetric polynomial** is a polynomial in $R[x_1, \ldots, x_n]$ that is invariant under permuting the variables.

The symmetric polynomials that are homogeneous of degree d form a finitely generated, free R-module $\Lambda_d(R)$. For example, if $n = 3$, then up to scalar multiplication, the only symmetric polynomial of degree 1 in x, y, z is $x + y + z$. In degree 2, here are two:

$$x^2 + y^2 + z^2, \qquad xy + xz + yz.$$

Every other symmetric polynomial that is homogeneous of degree 2 is a R-linear combination of these two, because the coefficients of x^2 and xy determine the coefficients of all other monomials. Similarly, the polynomials

$$\begin{aligned} m_3(x, y, z) &= x^3 + y^3 + z^3, \\ m_{21}(x, y, z) &= x^2y + xy^2 + x^2z + xz^2 + y^2z + yz^2, \\ m_{111}(x, y, z) &= xyz. \end{aligned}$$

are a basis for the space of degree-3 symmetric polynomials in $R[x, y, z]$. Note that each symmetric polynomial in this list is a sum of the monomials in a single orbit under the action of $\mathfrak{S}_3$, and it is indexed by the partition whose parts are the exponents of one (hence any) of its monomials. In general, for $\lambda = (\lambda_1, \ldots, \lambda_\ell)$, the **monomial symmetric polynomial** is defined by

$$m_\lambda(x_1, \ldots, x_n) = \sum_{\{a_1, \ldots, a_\ell\} \subseteq [n]} x_{a_1}^{\lambda_1} x_{a_2}^{\lambda_2} \cdots x_{a_\ell}^{\lambda_\ell}.$$

There is a problem with this definition: the sum is empty if $\ell > n$. So if we want to construct a basis for the symmetric polynomials indexed by partitions, n variables is not enough. To be able to handle arbitrary partitions, we need a *countably infinite* set of variables $\{x_1, x_2, \ldots\}$, which means that we need to work not with polynomials, but with **formal power series**.

9.2 Formal power series

Let R be an integral domain (typically $\mathbb{Z}$ or a field), and let $\mathbf{x} = \{x_1, x_2, \dots\}$ be a countably infinite set of commuting indeterminates. A **monomial** is a product $\mathbf{x}^\alpha = \prod_{i=1}^{\infty} x_i^{\alpha_i}$, where $\alpha_i \in \mathbb{N}$ for all i and $\sum_{i \in I} \alpha_i$ is finite (equivalently, all but finitely many of the α_i are zero). The sequence $\alpha = (\alpha_1, \alpha_2, \dots)$ is called the **exponent vector** of the monomial; listing the nonzero entries of α in decreasing order gives a partition $\lambda(\alpha)$. A **formal power series** is an expression

$$\sum_{\alpha} c_\alpha \mathbf{x}^\alpha$$

with $c_\alpha \in R$ for all α. Equivalently, a formal power series can be regarded as a function from monomials to R, mapping $\mathbf{x}^\alpha$ to c_α. (Also equivalently, the R-module of all formal power series is thus the direct product (not the direct sum) of infinitely many copies of R, one for each exponent vector.) We often use the notation

$$[\mathbf{x}^\alpha]F = \text{coefficient of monomial } \mathbf{x}^\alpha \text{ in the power series } F.$$

The set $R[[\mathbf{x}]]$ of all formal power series is an abelian group under addition, and in fact an R-module, namely the direct product of countably infinitely many copies of R, one for each exponent vector. [1] In fact, $R[[\mathbf{x}]]$ is a ring as well, with multiplication given by

$$\left(\sum_{\alpha \in \mathbb{N}^I} c_\alpha \mathbf{x}^\alpha\right)\left(\sum_{\beta \in \mathbb{N}^I} d_\beta \mathbf{x}^\beta\right) = \sum_{\gamma \in \mathbb{N}^I} \left(\sum_{(\alpha,\beta):\ \alpha+\beta=\gamma} c_\alpha d_\beta\right) \mathbf{x}^\gamma.$$

because the inner sum on the right-hand side has only finitely many terms for each γ, and is thus a well-defined element of R.

We are generally not concerned with whether (or where) a formal power series converges in the sense of calculus, since we rarely need to plug in real values for the indeterminates x_i (and when we do, analytic convergence is not usually an issue). All that matters is that every operation must produce a well-defined power series, in the sense that each coefficient is given by a finite computation in the base ring R. For example, multiplication of power series satisfies this criterion, as explained above.[2]

Familiar functions from analysis (like exp and log) can be regarded as formal power series, namely their Taylor series. However, we will typically study them using combinatorial rather than analytic methods. For instance, from this point of view, we would justify equating the function $1/(1-x)$ as equal to the power series $1 + x + x^2 + \cdots$ not by calculating derivatives of $1/(1-x)$, but rather by observing that the identity $(1-x)(1+x+x^2+\cdots) = 1$ holds in $\mathbb{Z}[[x]]$. (That said, combinatorics also gets a lot of mileage out of working with derivative operators — but treating them formally, as linear transformations that map monomials to other monomials, rather than analytically.) Very often, analytical identities among power series can be proved using combinatorial methods; see Problem 9.6 for an example.

9.3 Symmetric functions

We can now define symmetric functions properly, as elements of the ring of formal power series $\mathbb{C}[[\mathbf{x}]] = \mathbb{C}[[x_1, x_2, \dots]]$.

[1] By contrast, the polynomial ring $R[\mathbf{x}]$ is the direct *sum* of countably infinitely many copies of R.

[2] We would have a problem with multiplication if we allowed *two-way-infinite* series. For example, the square of $\sum_{n \in \mathbb{Z}} x^n$ is not well-defined.

Definition 9.3.1. Let $\lambda \vdash n$. The **monomial symmetric function** m_λ is the power series

$$m_\lambda = \sum_{\alpha:\ \lambda(\alpha)=\lambda} \mathbf{x}^\alpha$$

where, as before, α runs over all infinite lists of nonnegative integers in which all but finitely many entries are zero.

For example,

$$\begin{aligned} m_3 &= x_1^3 + x_2^3 + x_3^3 + \cdots && = \sum_{i=1}^{\infty} x_i^3, \\ m_{21} &= x_1^2x_2 + x_1x_2^2 + x_1^2x_3 + x_1x_2^3 + x_2^2x_3 + x_2x_3^2 + \cdots && = \sum_{i\neq j} x_i^2x_j, \\ m_{111} &= x_1x_2x_3 + x_1x_2x_4 + x_1x_3x_4 + x_2x_3x_4 + x_1x_2x_5 + \cdots && = \sum_{1\leq i<j<k} x_ix_jx_k. \end{aligned}$$

Definition 9.3.2. The **ring of symmetric functions over** R is

$$\Lambda = \Lambda_R(\mathbf{x}) = \bigoplus_{d\geq 0} \Lambda_d$$

where

$$\Lambda_d = \Lambda_{R,d}(\mathbf{x}) = \{\text{degree-}d \text{ symmetric functions in indeterminates } \mathbf{x} \text{ with coefficients in } R\}.$$

Each Λ_d is a finitely generated free R-module, with basis $\{m_\lambda : \lambda \vdash d\}$, and their direct sum Λ is a graded R-algebra. If we let $\mathfrak{S}_\infty$ be the group whose members are the permutations of $\{x_1, x_2, \dots\}$ with only finitely many non-fixed points (equivalently, $\mathfrak{S}_\infty = \bigcup_{n=1}^{\infty} \mathfrak{S}_n$), then Λ is the ring of formal power series that have bounded degree and that are invariant under the action of $\mathfrak{S}_\infty$.

The monomial symmetric functions are the most obvious basis for Λ from an algebraic point of view, in the sense that each m_λ is the orbit under $\mathfrak{S}_\infty$ of any single monomial in it. On the other hand, there are many other bases that arise more frequently in combinatorics. Understanding symmetric functions requires familiarity with these various bases and how they interact.

One piece of terminology: we say that a basis B of Λ is an **integral basis** if the symmetric functions with integer coefficients are precisely the integer linear combinations of elements of B. Evidently, $\{m_\lambda\}$ is an integral basis. This condition is stronger than being a vector space basis for Λ; for example, integral bases are not preserved by scaling.

9.4 Elementary symmetric functions

Definition 9.4.1. The ***k*th elementary symmetric function** e_k is the sum of all squarefree monomials of degree k. That is,

$$\begin{aligned} &e_0 = 1, \\ &e_k = \sum_{\substack{S\subseteq\mathbb{N}_{>0} \\ |S|=k}} \prod_{s\in S} x_s \ = \sum_{0<i_1<i_2<\cdots<i_k} x_{i_1}x_{i_2}\cdots x_{i_k} \qquad \text{for } k>0, \end{aligned}$$

Equivalently, $e_k = m_{1^k}$, where 1^k means the partition with k $1's$. For $\lambda = (\lambda_1, \ldots, \lambda_\ell) \in \mathsf{Par}$, we define

$$e_\lambda = e_{\lambda_1} \cdots e_{\lambda_\ell}.$$

In general, we say that a basis for Λ is **multiplicative** if it is defined on partitions in this way. (Note that $\{m_\lambda\}$ is *not* multiplicative, which is why Problem 9.1 is nontrivial.)

For example, $e_{11} = (x_1+x_2+x_3+\cdots)^2 = (x_1^2+x_2^2+\cdots)+2(x_1x_2+x_1x_3+x_2x_3+x_1x_4+\cdots) = m_2+2m_{11}$. In degree 3, we have

$$\begin{array}{rcll} e_{111} &=& (x_1+x_2+x_3+\cdots)^3 &= 6m_{111}+3m_{21}+m_3, \\ e_{21} &=& (x_1+x_2+x_3+\cdots)(x_1x_2+x_1x_3+x_2x_3+x_1x_4+\cdots) &= 3m_{111}+m_{21}, \\ e_3 &=& \sum_{i<j<k} x_ix_jx_k &= m_{111}. \end{array}$$

Apparently $\{e_3, e_{21}, e_{111}\}$ is an R-basis for Λ_3, because the transition matrix is unitriangular and therefore invertible over every R. This works for $n = 4$ as well, where

$$\begin{bmatrix} e_{1111} \\ e_{211} \\ e_{22} \\ e_{31} \\ e_4 \end{bmatrix} = \begin{bmatrix} 24 & 12 & 6 & 4 & 1 \\ 12 & 5 & 2 & 1 & 0 \\ 6 & 2 & 1 & 0 & 0 \\ 4 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} m_{1111} \\ m_{211} \\ m_{22} \\ m_{31} \\ m_4 \end{bmatrix}.$$

This matrix is again unitriangular, and notably is symmetric across the northwest/southeast diagonal — that is, the coefficient of e_λ in m_μ equals the coefficient of e_μ in m_λ.

Sage can do this computation, by the way:

```
## Input
n = 3
e = SymmetricFunctions(QQ).elementary()
m = SymmetricFunctions(QQ).monomial()
for lam in Partitions(n):
    m(e[lam])
## Output
m[1, 1, 1]
3*m[1, 1, 1] + m[2, 1]
6*m[1, 1, 1] + 3*m[2, 1] + m[3]
```

Let $\trianglerighteq$ denote the dominance partial order on partitions (see Definition 8.10.4). Also, for a partition λ, let $\tilde{\lambda}$ be its **conjugate**, given by transposing the Ferrers diagram (see the discussion after Example 1.2.4).

Theorem 9.4.2. *Let $\lambda, \mu \vdash n$, with $\ell = \ell(\lambda)$ and $k = \ell(\mu)$. Let $b_{\lambda,\mu}$ be the coefficient[3] of e_λ when expanded in the monomial basis, that is,*

$$e_\lambda = \sum_\mu b_{\lambda,\mu} m_\mu.$$

Then $b_{\lambda,\tilde{\lambda}} = 1$, and $b_{\lambda,\mu} = 0$ unless $\tilde{\lambda} \trianglerighteq \mu$. In particular, $\{e_\lambda : \lambda \vdash n\}$ is an integral basis for Λ_n.

[3] Stanley [Sta99, §7.4] writes $M_{\lambda,\mu}$.

Proof. Say that a **$\boldsymbol{\lambda}$-factorization** of a monomial is a factorization into monomials of degrees $\lambda_1, \dots, \lambda_\ell$. Let $\mathbf{x}^\mu = x_1^{\mu_1} \cdots x_k^{\mu_k}$. Then

$$\begin{aligned} b_{\lambda,\mu} &= \text{coefficient of } m_\mu \text{ in } m\text{-expansion of } e_\lambda \\ &= [\mathbf{x}^\mu] e_{\lambda_1} \cdots e_{\lambda_\ell} \\ &= \text{number of } \lambda\text{-factorizations of } \mathbf{x}^\mu \text{ into squarefree monomials } \mathbf{x}^{\alpha_1}, \dots, \mathbf{x}^{\alpha_\ell}. \end{aligned}$$

Represent such a λ-factorization of $\mathbf{x}^\mu$ by a tableau T of shape λ in which the ith row contains the variables in $\mathbf{x}^{\alpha_i}$, in increasing order. We will say such a tableau has **content $\boldsymbol{\mu}$**; i.e., its entries consist of μ_1 1's, μ_2 2's, etc.

For example, suppose that $\mu = (3,2,2,1,1)$ and $\lambda = (4,2,2,1)$. One λ-factorization of $\mathbf{x}^\mu$ and its associated tableau are

$$x_1^3 x_2^2 x_3^2 x_4^1 x_5^1 \;=\; (x_1 x_2 x_3 x_5)(x_1 x_3)(x_2 x_4)(x_1), \qquad T = \begin{array}{|c|c|c|c|} \hline 1 & 2 & 3 & 5 \\ \hline 1 & 3 \\ \cline{1-2} 2 & 4 \\ \cline{1-2} 1 \\ \cline{1-1} \end{array}.$$

Thus the entries of T correspond to variables, and its rows correspond to factors. Observe that all the 1's in T must be in the first column; all the 2's must be in the first or second column; etc. Thus, for every j, there must be collectively enough boxes in the first j columns of T to hold all the entries of T corresponding to the variables $x_1, \dots, x_j$. That is,

$$\forall j: \quad \tilde{\lambda}_1 + \cdots + \tilde{\lambda}_j \geq \mu_1 + \cdots + \mu_j \tag{9.1}$$

which is precisely the condition $\tilde{\lambda} \trianglerighteq \mu$. If this fails, then no λ-factorization of $\mathbf{x}^\mu$ can exist and $b_{\lambda,\mu} = 0$.

If $\tilde{\lambda} = \mu$, then every inequality in (9.1) is in fact an equality, which says that every entry in the jth column is in fact j. That is, there is exactly one λ-factorization of $\mathbf{x}^\mu$, and $b_{\lambda,\mu} = 1$.

Therefore, if we order partitions of n by any linear extension of dominance (such as lex order), then the matrix $[b_{\lambda,\mu}]$ will be upper unitriangular, hence invertible over any integral domain R. (This is the same argument as in Corollary 8.10.6.) It follows that the R-module spanned by the e_λ's is the same as that spanned by the m_μ's for any R, so $\{e_\lambda\}$ is an integral basis. □

Corollary 9.4.3 ("Fundamental Theorem of Symmetric Functions")**.** *The elementary symmetric functions $e_1, e_2, \dots$ are algebraically independent. Therefore, $\Lambda = R[e_1, e_2, \dots]$ as rings.*

Proof. Given any nontrivial polynomial relation among the e_i's, extracting the homogeneous pieces would give a nontrivial linear relation among the e_λ's, which does not exist. □

Corollary 9.4.4. *The transition matrix between the bases $\{e_\lambda\}$ and $\{m_\mu\}$ is symmetric; that is, $b_{\lambda,\mu} = b_{\mu,\lambda}$.*

Proof. We know from the proof of Theorem 9.4.2 that $b_{\lambda,\mu} = [m_\mu] e_\lambda$ is the number of λ-factorizations of $\mathbf{x}^\mu$ into squarefree monomials $\mathbf{x}^{\alpha_1}, \dots, \mathbf{x}^{\alpha_\ell}$. Moreover, these factorizations are in bijection with row-increasing tableaux T of shape λ and content μ: Specifically, $\mathbf{x}^{\alpha_j}$ is the product of the variables x_i such that i appears in the jth row of T. On the other hand, such tableaux are also in bijection with μ-factorizations of $\mathbf{x}^\lambda$ into squarefree monomials $\mathbf{x}^{\beta_1}, \dots, \mathbf{x}^{\beta_k}$, where $\mathbf{x}^{\beta_j}$ is the product of the variables x_i such that j appears in the ith row of T. The conclusion follows. □

For example, if $\mu = (3,2,2,1,1)$ and $\lambda = (4,2,2,1)$, then the tableau

$$T = \begin{array}{|c|c|c|c|}\hline 1 & 2 & 3 & 5 \\ \hline 1 & 3 \\ \cline{1-2} 2 & 4 \\ \cline{1-2} 1 \\ \cline{1-1}\end{array}$$

(which has shape λ and content μ) corresponds to the λ-factorization of x^μ given by

$$x_1^3 x_2^2 x_3^2 x_4^1 x_5^1 \;=\; (x_1x_2x_3x_5)(x_1x_3)(x_2x_4)(x_1)$$

and to the μ-factorization of x^λ given by

$$x_1^4 x_2^2 x_3^2 x_4^1 \;=\; (x_1x_2x_4)(x_1x_3)(x_1x_2)(x_3)(x_1).$$

The 1st factor is $x_1x_2x_4$ because the 1s in T occur in the 1st, 2nd and 4th rows; the 2nd factor is x_1x_3 because the 2s in T occur in the 1st, and 3rd rows; etc.

9.5 Complete homogeneous symmetric functions

Definition 9.5.1. The ***k*th complete homogeneous symmetric function h_k** is the sum of all monomials of degree k, extended multiplicatively to partitions:

$$\begin{aligned} h_0 &= 1, \\ h_k &= \sum_{0<i_1\le i_2\le\cdots\le i_k} x_{i_1}x_{i_2}\cdots x_{i_k} = \sum_{\lambda\vdash k} m_\lambda \qquad \text{for } k>0, \\ h_\lambda &= h_{\lambda_1}\cdots h_{\lambda_\ell} \qquad\qquad \text{for } \lambda=(\lambda_1,\ldots,\lambda_\ell)\in \mathsf{Par}. \end{aligned}$$

In degrees 2 and 3, the h_λ's can be expressed in the e- and m-bases as

$$\begin{bmatrix} h_{11} \\ h_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} e_{11} \\ e_2 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} m_{11} \\ m_2 \end{bmatrix}, \qquad \begin{bmatrix} h_{111} \\ h_{21} \\ h_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & -1 & 0 \\ 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} e_{111} \\ e_{21} \\ e_3 \end{bmatrix} = \begin{bmatrix} 6 & 3 & 1 \\ 3 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} m_{111} \\ m_{21} \\ m_3 \end{bmatrix},$$

The coefficient matrices above are all $\mathbb{Z}$-invertible, witnessing the fact that the h_λ's are also a free R-module basis for Λ and that $\Lambda = R[h_1, h_2, \ldots,]$ as a ring. We could figure out the transition matrices between the h_λ and e_λ, but instead will take a different approach that exploits the close relation between the two families. Consider the generating functions

$$E(t) \;=\; \sum_{k\ge 0} t^k e_k, \qquad H(t) \;=\; \sum_{k\ge 0} t^k h_k.$$

We regard $E(t)$ and $H(t)$ as formal power series in t whose coefficients are themselves formal power series in $\{x_i\}$. Observe that

$$E(t) \;=\; \sum_{k\ge 0} t^k e_k \;=\; \prod_{i\ge 1}(1+tx_i), \qquad H(t) \;=\; \sum_{k\ge 0} t^k h_k \;=\; \prod_{i\ge 1} \frac{1}{1-tx_i}. \tag{9.2}$$

Indeed, the coefficient of t^k in $E(t)$ is the sum of products of k different variables x_i, which is exactly e_k,. Similarly, in the formula for $H(t)$, each factor in the infinite product is a geometric series $1+tx_i+t^2x_i^2+\cdots$, so $[t^k]H(t)$ is the sum of all monomials of degree k. Now (9.2) implies that

$$H(t)E(-t) = \sum_{n=0}^{\infty}\sum_{k=0}^{n}(-1)^k e_k h_{n-k} t^k = 1$$

and extracting the coefficients of positive powers of t gives the **Jacobi-Trudi relations**: for every $n \geq 0$,

$$\sum_{k=0}^{n}(-1)^k e_k h_{n-k} = 0 \qquad \forall n > 0. \tag{9.3}$$

That is,

$$h_1 - e_1 = 0, \quad h_2 - e_1 h_1 + e_2 = 0, \quad h_3 - e_1 h_2 + e_2 h_1 - e_3 = 0, \quad \ldots$$

(where we have plugged in $h_0 = e_0 = 1$). The Jacobi-Trudi relations can be used iteratively to solve for the e_k in terms of the h_k:

$$\begin{aligned}
e_1 &= h_1 &&= h_1,\\
e_2 &= e_1 h_1 - h_2 &&= h_1^2 - h_2,\\
e_3 &= e_2 h_1 - e_1 h_2 + h_3 = h_1(h_1^2 - h_2) - h_2 h_1 + h_3 &&= h_1^3 - 2h_1 h_2 + h_3,\\
e_4 &= e_3 h_1 - e_2 h_2 + e_1 h_3 - h_4 &&= h_1^4 - 3h_1^2 h_2 + h_2^2 + 2h_1 h_3 - h_4,
\end{aligned} \tag{9.4}$$

etc. On the other hand, the Jacobi-Trudi relations are symmetric in the letters h and e, so the equations (9.4) hold if e and h are swapped. Therefore, the elementary and homogenous functions generate the same ring.

Corollary 9.5.2. *$\{h_\lambda : \lambda \vdash n\}$ is an integral basis for Λ_n. Moreover, $\Lambda_R \cong R[h_1, h_2, \ldots]$ as rings.*

Here is another way to see that the h's are an integral basis, which again exploits the symmetry of the Jacobi-Trudi relations. Define a ring endomorphism $\omega : \Lambda \to \Lambda$ by

$$\omega(e_i) = h_i \tag{9.5}$$

for all i, extended multiplicatively (so that $\omega(e_\lambda) = h_\lambda$) and linearly to all symmetric functions. This map, sometimes known as the *Hall transformation*[4] but more usually just referred to as ω, is well-defined since the elementary symmetric functions are algebraically independent. Now Corollary 9.5.2 follows from the following result:

Proposition 9.5.3. *ω is an involution: $\omega(\omega(f)) = f$ for all $f \in \Lambda$. In particular, ω is a ring automorphism.*

Proof. Applying ω to the Jacobi-Trudi relations (9.3), we see that for every $n \geq 1$,

$$\begin{aligned}
0 = \sum_{k=0}^{n}(-1)^{n-k}\omega(e_k)\omega(h_{n-k}) &= \sum_{k=0}^{n}(-1)^{n-k} h_k \omega(h_{n-k})\\
&= \sum_{k=0}^{n}(-1)^{k} h_{n-k}\omega(h_k) \qquad \text{(by replacing } k \text{ with } n-k)\\
&= (-1)^n \sum_{k=0}^{n}(-1)^{n-k} h_{n-k}\omega(h_k)
\end{aligned}$$

and comparing this last expression with the original Jacobi-Trudi relations gives $\omega(h_k) = e_k$ (e.g., because solving for $\omega(h_k)$ in terms of the h_k's gives exactly (9.4), with the e_k's replaced by $\omega(h_k)$'s). □

The involution ω will be extremely useful in its own right.

[4] Do not Google the phrase "Hall transformation." You have been warned.

9.6 Power-sum symmetric functions

Definition 9.6.1. The ***k*th power-sum symmetric function p_k** is the sum of the kth powers of all variables, extended multiplicatively to partitions:

$$p_k = m_k = \sum_{i=1}^{\infty} x_i^k,$$
$$p_\lambda = p_{\lambda_1} \cdots p_{\lambda_\ell} \qquad \text{for } \lambda = (\lambda_1, \ldots, \lambda_\ell) \in \mathsf{Par}.$$

For example, in degrees 2 and 3, you can work out that

$$\begin{bmatrix} p_{11} \\ p_2 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} m_{11} \\ m_2 \end{bmatrix} \qquad \text{and} \qquad \begin{bmatrix} p_{111} \\ p_{21} \\ p_3 \end{bmatrix} = \begin{bmatrix} 6 & 3 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} m_{111} \\ m_{21} \\ m_3 \end{bmatrix}.$$

Note that the transition matrices are invertible over $\mathbb{Q}$, but not over $\mathbb{Z}$: for example, $m_{11} = (p_{11} - p_2)/2$. Thus the power-sums are *not* an integral basis of symmetric functions (although, as we will shortly prove, they are a vector space basis for $\Lambda_{\mathbb{Q}}$).

We have seen this transition matrix before: its columns are characters of tabloid representations! (See (8.23).) This is the first explicit connection we can observe between representations of $\mathfrak{S}_n$ and symmetric functions, and it is the tip of an iceberg. It is actually not hard to prove.

In order to make this self-contained (if you haven't read the chapter on representations of $\mathfrak{S}_n$), I will repeat some definitions from §8.10. For a partition $\mu \vdash n$, a **tabloid** of shape μ is an equivalence class of tableaux under row-equivalence, or equivalently a set composition of $[n]$ into blocks whose sizes are the parts of μ. We represent tabloids as tableaux without vertical lines. If λ is another partition of n, then $\tau_\mu(C_\lambda)$ is defined as follows: if w is any permutation of cycle-type λ, then $\tau_\mu(C_\lambda)$ is the number of tabloids T of shape μ in which every cycle of λ is contained in a row of T. It should be easy to see that this number is independent of the choice of w. For example, if $\mu = (3,2)$ and $\lambda = (2,1,1,1)$, then $\tau_\mu(C_\lambda) = 4$: taking $w = (1\ 2)(3)(4)(5)$, the four tabloids are as follows.

1	2	3		1	2	4		1	2	5		3	4	5	
4	5			3	5			3	4			1	2		

Theorem 9.6.2. *For $\lambda \vdash n$, we have*

$$p_\lambda = \sum_{\mu \vdash n} \tau_\mu(C_\lambda) m_\mu$$

where $\tau_\mu(C_\lambda)$ means the character of the tabloid representation of shape μ on the conjugacy class C_λ of cycle-type λ, as in §8.10.

Proof. Let $\lambda = (\lambda_1, \ldots, \lambda_\ell)$ and $\mu = (\mu_1, \ldots, \mu_k)$. We adopt the notation (8.19). As in Theorem 9.4.2, let $\mathbf{x}^\mu = \prod_i x_i^{\mu_i}$. We calculate the coefficient

$$\begin{aligned} [\mathbf{x}^\mu] p_\lambda &= [\mathbf{x}^\mu] p_{\lambda_1} \cdots p_{\lambda_\ell} \\ &= \text{number of } \lambda\text{-factorizations of } \mathbf{x}^\mu \text{ as } x_{c_1}^{\lambda_1} \cdots x_{c_\ell}^{\lambda_\ell}. \end{aligned}$$

Here we will represent each such choice by a tabloid T in which the factor $x_{c_i}^{\lambda_i}$ contributes labels L_i to the c_ith row, so that T has shape μ. Thus the rows of T correspond to variables, while the entries correspond

to positions in the factorization. (The correspondence between rows and variables is something we have already seen in Corollary 9.4.4, but here the shape λ isn't easily visible in the tabloids, and the numbers correspond to the positions of individual variables rather than the positions of powers of variables.)

For example, let $\lambda = (2,1,1,1)$ and $\mu = (3,2)$. Since $\lambda_1 = 2$, the cells labeled 1 and 2 have to be in the same row, say row c_1, and they encode the monomial $x_{c_1}^2$ chosen from the first factor, which is p_2. The next part is $\lambda_2 = 1$, so the row c_2 containing the single cell labeled 3 tells us which monomial x_{c_2} to choose from the second factor, which is p_1. And so on.

In fact $[\mathbf{x}^\mu]p_\lambda = [x_1^3 x_2^2]p_\lambda = 4$. The four λ-factorizations of $\mathbf{x}^\mu$ are shown below with their corresponding tabloids.

$x_1^2 \cdot x_1 \cdot x_2 \cdot x_2$

x_1	1	2	3
x_2	4	5	

$x_1^2 \cdot x_2 \cdot x_1 \cdot x_2$

x_1	1	2	4
x_2	3	5	

$x_1^2 \cdot x_2 \cdot x_2 \cdot x_1$

x_1	1	2	5
x_2	3	4	

$x_2^2 \cdot x_1 \cdot x_1 \cdot x_1$

x_1	3	4	5
x_2	1	2	

These are precisely the tabloids in which each interval L_i is contained in a single row, and these are precisely those fixed by the permutation given in cycle notation as

$$w = (1 \ \cdots \ \lambda_1)(\lambda_1 + 1 \ \cdots \ \lambda_1 + \lambda_2) \ \cdots \ (\lambda_{[i-1]} + 1 \ \cdots \ \lambda_{[i]}) \ \cdots \ (n - \lambda_\ell + 1 \ \cdots \ n),$$

whose cycle-type is λ. (Compare equation (8.20). In the example above, w is the transposition $(1\ 2)$.) In particular, the number of such tabloids is by definition $\tau_\mu(C_\lambda)$. □

Corollary 9.6.3. *$\{p_\lambda\}$ is a basis for the symmetric functions (although not an integral basis).*

Proof. By Proposition 8.10.5, the transition matrix $[\tau_\mu(C_\lambda)]$ from the monomial basis to the power-sum basis is triangular with respect to dominance order, hence invertible (although not unitriangular). □

For those who are reading this chapter before Chapter 8, please note that Proposition 8.10.5 is purely combinatorial and requires no representation theory, so you might want to go look at it now — or better yet, prove it as an exercise; it's not hard.

9.7 Schur functions and skew Schur functions

The definition of Schur symmetric functions is very different from the m's, e's, h's and p's. It is not even clear at first that they are symmetric. But in fact the Schur functions turn out to be essential in the study of symmetric functions and in several ways are the "best" basis for Λ, particularly through the lens of representation theory.

Definition 9.7.1. A **column-strict tableau** T of shape λ, or **λ-CST** for short, is a labeling of the boxes of the Ferrers diagram of λ with integers (not necessarily distinct) that is

- weakly increasing across every row; and
- strictly increasing down every column.

The partition λ is called the **shape** of T, and the set of all column-strict tableaux of shape λ is denoted $\mathsf{CST}(\lambda)$. The **content** of a CST is the sequence $\alpha = (\alpha_1, \alpha_2, \dots)$, where α_i is the number of boxes labelled i,

and the **weight** of T is the monomial $\mathbf{x}^T = \mathbf{x}^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \cdots$ (the same information as the content, but in monomial form). For example:

1	1	3
2	3	

$x_1^2 x_2 x_3^2$

1	1	1
4	8	

$x_1^3 x_4 x_8$

1	2	3
1	4	

Not a CST

The terminology is not entirely standardized; column-strict tableaux are often called "semistandard tableaux" (as in, e.g. [Sta99]).

Definition 9.7.2. The **Schur function** corresponding to a partition λ is

$$s_\lambda = \sum_{T \in \mathsf{CST}(\lambda)} \mathbf{x}^T.$$

This is a well-defined power series since there are clearly only finitely many column-strict tableaux of a given shape and content. On the other hand, it is far from obvious that s_λ is symmetric. Let's see some examples.

Example 9.7.3. Suppose that $\lambda = (n)$ is the partition with one part, so that the corresponding Ferrers diagram has a single row. Each multiset of n positive integers (with repeats allowed) corresponds to exactly one CST, in which the numbers occur left to right in increasing order. Therefore

$$s_{(n)} = h_n = \sum_{\lambda \vdash n} m_\lambda. \tag{9.6}$$

At the other extreme, suppose that $\lambda = (1, 1, \ldots, 1)$ is the partition with n singleton parts, so that the corresponding Ferrers diagram has a single column. To construct a CST of this shape, we need any n distinct labels. Therefore

$$s_{(1,1,\ldots,1)} = e_n = m_{(1,1,\ldots,1)}. \tag{9.7}$$

◀

Example 9.7.4. Let $\lambda = (2, 1)$. We will express s_λ as a sum of the monomial symmetric functions m_3, m_{21}, m_{111}.

First, no tableau of shape λ can have three equal entries, so the coefficient of m_3 is 0.

Second, for weight $x_a x_b x_c$ with $a < b < c$, there are two possibilities, shown below.

a	b
c	

a	c
b	

Therefore, the coefficient of m_{111} is 2.

Third, for every $a \neq b \in \mathbb{N}_{>0}$, there is one tableau of shape λ and weight $x_a^2 x_b$: the one on the left if $a < b$, or the one on the right if $a > b$.

a	b
b	

b	b
a	

Therefore, $s_{(2,1)} = 2m_{111} + m_{21}$. Summarizing,

$$\begin{bmatrix} s_{111} \\ s_{21} \\ s_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} m_{111} \\ m_{21} \\ m_3 \end{bmatrix},$$

from which it follows that $\{s_3, s_{21}, s_{111}\}$ is a $\mathbb{Z}$-basis for Λ_3. ◀

It should be evident at this point that the Schur functions are **quasisymmetric**, i.e., that for every monomial $x_{i_1}^{a_1} \cdots x_{i_k}^{a_k}$ (where $i_1 < \cdots < i_k$), its coefficient in s_λ depends only on the *ordered* sequence $(a_1, \ldots, a_k)$. To see this, observe that if $j_1 < \cdots < j_k$, then replacing i_s with j_s for all s gives a bijection from λ-CSTs with weight $x_{i_1}^{a_1} \cdots x_{i_k}^{a_k}$ to λ-CSTs with weight $x_{j_1}^{a_1} \cdots x_{j_k}^{a_k}$. (Quasisymmetry is a weaker property than symmetry but is extremely important in its own right; see §9.19 for the beginning of the theory.)

In fact, the Schur functions are not just quasisymmetric but symmetric. Here is an elementary proof. It is enough to show that s_λ is invariant under transposing x_i and x_{i+1} for every $i \in \mathbb{N}_{>0}$, since those transpositions generate $\mathfrak{S}_\infty$. Let $T \in \mathsf{CST}(\lambda)$ and consider all the entries equal to i or $i+1$, ignoring columns that contain *both* i and $i+1$. The set of such entries in a single row looks like

i	$\cdots$	i	$i+1$	$\cdots$	$i+1$

Say that there are a instances of i and b instances of $i+1$. Then we can replace this part of the tableau with b instances of i and a instances of $i+1$. Doing this for every row gives a bijection between tableaux of weight $\cdots x_i^p x_{i+1}^q \cdots$ and those of weight $\cdots x_i^q x_{i+1}^p \cdots$, as desired.

Example 9.7.5. Consider the following tableau.

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

We will canonically construct a new tableau of the same shape in which the numbers of 5's and of 6's are switched. First, call a column *irrelevant* if it contains both a 5 and a 6; here the third column is irrelevant. Consider each group of relevant 5's and 6's in the same row as a single block.

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

Now swap the numbers of 5's and 6's in each block. For example, the topmost block has one 5 and three 6's, so change it to three 5's and one 6. (If a block has equal numbers of 5's and 6's, like the one in the second row, then don't do anything to it.)

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

Note that this construction

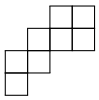
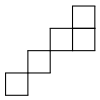
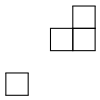
- preserves column-strictness;
- is an involution (because the ignored columns do not change);
- preserves the exponents on all other variables;
- swaps the numbers of k's and $(k+1)$s.

Thus, we have a bijection that says that s_λ is symmetric in each pair of variables $\{x_k, x_{k+1}\}$, hence is symmetric. ◀

An important generalization of a Schur function involves a generalization of the underlying Ferrers diagram of a tableau.

Definition 9.7.6. Let λ, μ be partitions with $\mu \subseteq \lambda$, i.e., $\lambda_i \geq \mu_i$ for all i. There is then an associated **skew partition** or **skew shape** λ/μ, defined via its **skew Ferrers diagram**, in which the ith row has boxes in columns $\mu_i + 1, \ldots, \lambda_i$. A **skew tableau of shape $\boldsymbol{\lambda/\mu}$** is a filling of the skew Ferrers diagram with numbers.

Some skew shapes are shown below; note that disconnected skew shapes are possible.

$4421/21 =$  $4421/321 =$  $4421/322 =$ 

The notion of a column-strict tableau carries over without change to skew shapes. Here is a CST of shape λ/μ, where $\lambda = (8,6,6,5,4,2)$ and $\mu = (5,3,3,3,2)$:

					2	2	3	
			1	1	3			
			2	3	4			
			4	4				
		2	6					
1	1							

The definition of Schur functions (Definition 9.7.1) can also be adapted to skew shapes.

Definition 9.7.7. Let $\mathsf{CST}(\lambda/\mu)$ denote the set of all column-strict skew tableaux of shape λ/μ, and as before weight each tableau $T \in \mathsf{CST}(\lambda/\mu)$ by the monomial $\mathbf{x}^T = \prod_i x_i^{\alpha_i(T)}$, where $\alpha_i(T)$ is the number of i's in T. The **skew Schur function** is then

$$s_{\lambda/\mu} = \sum_{T \in \mathsf{CST}(\lambda/\mu)} \mathbf{x}^T.$$

The elementary proof of symmetry of Schur functions carries over literally to skew Schur functions.

9.8 The Jacobi–Trudi determinant formula

We are next going to establish a formula for the Schur function s_λ as a determinant of a matrix whose entries are h_n's or e_n's (which also proves their symmetry). This takes more work, but the proof, due to the ideas of Lindström, Gessel, and Viennot, is beautiful, and the formula has many other useful consequences (such as what the involution ω does to Schur functions). This exposition follows closely that of [Sag01, §4.5].

Theorem 9.8.1. *[Jacobi–Trudi formula for Schur functions] Let $\lambda = (\lambda_1, \ldots, \lambda_\ell)$ be a partition. Then*

$$s_\lambda = \det \left[h_{\lambda_i - i + j} \right]_{i,j=1,\ldots,\ell} \tag{9.8}$$

and

$$s_{\tilde{\lambda}} = \det \left[e_{\lambda_i - i + j} \right]_{i,j=1,\ldots,\ell}. \tag{9.9}$$

In particular, the Schur functions are symmetric.

For example,

$$s_{311} = \begin{vmatrix} h_3 & h_4 & h_5 \\ h_0 & h_1 & h_2 \\ h_{-1} & h_0 & h_1 \end{vmatrix} = \begin{vmatrix} h_3 & h_4 & h_5 \\ 1 & h_1 & h_2 \\ 0 & 1 & h_1 \end{vmatrix} = h_{311} + h_5 - h_{41} - h_{32}.$$

Proof. We prove (9.8) in detail, and then discuss how the proof can be modified to prove (9.9).

Step 1: For $n \in \mathbb{N}$, express h_n as a generating function for lattice paths.

We will consider lattice paths P that start at some point on the x-axis in $\mathbb{Z}^2$ and move north or east one unit at a time. For every path that we consider, the number of eastward steps must be finite, but the number of northward steps is infinite. Thus the "ending point" is (x, ∞) for some $x \in \mathbb{N}$. Label each eastward step e of P by its y-coordinate $y(e)$. The **weight** of P is the monomial $x^P = \prod_e x_{y(e)}$. An example is shown in Figure 9.1.

The monomial x^P determines the path P up to horizontal shifting, and x^P can be any monomial. Thus we have a bijection, and it follows that for any $a \in \mathbb{N}$,

$$h_n = \sum_{\substack{\text{paths } P \text{ from} \\ (a,0) \text{ to } (a+n,\infty)}} x^P = \sum_{\substack{\text{paths } P \text{ with fixed starting} \\ \text{point with } n \text{ east steps}}} x^P. \tag{9.10}$$

Step 2: Express the generating function for families of lattice paths in terms of the h_k's.

For a partition λ of length ℓ, a **λ-path family $\mathbf{P} = (\pi, P_1, \ldots, P_\ell)$** consists of the following data:

- A permutation $\pi \in \mathfrak{S}_\ell$;
- Two sets of points $U = \{u_1, \ldots, u_\ell\}$ and $V = \{v_1, \ldots, v_\ell\}$, defined by

$$u_i = (\ell - i, 1), \qquad v_i = (\lambda_i + \ell - i, \infty);$$

- A list of lattice paths $P_1, \ldots, P_\ell$, where P_i is a path from $u_{\pi(i)}$ to v_i.

Figure 9.2 shows a λ-path family with $\lambda = (3, 3, 2, 1)$ and $\pi = 3124$. (In general the paths in a family are allowed to share edges, although that is not the case in this example.)

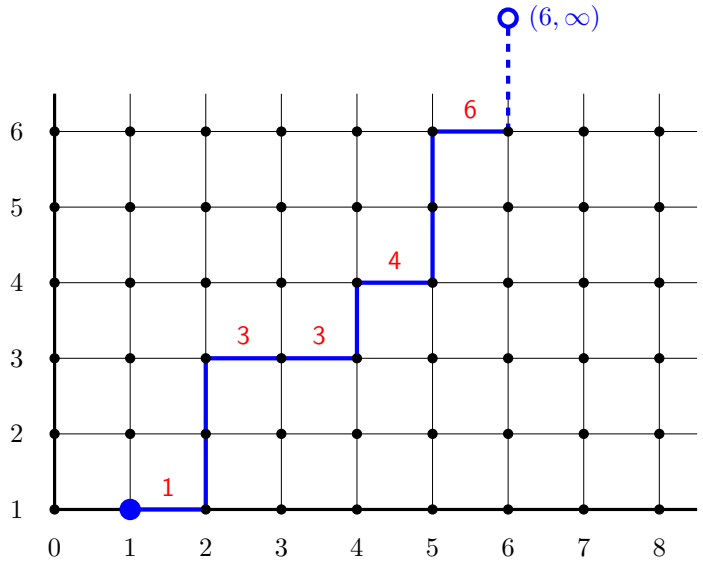


Figure 9.1: A lattice path P from $(1,1)$ to $(6,\infty)$ with weight $x^P = x_1x_3^2x_4x_6$.

Note that for each $i \in [\ell]$, the number of east steps in the path P_i from $u_{\pi(i)}$ to v_i is

$$(\lambda_i + \ell - i) - (\ell - \pi(i)) = \lambda_i - i + \pi(i).$$

Now the first miracle occurs: the signed generating function for path families is the determinant of a matrix whose entries are complete homogeneous symmetric functions! One key observation is that any collection of paths $P_1, \ldots, P_\ell$ in which P_i contains $\lambda_i - i + \pi(i)$ east steps gives rise to a λ-path family $(\pi, P_1, \ldots, P_\ell)$. In other words, if we know what π is, then P_i can be any path with the appropriate number of east steps.

For a path family $\mathbf{P} = (\pi, P_1, \ldots, P_\ell)$, let $x^{\mathbf{P}} = \prod_{i=1}^{\ell} x^{P_i}$ and $(-1)^{\mathbf{P}}$ be the sign of π. Then:

$$\begin{aligned}
\sum_{\mathbf{P}=(\pi,P_1,\ldots,P_\ell)} (-1)^{\mathbf{P}} x^{\mathbf{P}} &= \sum_{\pi \in \mathfrak{S}_\ell} \varepsilon(\pi) \sum_{\substack{\lambda\text{-path families} \\ \mathbf{P}=(\pi,P_1,\ldots,P_\ell)}} x^{P_1} \cdots x^{P_\ell} \\
&= \sum_{\pi \in \mathfrak{S}_\ell} \varepsilon(\pi) \prod_{i=1}^{\ell} \left(\sum_{\substack{\text{paths } P_i \text{ with} \\ \lambda_i - i + \pi(i) \text{ east steps}}} x^{P_i} \right) && \text{(by the key observation above)} \\
&= \sum_{\pi \in \mathfrak{S}_\ell} \varepsilon(\pi) \prod_{i=1}^{\ell} h_{\lambda_i - i + \pi(i)} && \text{(by (9.10))} \\
&= \det \left[h_{\lambda_i - i + j} \right]_{i,j=1,\ldots,\ell} && \text{(look! it's a determinant!)} \qquad (9.11)
\end{aligned}$$

Step 3: Simplify the generating function by cancellation.

Call a path family **good** if no two of its paths meet in a common vertex, and **bad** otherwise. Note that if $\mathbf{P}$ is good, then π must be the identity permutation, and in particular $(-1)^{\mathbf{P}} = 1$.

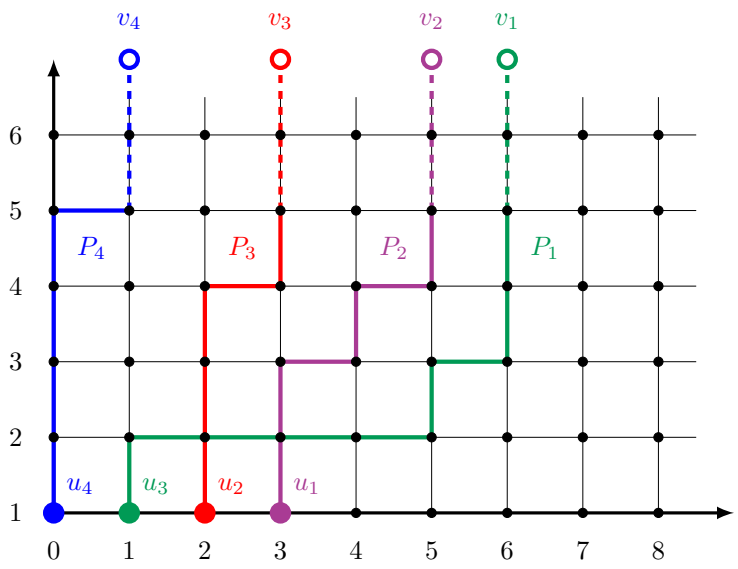


Figure 9.2: A $(3,3,2,1)$-path family (π, U, V) with $\pi = 3124$.

Define a sign-reversing, weight-preserving involution $\mathbf{P} \mapsto \mathbf{P}^\sharp$ on bad λ-path families as follows.

1. Of all the lattice points contained in two or more paths in $\mathbf{P}$, choose the point α with the lex-greatest pair of coordinates.
2. Of all the half-paths from α to some v_i, choose the two with the largest i. Interchange them. Call the resulting path family $\mathbf{P}^\sharp$.

An example is shown in Figure 9.3, where the two half-paths to be interchanged are highlighted in cyan.

Observe that:

- this operation is an involution on bad path families;
- $x^{\mathbf{P}} = x^{\mathbf{P}^\sharp}$; and
- $(-1)^{\mathbf{P}} = -(-1)^{\mathbf{P}^\sharp}$ (because the two are related by a transposition).

Therefore, by the calculation of (9.11), we have

$$|h_{\lambda_i - i + j}|_{i,j=1,\ldots,\ell} = \sum_{\mathbf{P}} (-1)^{\mathbf{P}} x^{\mathbf{P}} = \sum_{\mathbf{P} \text{ good}} x^{\mathbf{P}} \tag{9.12}$$

because all the bad families in the second sum cancel each other out.

Step 4: Enumerate weights of good path families.

For each good path family, label the east steps of each path by height as before. The labels weakly increase as we move north along each path. Moreover, for every j the jth east step of the path P_i occurs one unit east of that of P_{i+1}, so it must also occur strictly south of it (otherwise, the two paths would cross). Therefore, we can construct a column-strict tableau of shape λ by reading off the labels of each path, and this gives a bijection between good λ-path families and column-strict tableaux of shape λ. An example is shown in Figure 9.4.

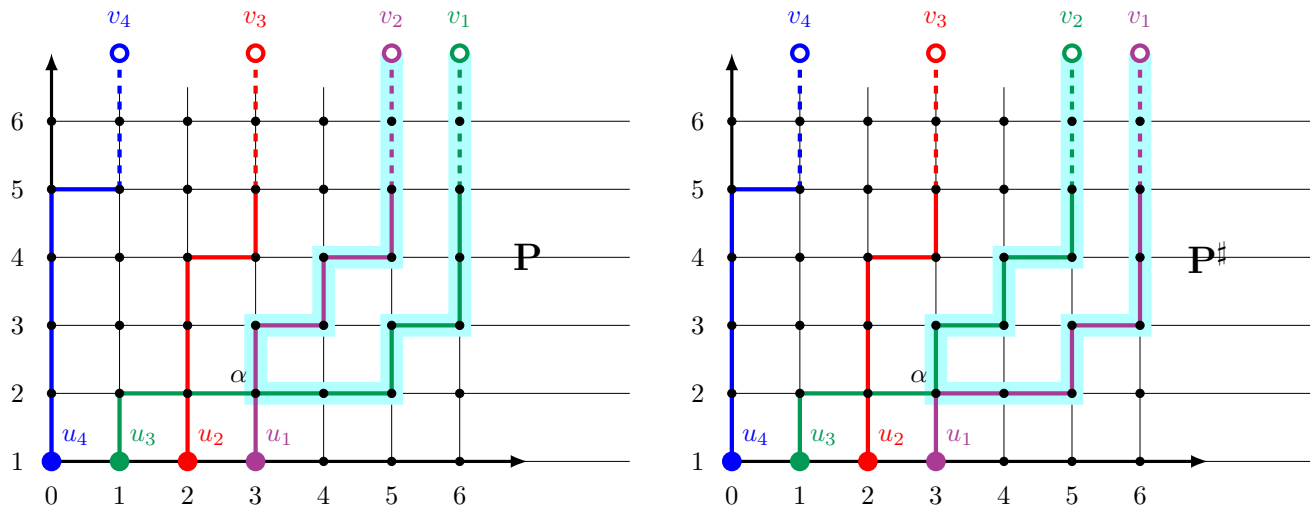


Figure 9.3: The involution $\mathbf{P} \leftrightarrow \mathbf{P}^\sharp$ on bad path families.

Consequently, (9.12) implies that

$$|h_{\lambda_i - i + j}|_{i,j=1,\ldots,\ell} = s_\lambda$$

which is (9.8). Isn't that amazing?

The proof of (9.9) is similar. The key difference is that instead of labeling each east step with its height, we number all the steps (north and east) consecutively, ignoring the first $i-1$ steps of P_i (those below the line $y = x + \ell - 1$, which must all be northward anyway). The weight of a path is still the the product of the variables corresponding to its east steps. This provides a bijection between lattice paths with k east steps and *squarefree* monomials of degree k, giving an analogue of (9.10), with h_n replaced by e_n. Bad path families cancel out by the same involution as before, and each good path family now gives rise to a tableau of shape λ in which rows strictly increase but columns weakly increase (see Figure 9.5). Transposing gives a column-strict tableau of shape $\tilde{\lambda}$, and (9.9) follows. □

Corollary 9.8.2. *For every partition λ, the involution ω interchanges s_λ and $s_{\tilde{\lambda}}$.*

Proof. We know that ω interchanges h_λ and e_λ, so it interchanges the RHS's, hence the LHS's, of (9.8) and (9.9). □

The next step is to prove that the Schur functions are a basis for the symmetric functions. Now that we know they are symmetric, they can be expressed in the monomial basis as

$$s_\lambda = \sum_{\mu \vdash n} K_{\lambda,\mu} m_\mu. \tag{9.13}$$

Thus $K_{\lambda,\mu}$ is the number of column-strict tableaux T with shape λ and content μ. These are called the **Kostka numbers**.

Theorem 9.8.3. *The Schur functions $\{s_\lambda : \lambda \vdash n\}$ are a $\mathbb{Z}$-basis for $\Lambda_\mathbb{Z}$.*

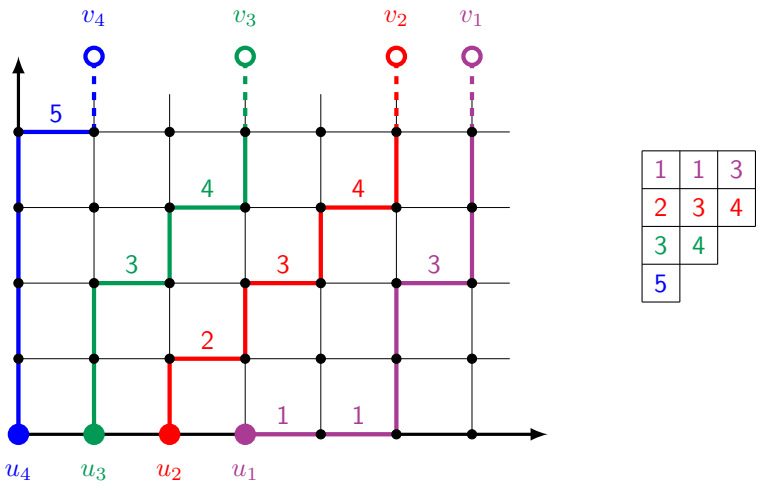


Figure 9.4: The bijection between good path families and column-strict tableaux.

Proof. Here comes one of those triangularity arguments. Consider the matrix of Kostka numbers $[K_{\lambda,\mu}]_{\lambda,\mu\vdash n}$. First, if $\lambda = \mu$, then there is exactly one possibility for T: fill the ith row full of i's. Therefore

$$\forall \lambda \vdash n : \quad K_{\lambda,\lambda} = 1. \tag{9.14}$$

Second, observe that if T is a CST of shape λ and content μ (so in particular $K_{\lambda,\mu} > 0$), then

- every 1 in T must appear in the 1st row;
- every 2 in T must appear in the 1st or 2nd row;
- ...
- every k in T must appear in one of the first k rows;
- ...

and therefore $\sum_{i=1}^{k} \mu_i \leq \sum_{i=1}^{k} \lambda_i$ for all k, which is just the statement that $\mu \trianglelefteq \lambda$ (see Definition 8.10.4). So the matrix $[K_{\lambda,\mu}]_{\lambda,\mu\vdash n}$ is unitriangular, hence $\mathbb{Z}$-invertible. Therefore the Schur functions are a vector space basis for $\Lambda_{\mathbb{Q}}$ and a free module basis for $\Lambda_{\mathbb{Z}}$, just as in the proof of Theorem 9.4.2. □

The lattice-path proof of Theorem 9.8.1 generalizes to skew shapes (although I haven't yet figured out exactly how) to give Jacobi-Trudi determinant formulas for skew Schur functions:

$$s_{\lambda/\mu} = \det\left[h_{\lambda_i-\mu_i-i+j}\right]_{i,j=1,\ldots,\ell}, \qquad s_{\tilde{\lambda}/\tilde{\mu}} = \det\left[e_{\lambda_i-\mu_i-i+j}\right]_{i,j=1,\ldots,\ell}. \tag{9.15}$$

9.9 The Cauchy kernel and the Hall inner product

The next step in studying the ring of symmetric functions Λ is to define an inner product on it. For this we will need the **Cauchy kernel** and the **dual Cauchy kernel**, which are formal power series in two sets of variables $\mathbf{x} = \{x_1, x_2, \ldots\}$, $\mathbf{y} = \{y_1, y_2, \ldots\}$, defined as the following infinite products:

$$\Omega = \prod_{i,j\geq 1} (1 - x_i y_j)^{-1}, \qquad \Omega^* = \prod_{i,j\geq 1} (1 + x_i y_j).$$

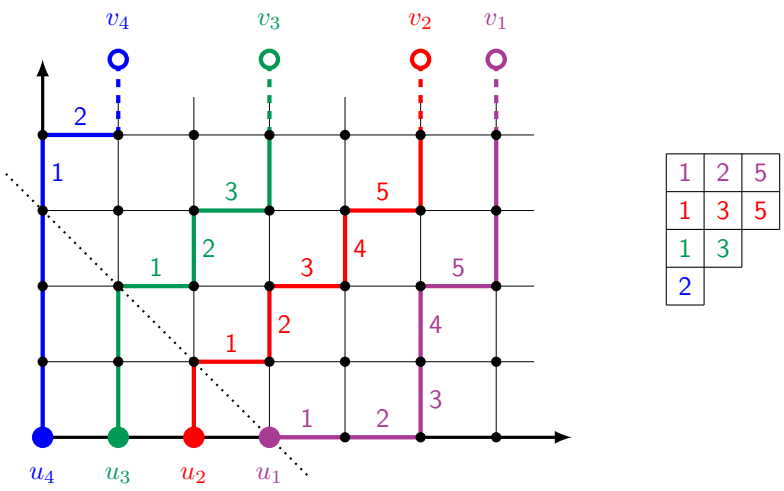


Figure 9.5: The dual bijection between good path families and row-strict tableaux.

The power series Ω and Ω^* are well-defined because the coefficient of any monomial $\mathbf{x}^\alpha \mathbf{y}^\beta$ is the number of ways of factoring it into monomials of the form $x_i y_j$, which is clearly finite (in particular it is zero if $|\alpha| \neq |\beta|$). Moreover, Ω and Ω^* are evidently **bisymmetric**[5], i.e., symmetric with respect to each of the variable sets $\mathbf{x} = \{x_1, x_2, \dots\}$ and $\mathbf{y} = \{y_1, y_2, \dots\}$. Thus we can write Ω and Ω^* as power series in some basis for $\Lambda(\mathbf{x})$ and ask which elements of $\Lambda(\mathbf{y})$ show up as coefficients.

For a partition $\lambda \vdash n$, let[6] $r_i = r_i(\lambda)$ be the number of i's in λ. Define scalars

$$\begin{aligned} z_\lambda &= 1^{r_1} 2^{r_2} \cdots n^{r_n}\, r_1!\, r_2! \cdots r_n!, \\ \varepsilon_\lambda &= (-1)^{n-\ell(\lambda)} = (-1)^{r_2 + r_4 + r_6 + \cdots}. \end{aligned} \tag{9.16}$$

(For the last equality, note that $n - \ell(\lambda) = (\sum_i i r_i) - (\sum_i r_i) = \sum_i (i-1) r_i$, which is congruent to $r_2 + r_4 + \cdots$ modulo 2.)

For example, if $\lambda = (3,3,2,1,1,1)$ then $z_\lambda = (1^3 3!)(2^1 1!)(3^2 2!) = 216$ and $\varepsilon_\lambda = -1$.

Proposition 9.9.1. *Let $\lambda \vdash n$ and let C_λ be the corresponding conjugacy class in $\mathfrak{S}_n$. Then $|C_\lambda| = n!/z_\lambda$, and ε_λ is the sign of each permutation in C_λ.*

We omit the proof, which is straightforward.

Proposition 9.9.2. *We have*

$$\Omega = \sum_\lambda h_\lambda(\mathbf{x}) m_\lambda(\mathbf{y}) = \sum_\lambda \frac{p_\lambda(\mathbf{x}) p_\lambda(\mathbf{y})}{z_\lambda} \quad \text{and} \tag{9.17}$$

$$\Omega^* = \sum_\lambda e_\lambda(\mathbf{x}) m_\lambda(\mathbf{y}) = \sum_\lambda \varepsilon_\lambda \frac{p_\lambda(\mathbf{x}) p_\lambda(\mathbf{y})}{z_\lambda} \tag{9.18}$$

[5] Technically, Ω lives not in the ring $\Lambda(\mathbf{x}, \mathbf{y})$ of bisymmetric power series, but rather its completion, since it contains terms of arbitrarily high degree. If you don't know what "completion" means then don't worry about it. The key point is that Ω is still determined by the coefficients of the bisymmetric series $u_\lambda(\mathbf{x}) v_\mu(\mathbf{y})$ for any bases $\{u_\lambda\}$, $\{v_\mu\}$ of Λ — it is just no longer true that all but finitely many of these coefficients are zero.

[6] In [Sta99], Stanley uses m_i, presumably as a mnemonic for "multiplicity," where I use r_i. I have changed the notation in order to avoid conflict with the notation for monomial symmetric functions.

where the sums run over all partitions λ.

Proof. Recall from (9.2) that $\prod_{i\geq 1}(1-x_it)^{-1} = \sum_{k\geq 0} h_k(\mathbf{x})t^k$. Therefore

$$\begin{aligned}
\prod_{i,j\geq 1}(1-x_iy_j)^{-1} &= \prod_{j\geq 1}\left(\prod_{i\geq 1}(1-x_it)^{-1}\,\Big|_{t=y_j}\right) \\
&= \prod_{j\geq 1}\sum_{k\geq 0} h_k(\mathbf{x})\; y_j^k \\
&= \sum_{\alpha=(\alpha_1,\alpha_2,\dots):\ \sum_i \alpha_i<\infty} \mathbf{y}^\alpha \prod_{i=1}^{\infty} h_{\alpha_i}(\mathbf{x}) \\
&= \sum_{\alpha} \mathbf{y}^\alpha h_{\lambda(\alpha)}(\mathbf{x})
\end{aligned}$$

(where $\lambda(\alpha)$ means the partition whose parts are $\alpha_1,\dots,\alpha_\ell$, sorted in weakly decreasing order)

$$= \sum_{\lambda} h_\lambda(\mathbf{x})m_\lambda(\mathbf{y})$$

as desired.

For the second equality in (9.17), recall the standard power series expansions

$$\log(1+q) = \sum_{n\geq 1}(-1)^{n+1}\frac{q^n}{n}, \qquad \log(1-q) = -\sum_{n\geq 1}\frac{q^n}{n}, \qquad \exp(q) = \sum_{n\geq 0}\frac{q^n}{n!}. \tag{9.19}$$

These are formal power series that obey the rules you would expect; for instance, $\log(\prod_i q_i) = \sum_i(\log q_i)$ and $\exp\log(q) = q$. (The proof of the second of these is left to the reader as Problem 9.6.) In particular,

$$\begin{aligned}
\log\Omega = \log\prod_{i,j\geq 1}(1-x_iy_j)^{-1} &= -\sum_{i,j\geq 1}\log(1-x_iy_j) \\
&= \sum_{i,j\geq 1}\sum_{n\geq 1}\frac{x_i^n y_j^n}{n} \qquad \text{(by (9.19))} \\
&= \sum_{n\geq 1}\frac{1}{n}\left(\sum_{i\geq 1}x_i^n\right)\left(\sum_{j\geq 1}y_j^n\right) \\
&= \sum_{n\geq 1}\frac{p_n(\mathbf{x})p_n(\mathbf{y})}{n}
\end{aligned}$$

and now exponentiating both sides and applying the power series expansion for exp, we get

$$\begin{aligned}
\Omega &= \exp\left(\sum_{n\geq 1}\frac{p_n(\mathbf{x})p_n(\mathbf{y})}{n}\right) = \sum_{k\geq 0}\frac{1}{k!}\left(\sum_{n\geq 1}\frac{p_n(\mathbf{x})p_n(\mathbf{y})}{n}\right)^k \\
&= \sum_{k\geq 0}\frac{1}{k!}\left[\sum_{\lambda:\,\ell(\lambda)=k}\frac{k!}{r_1!\,r_2!\cdots}\left(\frac{p_1(\mathbf{x})p_1(\mathbf{y})}{1}\right)^{r_1}\left(\frac{p_2(\mathbf{x})p_2(\mathbf{y})}{2}\right)^{r_2}\cdots\right]
\end{aligned}$$

(by the multinomial expansion, where r_i means $r_i(\lambda)$)

$$= \sum_\lambda \frac{\prod_{i=1}^{\infty}(p_i(\mathbf{x})p_i(\mathbf{y}))^{r_i}}{\prod_{i=1}^{\infty} i^{r_i} r_i!} = \sum_\lambda \frac{p_\lambda(\mathbf{x})p_\lambda(\mathbf{y})}{z_\lambda}.$$

□

The proofs of the identities for the dual Cauchy kernel are analogous, and are left to the reader as Problem 9.7.

As a first benefit, we can express the homogeneous and elementary symmetric functions in the power-sum basis.

Corollary 9.9.3. *For all n, we have:*

1. $h_n = \sum_{\lambda \vdash n} p_\lambda / z_\lambda$;
2. $e_n = \sum_{\lambda \vdash n} \varepsilon_\lambda p_\lambda / z_\lambda$;
3. $\omega(p_\lambda) = \varepsilon_\lambda p_\lambda$, *where ω is the involution defined in* (9.5).

Proof. (1) Start with the identity of (9.17):

$$\sum_\lambda h_\lambda(\mathbf{x}) m_\lambda(\mathbf{y}) = \sum_\lambda \frac{p_\lambda(\mathbf{x})p_\lambda(\mathbf{y})}{z_\lambda}.$$

Set $y_1 = t$, and $y_k = 0$ for all $k > 1$. This kills all terms on the left side for which λ has more than one part, leaving only those where $\lambda = (n)$ for some n. Meanwhile, on the right side, $p_\lambda(\mathbf{y})$ specializes to $t^{|\lambda|}$, so we get

$$\sum_n h_n(\mathbf{x}) t^n = \sum_\lambda \frac{p_\lambda(\mathbf{x}) t^{|\lambda|}}{z_\lambda}$$

and extracting the coefficient of t^n gives the desired expression for h_n.

(2) Start with (9.18) and do the same thing.

(3) Let ω act on symmetric functions in $\mathbf{x}$ while fixing those in $\mathbf{y}$. Using (9.17) and (9.18), we obtain

$$\sum_\lambda \frac{p_\lambda(\mathbf{x})\, p_\lambda(\mathbf{y})}{z_\lambda} = \sum_\lambda h_\lambda(\mathbf{x}) m_\lambda(\mathbf{y}) = \omega\left(\sum_\lambda e_\lambda(\mathbf{x}) m_\lambda(\mathbf{y})\right) = \omega\left(\sum_\lambda \varepsilon_\lambda \frac{p_\lambda(\mathbf{x})p_\lambda(\mathbf{y})}{z_\lambda}\right)$$
$$= \sum_\lambda \frac{\varepsilon_\lambda \omega(p_\lambda(\mathbf{x}))\, p_\lambda(\mathbf{y})}{z_\lambda}$$

and equating the red coefficients of $p_\lambda(\mathbf{y})/z_\lambda$ yields the desired result. □

Definition 9.9.4. The **Hall inner product** on symmetric functions is defined by declaring $\{h_\lambda\}$ and $\{m_\lambda\}$ to be dual bases. That is, we define

$$\langle h_\lambda, m_\mu \rangle_\Lambda = \delta_{\lambda\mu}$$

and extend by linearity to all of Λ.

Thus the Cauchy kernel can be regarded as a generating function for pairs (h_λ, m_μ), weighted by their inner product. In fact it can be used more generally to compute Hall inner products:

Proposition 9.9.5. *The Hall inner product has the following properties:*

1. *If $\{u_\lambda\}$ and $\{v_\mu\}$ are graded bases for Λ indexed by partitions, such that $\Omega = \sum_\lambda u_\lambda(\mathbf{x})v_\lambda(\mathbf{y})$, then they are dual bases with respect to the Hall inner product; i.e., $\langle u_\lambda, v_\mu\rangle = \delta_{\lambda\mu}$.*
2. *In particular, $\{p_\lambda\}$ and $\{p_\lambda/z_\lambda\}$ are dual bases, and $\{p_\lambda/\sqrt{z_\lambda}\}$ is self-dual, i.e., orthonormal.*
3. *$\langle\cdot,\cdot\rangle$ is a genuine inner product (in the sense of being a* nondegenerate *bilinear form).*
4. *$\langle\cdot,\cdot\rangle$ is positive-definite: $\langle f, f\rangle \geq 0$ for all $f \in \Lambda$, with equality if and only if $f = 0$.*
5. *The involution ω is an isometry with respect to the Hall inner product, i.e.,*

$$\langle a, b\rangle = \langle\omega(a), \omega(b)\rangle.$$

Proof. Assertion (1) is a matter of linear algebra, and is left to the reader (Problem 9.3). Assertion (2) follows from (1) together with (9.18), and (3) from the fact that $\Lambda_\mathbb{R}$ admits an orthonormal basis, namely $\{p_\lambda/z_\lambda\}$. Expanding an arbitrary symmetric function f in the power-sum basis quickly yields (4). Finally, the quickest proof of (5) uses the power-sum basis: by Corollary 9.9.3(3), we have

$$\langle\omega p_\lambda, \omega p_\mu\rangle = \langle\varepsilon_\lambda p_\lambda, \varepsilon_\mu p_\mu\rangle = \varepsilon_\lambda\varepsilon_\mu\langle p_\lambda, p_\mu\rangle = \delta_{\lambda\mu} = \langle p_\lambda, p_\mu\rangle$$

because $\varepsilon_\lambda \in \{1, -1\}$ for all λ. □

The orthonormal basis $\{p_\lambda/\sqrt{z_\lambda}\}$ is not particularly nice from a combinatorial point of view, because it involves irrational coefficients. It turns out that there is a better orthonormal basis: the Schur functions! By Proposition 9.9.5, it suffices to show that $\Omega = \sum_\lambda s_\lambda(\mathbf{x})s_\lambda(\mathbf{y})$, for which we will need a marvelous bijection.

9.10 The Robinson-Schensted-Knuth correspondence

Recall from Example 1.2.4 that a **standard [Young] tableau of shape $\boldsymbol{\lambda}$** is a filling of the Ferrers diagram of λ with the numbers $1, 2, \ldots, n$ that is increasing left-to-right and top-to-bottom. We write $\mathsf{SYT}(\lambda)$ for the set of all standard tableaux of shape λ, and set $f^\lambda = |\mathsf{SYT}(\lambda)|$ (this symbol f^λ is traditional).

For example, if $\lambda = (3, 3)$, then $f^\lambda = 5$; the members of $\mathsf{SYT}(\lambda)$ are as follows.

1	3	5
2	4	6

1	3	4
2	5	6

1	2	5
3	4	6

1	2	4
3	5	6

1	2	3
4	5	6

The **Robinson-Schensted-Knuth (RSK) correspondence**[7] is a bijection between permutations of length n and pairs of standard tableaux of the same shape $\lambda \vdash n$:

$$\mathfrak{S}_n \xrightarrow{\text{RSK}} \bigcup_{\lambda\vdash n} \mathsf{SYT}(\lambda) \times \mathsf{SYT}(\lambda).$$

The main step in the RSK algorithm is **row-insertion**.

Definition 9.10.1. Let T be a column-strict tableau and let $x \in \mathbb{N}_{>0}$. The **row-insertion** $T \leftarrow x$ is defined as follows:

- If $T = \emptyset$, then $T \leftarrow x = \boxed{x}$.

[7] Different versions of the algorithm are referred to by various subsets of these three names; I am not drawing that distinction.

- If $x \geq u$ for all entries u in the top row of T, then append x to the end of the top row.
- Otherwise, find the leftmost entry u such that $x < u$. Replace u with x, and then insert u into the subtableau consisting of the second and succeeding rows. In this case we say that x **bumps** u.
- Repeat until the bumping stops.

Got that? Now, for $w = w_1 w_2 \cdots w_n \in \mathfrak{S}_n$, let P be the tableau $((\emptyset \leftarrow w_1) \leftarrow w_2) \leftarrow \cdots \leftarrow w_n \in \mathfrak{S}_n$. Let Q be the standard tableau of the same shape as P that records where a new box appears in the underlying Ferrers diagram at each step of the algorithm. The tableaux P and Q are respectively called the **insertion tableau** and the **recording tableau**, and the map $w \mapsto (P, Q)$ is the **RSK correspondence**.

Example 9.10.2. Let $w = 57214836 \in \mathfrak{S}_8$. Start with a pair (P, Q) of empty tableaux.

Step 1: Row-insert $w_1 = 5$ into P. We do this in the obvious way. Since it is the first cell added, we add a cell containing 1 to Q.

$$P = \begin{array}{|c|}\hline 5 \\ \hline \end{array} \qquad Q = \begin{array}{|c|}\hline 1 \\ \hline \end{array} \tag{9.20a}$$

Step 2: Row-insert $w_2 = 7$ into P. Since $5 < 7$, we can do this by appending the new cell to the top row, and adding a cell labeled 2 to Q to record where we have put the new cell in P.

$$P = \begin{array}{|c|c|}\hline 5 & 7 \\ \hline \end{array} \qquad Q = \begin{array}{|c|c|}\hline 1 & 2 \\ \hline \end{array} \tag{9.20b}$$

Step 3: Row-insert $w_3 = 2$ into P. This is a bit trickier. We cannot just append a 2 to the first row of P, because the result would not be a standard tableau. The 2 has to go in the top left cell, but that already contains a 5. Therefore, the 2 "bumps" the 5 out of the first row into a new second row. Again, we record the location of the new cell by adding a cell labeled 3 to Q.

$$P = \begin{array}{|c|c|}\hline 2 & 7 \\ \hline 5 \\ \cline{1-1} \end{array} \qquad Q = \begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 \\ \cline{1-1} \end{array} \tag{9.20c}$$

Step 4: Row-insert $w_4 = 1$ into P. This time, the new 1 bumps the 2 out of the first row. The 2 has to go into the second row, but again we cannot simply append it to the right. Instead, the 2 bumps the 5 out of the second row into the (new) third row.

$$P = \begin{array}{|c|c|}\hline 1 & 7 \\ \hline 2 \\ \cline{1-1} 5 \\ \cline{1-1} \end{array} \qquad Q = \begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 \\ \cline{1-1} 4 \\ \cline{1-1} \end{array} \tag{9.20d}$$

Step 5: Row-insert $w_5 = 4$ into P. The 4 bumps the 7 out of the first row. The 7, however, can comfortably fit at the end of the second row, without any more bumping.

$$P = \begin{array}{|c|c|}\hline 1 & 4 \\ \hline 2 & 7 \\ \cline{1-2} 5 \\ \cline{1-1} \end{array} \qquad Q = \begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 & 5 \\ \cline{1-2} 4 \\ \cline{1-1} \end{array} \tag{9.20e}$$

Step 6: Row-insert $w_6 = 8$ into P. The 8 just goes at the end of the first row.

$$P = \begin{array}{|c|c|c|}\hline 1 & 4 & 8 \\ \hline 2 & 7 \\ \cline{1-2} 5 \\ \cline{1-1} \end{array} \qquad Q = \begin{array}{|c|c|c|}\hline 1 & 2 & 6 \\ \hline 3 & 5 \\ \cline{1-2} 4 \\ \cline{1-1} \end{array} \tag{9.20f}$$

Step 7: Row-insert $w_7 = 3$ into P. 3 bumps 4, and then 4 bumps 7.

$$P = \begin{array}{|c|c|c|}\hline 1 & 3 & 8 \\ \hline 2 & 4 \\ \cline{1-2} 5 & 7 \\ \cline{1-2} \end{array} \qquad Q = \begin{array}{|c|c|c|}\hline 1 & 2 & 6 \\ \hline 3 & 5 \\ \cline{1-2} 4 & 7 \\ \cline{1-2} \end{array} \tag{9.20g}$$

Step 8: Row-insert $w_8 = 6$ into P. 6 bumps 8 into the second row.

$$P = \begin{array}{|c|c|c|}\hline 1 & 3 & 6 \\ \hline 2 & 4 & 8 \\ \cline{1-3} 5 & 7 \\ \cline{1-2} \end{array} \qquad Q = \begin{array}{|c|c|c|}\hline 1 & 2 & 6 \\ \hline 3 & 5 & 8 \\ \cline{1-3} 4 & 7 \\ \cline{1-2} \end{array} \tag{9.20h}$$

◀

A crucial feature of the RSK correspondence is that it can be reversed. That is, given a pair (P, Q), we can recover the permutation that gave rise to it.

Example 9.10.3. Suppose that we are given the pair of tableaux in (9.20h). What was the previous step? To get the previous Q, we just delete the 8. As for P, the last cell added must be the one containing 8. This is in the second row, so somebody must have bumped 8 out of the first row. That somebody must be the largest number less than 8, namely 6. So 6 must have been the number inserted at this stage, and the previous pair of tableaux must have been those in (9.20g). ◀

Example 9.10.4. Let P be the standard tableau (with 18 boxes) shown in (a) below. Suppose that we know that the cell labeled 16 was the last one added (because the corresponding cell in Q contains an 18). Then the "bumping path" must be as indicated in the center figure (b). (That is, the 16 was bumped by the 15, which was bumped by the 13, and so on.) Each number in the bumping path is the rightmost one in its row that is less than the next lowest number in the path. The previous tableau in the RSK algorithm can now be found by "unbumping": push every number in the bumping path up and toss out the top one, to obtain the tableau on the right (c).

(a)
$$\begin{array}{|c|c|c|c|c|c|}\hline 1 & 2 & 5 & 8 & 10 & 18 \\ \hline 3 & 4 & 11 & 12 & 19 \\ \cline{1-5} 6 & 7 & 13 \\ \cline{1-3} 9 & 15 & 17 \\ \cline{1-3} 14 & 16 \\ \cline{1-2} \end{array}$$

(b)
$$\begin{array}{|c|c|c|c|c|c|}\hline 1 & 2 & 5 & 8 & 10 & 18 \\ \hline 3 & 4 & 11 & 12 & 19 \\ \cline{1-5} 6 & 7 & 13 \\ \cline{1-3} 9 & 15 & 17 \\ \cline{1-3} 14 & 16 \\ \cline{1-2} \end{array}$$

(c)
$$\begin{array}{|c|c|c|c|c|c|}\hline 1 & 2 & 5 & 8 & 12 & 18 \\ \hline 3 & 4 & 11 & 13 & 19 \\ \cline{1-5} 6 & 7 & 15 \\ \cline{1-3} 9 & 16 & 17 \\ \cline{1-3} 14 \\ \cline{1-1} \end{array}$$

Iterating this procedure allows us to recover w from the pair (P, Q). ◀

We have proved the following fact:

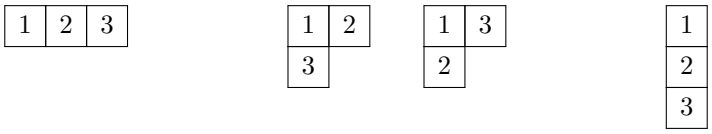
Theorem 9.10.5. *The RSK correspondence is a bijection*

$$\mathfrak{S}_n \xrightarrow{\mathsf{RSK}} \bigcup_{\lambda \vdash n} \mathsf{SYT}(\lambda) \times \mathsf{SYT}(\lambda).$$

Corollary 9.10.6. *For every n we have* $\sum_{\lambda \vdash n} (f^\lambda)^2 = n!$.

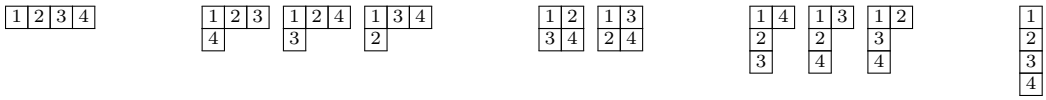
(What does this remind you of?)

Example 9.10.7. The SYT's with $n = 3$ boxes are as follows:



Note that $f^{(3)} = f^{(1,1,1)} = 1$ and $f^{(2,1)} = 2$, and $1^2 + 2^2 + 1^2 = 6 = 3!$. ◀

Example 9.10.8. The SYT's with $n = 4$ boxes are as follows:



So

$$f^{(4)} = 1, \qquad f^{(3,1)} = 3, \qquad f^{(2,2)} = 2, \qquad f^{(2,1,1)} = 3, \qquad f^{(1,1,1,1)} = 1.$$

and the sum of the squares of these numbers is 24. ◀

We have seen these numbers before — they are the dimensions of the irreps of $\mathfrak{S}_3$ and $\mathfrak{S}_4$, as calculated in Examples 8.7.2 and 8.7.3. Hold that thought!

Another neat fact about the RSK correspondence is this:

Proposition 9.10.9. *Let $w \in \mathfrak{S}_n$. If $\mathsf{RSK}(w) = (P, Q)$, then $\mathsf{RSK}(w^{-1}) = (Q, P)$. In particular, the number of involutions in $\mathfrak{S}_n$ is $\sum_{\lambda \vdash n} f^\lambda$.*

The proof is in [Sta99, §7.13]; I hope to understand and write it up some day. It is certainly not obvious from the standard RSK algorithm, where it looks like P and Q play inherently different roles. In fact, they are more symmetric than they look. There are alternative descriptions of RSK from which the symmetry is more apparent, also described in [Sta99, §7.13] and in [Ful97, §4.2]. I describe the former (without proof) in §9.18.

The RSK correspondence can be extended to more general tableaux. This turns out to be the key to expanding the Cauchy kernel in terms of Schur functions.

Definition 9.10.10. A **generalized permutation** of length n is a $2 \times n$ matrix

$$w = \begin{pmatrix} \mathbf{q} \\ \mathbf{p} \end{pmatrix} = \begin{pmatrix} q_1 & q_2 & \cdots & q_n \\ p_1 & p_2 & \cdots & p_n \end{pmatrix} \tag{9.21}$$

where $\mathbf{q} = (q_1, \ldots, q_n), \mathbf{p} = (p_1, \ldots, p_n) \in \mathbb{N}^n_{>0}$, and the $(q_1, p_1), \ldots, (q_n, p_n)$ are in lex order. (That is, $q_1 \leq \cdots \leq q_n$, and if $q_i = q_{i+1}$ then $p_i \leq p_{i+1}$.) The **weight** of w is the monomial $\mathbf{x}^P \mathbf{y}^Q = x_{p_1} \cdots x_{p_n} y_{q_1} \cdots y_{q_n}$. The set of all generalized permutations will be denoted GP, and the set of all generalized permutations of length n will be denoted $\mathsf{GP}(n)$.

If $q_i = i$ for all i and the p_i's are pairwise distinct elements of $[n]$, then $w = \binom{\mathbf{q}}{\mathbf{p}}$ is just an ordinary permutation in $\mathfrak{S}_n$, written in two-line notation.

The **generalized RSK algorithm (gRSK)** is defined in exactly the same way as original RSK, except that the inputs are now allowed to be generalized permutations rather than ordinary permutations. At the ith stage, we row-insert p_i in the insertion tableau P and place q_i in the recording tableau Q in the new cell added.

Example 9.10.11. Consider the generalized permutation

$$w = \binom{\mathbf{q}}{\mathbf{p}} = \begin{pmatrix} 1 & 1 & 2 & 4 & 4 & 4 & 5 & 5 & 5 \\ 2 & 4 & 1 & 1 & 3 & 3 & 2 & 2 & 4 \end{pmatrix} \in \mathsf{GP}(9).$$

The result of the gRSK algorithm is as follows. The unnamed tableau on the right records the order in which the boxes were added.

$$P = \begin{array}{|c|c|c|c|c|}\hline 1 & 1 & 2 & 2 & 4 \\ \hline 2 & 3 & 3 \\ \cline{1-3} 4 \\ \cline{1-1} \end{array} \qquad Q = \begin{array}{|c|c|c|c|c|}\hline 1 & 1 & 4 & 4 & 5 \\ \hline 2 & 4 & 5 \\ \cline{1-3} 5 \\ \cline{1-1} \end{array} \qquad \begin{array}{|c|c|c|c|c|}\hline 1 & 2 & 5 & 6 & 9 \\ \hline 3 & 4 & 8 \\ \cline{1-3} 7 \\ \cline{1-1} \end{array}$$

◀

The tableaux P, Q arising from gRSK will always have the same shape as each other, and will be weakly increasing eastward and strictly increasing southward — that is, they will be *column-strict tableaux*, precisely the things for which the Schur functions are generating functions. Column-strictness of P follows from the definition of insertion. As for Q, it is enough to show that no label k appears more than once in the same column. Indeed, all instances of k in $\mathbf{q}$ occur consecutively (say as $q_i, \ldots, q_j$), and the corresponding entries of $\mathbf{p}$ are weakly increasing, so none of them will bump any other (in fact their bumping paths will not cross), which means that each k appears to the east of all previous k's.

This observation also suffices to show that the generalized permutation w can be recovered from the pair (P, Q): the rightmost instance of the largest entry in Q must have been the last box added. Hence the corresponding box of P can be "unbumped" to recover the previous P and thus the last column of w. Iterating this process allows us to recover w. Therefore, generalized RSK gives a bijection

$$\mathsf{GP}(n) \xrightarrow{\mathsf{RSK}} \bigcup_{\lambda \vdash n} \{(P, Q) \colon P, Q \in \mathsf{CST}(\lambda)\} \tag{9.22}$$

in which a generalized permutation $\binom{\mathbf{q}}{\mathbf{p}}$ maps to a pair of tableaux P, Q with weight monomials $\mathbf{x}^P$ and $\mathbf{y}^Q$.

On the other hand, a generalized permutation $w = \binom{\mathbf{q}}{\mathbf{p}} \in \mathsf{GP}(n)$ is determined by the number a_{ij} of occurrences of every column $\binom{q_i}{p_i}$. Therefore, the generating function for generalized permutations by weights is precisely the Cauchy kernel:

$$\Omega = \prod_{i,j \geq 1} \frac{1}{1 - x_i y_j} = \sum_{(a_{ij})} \prod_{i,j} (x_i y_j)^{a_{ij}} = \sum_{\binom{\mathbf{q}}{\mathbf{p}} \in \mathsf{GP}} \mathbf{x}^{\mathbf{q}} \mathbf{y}^{\mathbf{p}}$$

where the first sum ranges over all doubly-indexed tables of nonnegative integers $(a_{ij})_{i,j \geq 1}$ with finite support (i.e., such that only finitely many of the a_{ij} are nonzero). We now apply gRSK, and the rest is

straightforward:

$$\begin{aligned}\Omega &= \sum_{\lambda} \sum_{P,Q \in \mathsf{CST}(\lambda)} \mathbf{x}^P \mathbf{y}^Q \qquad \text{(by gRSK)} \\ &= \sum_{\lambda} \left(\sum_{P \in \mathsf{CST}(\lambda)} \mathbf{x}^P \right) \left(\sum_{Q \in \mathsf{CST}(\lambda)} \mathbf{y}^Q \right) \\ &= \sum_{\lambda} s_\lambda(\mathbf{x}) s_\lambda(\mathbf{y}).\end{aligned}$$

Combined with assertion (a) of Proposition 9.9.5, we have proven:

Theorem 9.10.12. *The Schur functions form an orthonormal $\mathbb{Z}$-basis for Λ under the Hall inner product.*

Corollary 9.10.13. *For every $\mu \in \mathsf{Par}$ we have*

$$h_\mu = \sum_\lambda K_{\lambda\mu} s_\lambda \qquad \textit{and} \qquad e_\mu = \sum_\lambda K_{\tilde{\lambda}\mu} s_\lambda.$$

Proof. Equation (9.13) implies that

$$\langle s_\lambda,\, h_\mu \rangle = \left\langle \sum_{\mu \vdash n} K_{\lambda,\mu} m_\mu,\, h_\mu \right\rangle = K_{\lambda,\mu} = \left\langle s_\lambda,\, \sum_\lambda K_{\lambda\mu} s_\lambda \right\rangle \tag{9.23}$$

for every λ. So the two blue expressions are equal. Applying ω to both sides (using Prop. 9.9.5(5)) we get

$$\langle s_\lambda,\, e_\mu \rangle = \big\langle \omega(s_\lambda),\, \omega(e_\mu) \big\rangle = \langle s_{\tilde{\lambda}},\, h_\mu \rangle = K_{\tilde{\lambda},\mu} = \left\langle s_\lambda,\, \sum_\lambda K_{\tilde{\lambda}\mu} s_\lambda \right\rangle \tag{9.24}$$

and the two red expressions are equal. □

9.11 The Frobenius characteristic

As in Section 8.6, denote by $C\ell(\mathfrak{S}_n)$ the vector space of $\mathbb{C}$-valued class functions on the symmetric group $\mathfrak{S}_n$; also, let $C\ell(\mathfrak{S}_0) = \mathbb{C}$. Define a graded vector space

$$C\ell(\mathfrak{S}) \;=\; \bigoplus_{n \geq 0} C\ell(\mathfrak{S}_n)$$

We now want to make $C\ell(\mathfrak{S})$ into a graded ring. To start, we declare that the elements of $C\ell(\mathfrak{S}_0)$ behave like scalars. For $n_1, n_2 \in \mathbb{N}_{>0}$ and $f_i \in C\ell(\mathfrak{S}_{n_i})$, we would like to define a product $f_1 f_2 \in C\ell(\mathfrak{S}_n)$, where $n = n_1 + n_2$. First, define a function $f_1 \times f_2 : \mathfrak{S}_{n_1} \times \mathfrak{S}_{n_2} \to \mathbb{C}$ by

$$(f_1 \times f_2)(w_1, w_2) = f_1(w_1) f_2(w_2);$$

this is a class function because the conjugacy classes in $G \times H$ are just the Cartesian products of conjugacy classes in G with those in H (this is a general fact about products of groups). The next step is to lift to $\mathfrak{S}_n$. Identify $\mathfrak{S}_{n_1} \times \mathfrak{S}_{n_2}$ with the Young subgroup $\mathfrak{S}_{n_1,n_2} \subseteq \mathfrak{S}_n$ fixing each of the sets $\{1, 2, \ldots, n_1\}$ and

$\{n_1+1, n_1+2, \ldots, n_1+n_2\}$. (See (8.19).) We now define the product $f_1 f_2 \in C\ell(\mathfrak{S}_n)$ by the formula for induced characters (Proposition 8.9.4):

$$f_1 f_2 \;=\; \mathrm{Ind}_{\mathfrak{S}_{n_1,n_2}}^{\mathfrak{S}_n}(f_1 \times f_2) \;=\; \frac{1}{n_1!\,n_2!} \sum_{\substack{g\in\mathfrak{S}_n:\\ g^{-1}wg\in\mathfrak{S}_{n_1,n_2}}} (f_1\times f_2)(g^{-1}wg).$$

There is no guarantee that $f_1 f_2$ is a character of $\mathfrak{S}_n$ (unless f_1 and f_2 are characters), but at least this operation is a well-defined map on class functions, and it makes $C\ell(\mathfrak{S})$ into a commutative graded $\mathbb{C}$-algebra. (It is pretty clearly bilinear and commutative; it is nontrivial but not hard to check that it is associative.)

For a partition $\lambda \vdash n$, let $1_\lambda \in C\ell(\mathfrak{S}_n)$ be the indicator function on the conjugacy class $C_\lambda \subseteq \mathfrak{S}_n$, and let

$$\mathfrak{S}_\lambda \;=\; \mathfrak{S}_{\{1,\ldots,\lambda_1\}} \times \mathfrak{S}_{\{\lambda_1+1,\ldots,\lambda_1+\lambda_2\}} \times \cdots \times \mathfrak{S}_{\{n-\lambda_\ell+1,\ldots,n\}} \;\subseteq\; \mathfrak{S}_n.$$

For a permutation $w \in \mathfrak{S}_n$, let $\lambda(w)$ denote the cycle-type of w (so $\lambda(w)$ is a partition). Define a function $\psi : \mathfrak{S}_n \to \Lambda_n$ by

$$\psi(w) = p_{\lambda(w)}. \tag{9.25}$$

Note that ψ is a class function (albeit with values in Λ rather than in $\mathbb{C}$).

Definition 9.11.1. The **Frobenius characteristic** is the map

$$\mathbf{ch}:\; C\ell_{\mathbb{C}}(\mathfrak{S}) \to \Lambda_{\mathbb{C}}$$

defined on $f \in C\ell(\mathfrak{S}_n)$ by

$$\mathbf{ch}(f) \;=\; \langle f, \psi\rangle_{\mathfrak{S}_n} \;=\; \frac{1}{n!}\sum_{w\in\mathfrak{S}_n} \overline{f(w)}\, p_{\lambda(w)} \;=\; \sum_{\lambda\vdash n} \overline{f(C_\lambda)}\,\frac{p_\lambda}{z_\lambda}$$

where the last equality follows from Proposition 9.9.1.

Theorem 9.11.2. *The Frobenius characteristic* **ch** *has the following properties:*

1. $\mathbf{ch}(1_\lambda) = p_\lambda/z_\lambda$.
2. **ch** *is an isometry, i.e., it preserves inner products:*
$$\langle f, g\rangle_{\mathfrak{S}_n} \;=\; \langle \mathbf{ch}(f), \mathbf{ch}(g)\rangle_\Lambda.$$
3. **ch** *is a ring isomorphism (in fact, an isomorphism of graded $\mathbb{C}$-algebras).*
4. $\mathbf{ch}(\mathrm{Ind}_{\mathfrak{S}_\lambda}^{\mathfrak{S}_n}\chi_{\mathrm{triv}}) = \mathbf{ch}(\tau_\lambda) = h_\lambda$.
5. $\mathbf{ch}(\mathrm{Ind}_{\mathfrak{S}_\lambda}^{\mathfrak{S}_n}\chi_{\mathrm{sign}}) = e_\lambda$.
6. *Let χ be any character of $\mathfrak{S}_n$ and let χ_{sign} be the sign character on $\mathfrak{S}_n$. Then $\mathbf{ch}(\chi\otimes\chi_{\mathrm{sign}}) = \omega(\mathbf{ch}(\chi))$, where ω is the involution of 9.5.*
7. **ch** *restricts to an isomorphism $C\ell_V(\mathfrak{S}) \to \Lambda_{\mathbb{Z}}$, where $C\ell_V(\mathfrak{S})$ is the $\mathbb{Z}$-module generated by irreducible characters (i.e., the space of* ***virtual characters****).*
8. *The irreducible characters of $\mathfrak{S}_n$ are $\{\mathbf{ch}^{-1}(s_\lambda) : \lambda\vdash n\}$.*

Proof. (1): Immediate from the definition. It follows that **ch** is (at least) a graded $\mathbb{C}$-vector space isomorphism, since $\{1_\lambda :\ \lambda\vdash n\}$ and $\{p_\lambda/z_\lambda :\ \lambda\vdash n\}$ are $\mathbb{C}$-bases for $C\ell(\mathfrak{S}_n)$ and Λ_n respectively.

(2): It suffices to check the identity on a basis of $C\ell(\mathfrak{S})$:

$$\langle 1_\lambda,\, 1_\mu\rangle_{\mathfrak{S}_n} \;=\; \frac{1}{n!}\sum_{w\in\mathfrak{S}_n} \overline{1_\lambda(w)}1_\mu(w) \;=\; \frac{1}{n!}|C_\lambda|\delta_{\lambda\mu} \;=\; \delta_{\lambda\mu}/z_\lambda \;=\; \langle p_\lambda/z_\lambda,\, p_\mu/z_\mu\rangle_\Lambda \;=\; \big\langle \mathbf{ch}(1_\lambda), \mathbf{ch}(1_\mu)\big\rangle_\Lambda$$

where the penultimate equality is (9.17) (from expanding the Cauchy kernel in the power-sum bases).

(3): Let $n = j + k$ and let $f \in C\ell(\mathfrak{S}_{[j]})$ and $g \in C\ell(\mathfrak{S}_{[j+1,n]})$ (so that elements of these two groups commute, and the cycle-type of a product is just the multiset union of the cycle-types). Then:

$$\begin{aligned}
\mathbf{ch}(fg) &= \left\langle \mathrm{Ind}_{\mathfrak{S}_j \times \mathfrak{S}_k}^{\mathfrak{S}_n}(f \times g),\ \psi \right\rangle_{\mathfrak{S}_n} && \text{(where } \psi \text{ is defined as in (9.25))} \\
&= \left\langle f \times g,\ \mathrm{Res}_{\mathfrak{S}_j \times \mathfrak{S}_k}^{\mathfrak{S}_n} \psi \right\rangle_{\mathfrak{S}_j \times \mathfrak{S}_k} && \text{(by Frobenius reciprocity)} \\
&= \frac{1}{j!\,k!} \sum_{(w,x) \in \mathfrak{S}_j \times \mathfrak{S}_k} \overline{(f \times g)(w,x)} \cdot p_{\lambda(wx)} \\
&= \left(\frac{1}{j!} \sum_{w \in \mathfrak{S}_j} \overline{f(w)}\, p_{\lambda(w)} \right) \left(\frac{1}{k!} \sum_{x \in \mathfrak{S}_k} \overline{g(x)}\, p_{\lambda(x)} \right) && \text{(because the power-sum basis is multiplicative)} \\
&= \mathbf{ch}(f)\ \mathbf{ch}(g).
\end{aligned}$$

Thus it is a ring isomorphism. It is a $\mathbb{C}$-algebra isomorphism because it maps the unit to the unit and is graded by definition.

(4), (5): Denote by χ^n_{triv} and χ^n_{sign} the trivial and sign characters on $\mathfrak{S}_n$. We calculate in parallel:

$$\begin{aligned}
\mathbf{ch}(\chi^n_{\mathrm{triv}}) &= \langle \chi^n_{\mathrm{triv}}, \psi \rangle_{\mathfrak{S}_n} & \mathbf{ch}(\chi^n_{\mathrm{sign}}) &= \langle \chi^n_{\mathrm{sign}}, \psi \rangle_{\mathfrak{S}_n} && \text{(by definition of } \mathbf{ch}) \\
&= \frac{1}{n!} \sum_{w \in \mathfrak{S}_n} p_{\lambda(w)} & &= \frac{1}{n!} \sum_{w \in \mathfrak{S}_n} \varepsilon_{\lambda(w)} p_{\lambda(w)} && \text{(by def'n of } \psi \text{ and } \langle \cdot, \cdot \rangle_{\mathfrak{S}_n}) \\
&= \sum_{\lambda \vdash n} \frac{|C_\lambda|}{n!} p_\lambda & &= \sum_{\lambda \vdash n} \varepsilon_\lambda \frac{|C_\lambda|}{n!} p_\lambda \\
&= \sum_{\lambda \vdash n} \frac{p_\lambda}{z_\lambda} & &= \sum_{\lambda \vdash n} \varepsilon_\lambda \frac{p_\lambda}{z_\lambda} \\
&= h_n & &= e_n && \text{(by Corollary 9.9.3).}
\end{aligned}$$

Now

$$h_\lambda = \prod_{i=1}^{\ell} h_{\lambda_i} = \prod_{i=1}^{\ell} \mathbf{ch}(\chi^{\lambda_i}_{\mathrm{triv}}) = \mathbf{ch}\left(\prod_{i=1}^{\ell} \chi^{\lambda_i}_{\mathrm{triv}} \right) = \mathbf{ch}(\mathrm{Ind}_{\mathfrak{S}_\lambda}^{\mathfrak{S}_n} \chi^n_{\mathrm{triv}})$$

(the third equality since **ch** is a ring homomorphism) and likewise $e_\lambda = \mathbf{ch}(\mathrm{Ind}_{\mathfrak{S}_\lambda}^{\mathfrak{S}_n} \chi^n_{\mathrm{sign}})$.

(6): Left as an exercise.

(7), (8): Each of (4) and (5) says that $\mathbf{ch}^{-1}(\Lambda_{\mathbb{Z}})$ is contained in the space of virtual characters, because $\{h_\lambda\}$ and $\{e_\lambda\}$ are $\mathbb{Z}$-module bases for $\Lambda_{\mathbb{Z}}$, and their inverse images under **ch** are genuine characters. On the other hand, $\{s_\lambda\}$ is also a $\mathbb{Z}$-basis, so each $\sigma_\lambda := \mathbf{ch}^{-1}(s_\lambda)$ is a character. Moreover, since **ch** is an isometry we have

$$\langle \sigma_\lambda, \sigma_\mu \rangle_{\mathfrak{S}_n} = \langle s_\lambda, s_\mu \rangle_\Lambda = \delta_{\lambda\mu}$$

which must mean that $\{\sigma_\lambda \colon \lambda \vdash n\}$ is a $\mathbb{Z}$-basis for $C\ell_V(\mathfrak{S}_n)$, and that each σ_λ is either an irreducible character or its negative. Thus, up to sign changes and permutations, the class functions σ_λ are just the characters χ_λ of the Specht modules indexed by λ (see §8.10). That is, $\sigma_\lambda = \pm\chi_{\pi(\lambda)}$, where π is a permutation of Par preserving size.

In fact, we claim that $\sigma_\lambda = \chi_\lambda$ for all λ. First, we confirm that the signs are positive. We can write each

Schur function as

$$s_\lambda = \sum_{\mu \vdash n} b_{\lambda,\mu} \frac{p_\mu}{z_\mu} \tag{9.26}$$

for some integers $b_{\lambda,\mu}$. Applying $\mathbf{ch}^{-1}$ gives

$$\sigma_\lambda = \sum_{\mu \vdash n} b_{\lambda,\mu} 1_\mu, \tag{9.27}$$

so that $b_{\lambda,\mu} = \pm\chi_{\pi(\lambda)}(C_\mu)$. In particular, taking $\mu = (1^n)$, the cycle-type of the identity permutation, we have

$$b_{\lambda,(1^n)} = \pm \dim \chi_{\pi(\lambda)}. \tag{9.28}$$

On the other hand, the only power-sum symmetric function that contains the squarefree monomial $x_1 x_2 \cdots x_n$ is $p_{(1^n)}$ (with coefficient $z_{(1^n)} = n!$). Extracting the coefficients of that monomial on both sides of (9.26) gives

$$f^\lambda = b_{\lambda,\mu}. \tag{9.29}$$

In particular, comparing (9.28) and (9.29), we see that the sign $\pm$ is positive for every λ. (We also have a strong hint that π is the identity permutation, because $\dim \chi_{\pi(\lambda)} = f^\lambda$.) □

We can now tie up a loose end from §8.10:

Corollary 9.11.3. *The Kostka numbers $K_{\lambda,\mu}$ give the multiplicities of the irreps of $\mathfrak{S}_n$ in the tabloid representations (V_μ, ρ_μ). In particular, the tabloid representations form a basis for the free abelian group of virtual characters.*

Proof. We calculate the multiplicity of each irrep in the tabloid representation using characters:

$$\begin{aligned}
\langle \mathbf{ch}^{-1}(s_\lambda),\, \tau_\mu \rangle_{\mathfrak{S}_n} &= \langle s_\lambda,\, \mathbf{ch}(\tau_\mu) \rangle_\Lambda && \text{(Thm. 9.11.2(2))} \\
&= \langle s_\lambda,\, h_\mu \rangle_\Lambda && \text{(Thm. 9.11.2(4))} \\
&= \sum_{\nu \vdash n} K_{\lambda,\nu} \langle m_\nu,\, h_\mu \rangle_\Lambda && \text{(eqn. (9.13) and bilinearity)} \\
&= \sum_{\nu \vdash n} K_{\lambda,\nu} \delta_{\nu\mu} && \text{(Definition 9.9.4)} \\
&= K_{\lambda,\mu}.
\end{aligned}$$

The second assertion follows because $K_{\lambda,\lambda} = 1$ for all λ by (9.14). □

9.12 What's next

The Frobenius characteristic allows us to translate back and forth between symmetric functions and characters of symmetric groups. In particular, many questions about representations of $\mathfrak{S}_n$ can now be answered in terms of tableau combinatorics. Here are a few fundamental things we would like to know at this point.

1. Irreducible characters. What is the value of the irreducible character $\chi_\lambda = \mathbf{ch}^{-1}(s_\lambda)$ on the conjugacy class C_μ? In other words, what is the character table of $\mathfrak{S}_n$? We have worked out some examples (e.g., $n = 3$, $n = 4$) and know that the values are all integers, since the Schur functions are an integral basis for Λ_n. A precise combinatorial formula is given by the **Murnaghan–Nakayama Rule**. (According to Stanley [Sta99,

p.410], this formula was first published by Littlewood and Richardson in 1934, predating Murnaghan (1937) and Nakayama (1941).)

2. Dimensions of irreducible characters. A special case of the Murnaghan–Nakayama Rule is that the irreducible representation with character χ_λ has dimension f^λ, the number of standard tableaux of shape λ. What are the numbers f^λ? There is a beautiful interpretation called the **hook-length formula of Frame, Robinson and Thrall**, which again has many, many proofs in the literature.

3. Littlewood–Richardson numbers. Now that we know how important the Schur functions are from a representation-theoretic standpoint, how do we multiply them? That is, suppose that μ, ν are partitions with $|\mu| = q$, $|\nu| = r$. Then $s_\mu s_\nu \in \Lambda_{q+r}$, so it has a unique expansion as a linear combination of Schur functions:

$$s_\mu s_\nu = \sum_\lambda c^\lambda_{\mu,\nu} s_\lambda, \qquad c^\lambda_{\mu,\nu} \in \mathbb{Z}. \tag{9.30}$$

The $c^\lambda_{\mu,\nu} \in \mathbb{Z}$ are called the **Littlewood–Richardson numbers**. They are the structure coefficients for Λ, regarded as an algebra generated as a vector space by the Schur functions. The $c^\lambda_{\mu,\nu}$ must be integers, because $s_\mu s_\nu$ is certainly a $\mathbb{Z}$-linear combination of the monomial symmetric functions, and the Schur functions are a $\mathbb{Z}$-basis.

Equation (9.30) is equivalent to

$$c^\lambda_{\mu,\nu} = \langle s_\mu s_\nu,\ s_\lambda \rangle_\Lambda$$

and applying $\mathbf{ch}^{-1}$ gives an interpretation of the $c^\lambda_{\mu,\nu}$ in terms of characters, namely

$$c^\lambda_{\mu,\nu} \ = \ \left\langle \operatorname{Ind}^{\mathfrak{S}_n}_{\mathfrak{S}_q \times \mathfrak{S}_r}(\chi_\mu \otimes \chi_\nu),\ \chi_\lambda \right\rangle_{\mathfrak{S}_n} \ = \ \left\langle \chi_\mu \otimes \chi_\nu,\ \operatorname{Res}^{\mathfrak{S}_n}_{\mathfrak{S}_q \times \mathfrak{S}_r}(\chi_\lambda) \right\rangle_{\mathfrak{S}_q \times \mathfrak{S}_r}$$

where the second equality comes from Frobenius reciprocity.

Any combinatorial interpretation for the numbers $c^\lambda_{\mu\nu}$ is called a **Littlewood–Richardson rule**; there are many of them.

4. Transition matrices. What are the coefficients of the transition matrices between different bases of Λ_n? We have worked out a few cases using the Cauchy kernel, and we have defined the Kostka numbers to be the transition coefficients from the m's to the s's (this is just the definition of the Schur functions).

9.13 Alternants and the classical definition of Schur functions

This section closely follows [Sta99, §7.15].

For this section, we will regard symmetric functions as polynomials rather than power series, for a reason that will quickly become apparent.

Definition 9.13.1. Let $\mathfrak{S}_n$ act on $R = \Bbbk[x_1, \ldots, x_n]$ by permuting variables. A polynomial $a \in R$ is **alternating**, or an **alternant**, if $w(a) = \varepsilon(w)a$ for all $w \in \mathfrak{S}_n$. Equivalently, interchanging any two variables x_i, x_j maps a to $-a$.

In particular, every alternant is divisible by $x_j - x_i$ for each $i < j$, hence by the **Vandermonde determinant**

$$V = \prod_{1 \le i < j \le n} (x_i - x_j) = \begin{vmatrix} x_1^{n-1} & x_1^{n-2} & \cdots & x_1 & 1 \\ x_2^{n-1} & x_2^{n-2} & \cdots & x_2 & 1 \\ \vdots & \vdots & \vdots & & \vdots \\ x_n^{n-1} & x_n^{n-2} & \cdots & x_n & 1 \end{vmatrix}.$$

(Why does this equality hold? Interchanging x_i with x_j swaps two rows of the determinant, hence changes its sign. Therefore the determinant is divisible by the product on the left. On the other hand, both polynomials are homogeneous of degree $\binom{n}{2} = 0+1+\cdots+(n-1)$, and the coefficients of $x_1^{n-1}x_2^{n-2}\cdots x_{n-1}^1 x_n^0$ are both $+1$, so equality must hold.) This is why we are working with polynomials: replacing $1 \leq i < j \leq n$ with $1 \leq i < j \leq \infty$ in the definition of V does not result in a well-defined power series.

We can construct more general alternants by changing the powers of variables that occur in each column of the Vandermonde determinant: for $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$, we define

$$a_\alpha = a_\alpha(x_1, \ldots, x_n) = \left|x_i^{\alpha_j}\right|_{i,j=1}^n = \sum_{w \in \mathfrak{S}_n} \varepsilon(w) w(\mathbf{x}^\alpha) \tag{9.31}$$

where $\mathbf{x}^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$. Note that $a_\alpha = 0$ if (and only if) α contains some entry more than once. Moreover, permuting the entries of α only changes a_α up to sign, so we might as well list them in decreasing order. For these two reasons, we can write α as the componentwise sum $\lambda + \delta$, where $\lambda = (\lambda_1 \geq \cdots \geq \lambda_n \geq 0) \in \mathsf{Par}$ and $\delta = (n-1, n-2, \ldots, 1, 0)$. That is, $\alpha_j = \lambda_j + \delta_j = \lambda_j + n - j$, and thus

$$a_{\lambda+\delta} = \left|x_i^{\lambda_j + n - j}\right|_{i,j=1}^n.$$

In particular, $a_\delta = V$. As observed above, every alternant is divisible by a_δ, so the quotient $a_{\lambda+\delta}/a_\delta$ is a polynomial; moreover, it is a *symmetric* polynomial, since each $w \in \mathfrak{S}_n$ scales it by $\varepsilon(w)/\varepsilon(w) = 1$.

Theorem 9.13.2. *For all λ, we have $a_{\lambda+\delta}/a_\delta = s_\lambda(x_1, \ldots, x_n)$.*

Proof. In light of the second assertion of Corollary 9.10.13 (specialized to the first n variables) and the invertibility of the matrix $[K_{\lambda\mu}]$, it is equivalent to show that for every $\mu = (\mu_1, \ldots, \mu_k)$ we have

$$e_\mu = \sum_\lambda K_{\tilde{\lambda}\mu} a_{\lambda+\delta}/a_\delta$$

or equivalently

$$a_\delta e_\mu = \sum_\lambda K_{\tilde{\lambda}\mu} a_{\lambda+\delta}.$$

Both sides of the equation are alternating, so it is enough to show that for every λ, the monomial $\mathbf{x}^{\lambda+\delta}$ has the same coefficient on both sides of this equation. On the RHS this coefficient is $K_{\tilde{\lambda}\mu}$ since the monomial only appears in the λ summand. On the LHS, the coefficient $[\mathbf{x}^{\lambda+\delta}]a_\delta e_\mu$ is the sum of $\varepsilon(w)$ over all factorizations

$$\mathbf{x}^{\lambda+\delta} = w(\mathbf{x}^\delta)\, \mathbf{x}^{\beta^1} \cdots \mathbf{x}^{\beta^k} = x_{w(1)}^0\, x_{w(2)}^1 \cdots x_{w(n)}^{n-1}\, \mathbf{x}^{\beta^1} \cdots \mathbf{x}^{\beta^k}.$$

where each $\mathbf{x}^{\beta^i}$ is a squarefree monomial of degree μ_i. Denote such a factorization by $f(w, \beta) = f(w, \beta^1, \ldots, \beta^k)$, and denote by F the set of all such factorizations. Thus we are trying to prove that

$$\sum_{f(w,\beta) \in F} \varepsilon(w) = K_{\tilde{\lambda}\mu}. \tag{9.32}$$

Let $f(w,\beta)_j$ denote the partial product $w(\mathbf{x}^\delta)\mathbf{x}^{\beta^1}\cdots\mathbf{x}^{\beta^j}$. For a monomial M, let $\mathrm{pow}_{x_i}(M)$ denote the power of x_i that appears in M.

We now describe a sign-reversing involution on the set F. Suppose that $f(w,\beta)$ is a factorization such that for some $j \in [k]$ and some $a \neq b$

$$\mathrm{pow}_a(f(w,\beta)_j) = \mathrm{pow}_b(f(w,\beta)_j).$$

Choose $(j, \{a, b\})$ to be lexicographically minimal. Then interchanging x_a and x_b in every $\mathbf{x}^{\beta^i}$ and multiplying w by the transposition $(a\ b)$ produces another element of F and preserves the equality condition and the pair $(j, \{a, b\})$, while flipping the sign of w.

For example, let $n = 3$, $\lambda = (2, 2, 1)$, $\alpha = (4, 3, 1)$, $\mu = (2, 2, 1)$. The set F contains eight factorizations of $\mathbf{x}^\alpha = x_1^4 x_2^3 x_3$, including three cancelling pairs:

w	$\varepsilon(w)$	$w(\mathbf{x}^\delta)$	$\mathbf{x}^{\beta^1}$	$\mathbf{x}^{\beta^2}$	$\mathbf{x}^{\beta^3}$	$j, \{a, b\}$
123	1	$x_1^2 x_2$	$x_1 x_2$	$x_1 x_2$	x_3	$-$
123	1	$x_1^2 x_2$	$x_1 x_2$	$x_1 x_3$	x_2	$-$
123	1	$x_1^2 x_2$	$x_1 x_3$	$x_1 x_2$	x_2	$1, \{2, 3\}$
132	-1	$x_1^2 x_3$	$x_1 x_2$	$x_1 x_2$	x_2	$1, \{2, 3\}$
123	1	$x_1^2 x_2$	$x_1 x_2$	$x_2 x_3$	x_1	$2, \{1, 2\}$
213	-1	$x_2^2 x_1$	$x_1 x_2$	$x_1 x_3$	x_1	$2, \{1, 2\}$
123	1	$x_1^2 x_2$	$x_2 x_3$	$x_1 x_2$	x_1	$1, \{1, 2\}$
213	-1	$x_2^2 x_1$	$x_1 x_3$	$x_1 x_2$	x_1	$1, \{1, 2\}$

The uncanceled factorizations $f(w, \beta)$ are those for which, in every partial product $f(w, \beta)_j$ all variables occur with different powers. But in fact this condition implies $w = \mathrm{Id}$, for otherwise, there are indices $a < b$ for which

$$\begin{aligned} \mathrm{pow}_a(w(\mathbf{x}^\delta)) = \mathrm{pow}_a(f(w,\beta)_0) \ &< \ \mathrm{pow}_b(f(w,\beta)_0) = \mathrm{pow}_b(w(\mathbf{x}^\delta)) \qquad \text{but certainly} \\ \mathrm{pow}_a(\mathbf{x}^{\delta+\lambda}) = \mathrm{pow}_a(f(w,\beta)_k) \ &> \ \mathrm{pow}_b(f(w,\beta)_k) = \mathrm{pow}_b(\mathbf{x}^{\delta+\lambda}) \end{aligned}$$

but since the $\mathbf{x}^{\beta^i}$ are all squarefree, there must be some j such that

$$\mathrm{pow}_a(f(w,\beta)_j) \ = \ \mathrm{pow}_b(f(w,\beta)_j)$$

(basically, by the intermediate value theorem).

In particular, the coefficient $[\mathbf{x}^{\lambda+\delta}]a_\delta e_\mu$ is positive: it is the number of factorizations of $\mathbf{x}^\lambda$ into squarefree monomials $\mathbf{x}^{\beta^1}, \ldots, \mathbf{x}^{\beta^k}$ of degrees $\mu_1, \ldots, \mu_k$ so that for all $j \le k$ we have

$$\mathrm{pow}_1(\mathbf{x}^{\beta^1} \cdots \mathbf{x}^{\beta^j}) \ge \mathrm{pow}_2(\mathbf{x}^{\beta^1} \cdots \mathbf{x}^{\beta^j}) \ge \cdots \ge \mathrm{pow}_n(\mathbf{x}^{\beta^1} \cdots \mathbf{x}^{\beta^j}). \tag{9.33}$$

Thus each variable x_j must occur in λ_i of the monomials $\mathbf{x}^{\beta^i}$. We record the list of monomials by a tableau of content μ whose entries correspond to monomials $\mathbf{x}^{\beta^i}$ and whose columns correspond to variables x_j: column j contains an i if x_j occurs in $\mathbf{x}^{\alpha_i}$. Thus the tableau has shape $\tilde{\lambda}$. We can arrange each column in increasing order, so the the entry in (i, j) tells us the ith monomial divisible by x_j. Continuing our example, the two factorizations of $\mathbf{x}^\lambda$ that remain uncancelled (see the preceding table) give rise to tableaux as follows:

$x_1x_2 \cdot x_1x_2 \cdot x_3$ $\qquad$ $x_1x_2 \cdot x_1x_3 \cdot x_2$

1	1	3
2	2	

1	1	2
2	3	

There are no repeats in columns because no variable occurs more than once in any $\mathbf{x}^{\beta^i}$. Moreover, if the ith row has a strict decrease $a > b$ between the jth and $(j+1)$st columns, then this means that the ith

occurrence of x_j occurs later than the ith occurrence of x_i — i.e., there are more x_{j+1}'s then x_j in the first b monomials, which contradicts (9.33). Hence the tableau is column-strict. Moreover, every column-strict tableau of shape $\tilde{\lambda}$ and content μ gives rise to a factorization that contributes 1 to the coefficient $[\mathbf{x}^{\lambda+\delta}]a_\delta e_\mu$. We conclude that the coefficient is $K_{\tilde{\lambda}\mu}$ as desired. □

9.14 The Murnaghan–Nakayama Rule

We know from Theorem 9.11.2 that the irreducible characters of $\mathfrak{S}_n$ are $\chi_\lambda = \mathbf{ch}^{-1}(s_\lambda)$ for $\lambda \vdash n$. We want to compute these numbers. Via the Frobenius characteristic, this problem is equivalent to expanding the Schur functions (which correspond to irreducible characters) as linear combinations of the power-sums (which correspond to indicator functions of conjugacy classes). We will need the description of Schur functions as quotients of alternants in §9.13, and the key step will be expressing a product $s_\nu p_r$ as a linear combination of Schur functions (equation (9.35)).

We first state the result, then prove it. The relevant combinatorial objects are **ribbons** and **ribbon tableau**. A **ribbon** is a connected[8] skew shape R with no 2×2 block, or equivalently with no square both north and west of another square. The **size** $|R|$ is as usual the number of squares in the ribbon, and its **height** $h(R)$ is the number of rows.[9]

A **ribbon tableau** is a decomposition of a Ferrers diagram into ribbons $R_1, \ldots, R_k$ such that for each $i \leq k$, the union of the first i ribbons forms a Ferrer diagram. Here is an example of a ribbon tableau of shape $\lambda = (8, 7, 6, 6, 4)$ into $k = 6$ ribbons.

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

Note that each row and column is weakly increasing, and that for each $i \leq k$, the union $R_1 \cup \cdots \cup R_i$ is a partition. In this context ribbons are often called **border strips** or **rim hooks**.

The list ρ of sizes of the ribbons is the **content** of the ribbon tableau; here $\rho = (6, 3, 4, 7, 4, 7)$. Let $RT(\lambda, \rho)$ denote the set of ribbon tableaux of shape λ and content ρ, and for $T = (R_1, \ldots, R_k) \in RT(\lambda, \rho)$ put

$$(-1)^T = \prod_{i=1}^{k} (-1)^{1+\mathrm{ht}(R_i)}.$$

For example, the heights of $R_1, \ldots, R_6$ in the ribbon tableau T shown above are 4, 3, 3, 3, 2, 4. There are an odd number of even heights, so $(-1)^T = -1$.

We first make an observation about ribbons.

Lemma 9.14.1. *Consider a pair of partitions λ, ν, where λ is obtained from ν by adding a ribbon R. Suppose that the squares of R appear in rows $i, i+1, \ldots, j$, so that $\lambda_k = \nu_k$ for $k < i$ or $k > j$. Then:*

[8] "Connected" means "connected with respect to sharing edges, not just diagonals", or equivalently "the topological interior is connected"; for example, the skew shape $21/1 =$  is *not* considered to be connected.

[9] Stanley [Sta99, §7.17] defines the height as *one less than* the number of rows, which simplifies the formulas but seems less natural to me.

1. $\lambda_i \in [\nu_i + 1, \nu_{i-1}]$.
2. *For each* $k \in [i+1, j]$ *we have* $\lambda_k = \nu_{k-1} + 1$.

Proof. For (1), we have $\lambda_i \leq \lambda_{i-1} = \nu_{i-1}$; on the other hand, λ_i is obtained by adding at least one box to ν_i. (In particular this interval cannot be empty — if we were able to add the entire ribbon R, we could have just added just one box in the ith row of ν and nothing else, so it must be the case that $\nu_{i-1} > \nu_i$.)

(2) asserts that the last box in the kth row of λ must be one column east and one row south of the last box in the $(k-1)$st row of ν. Indeed, any further west and R would not be connected; any further east and R would have a 2×2 block. □

Now we can state and prove the main result.

Theorem 9.14.2 (Murnaghan–Nakayama Rule)**.** *For all* $\lambda, \mu \vdash n$*, the value of the irreducible character* χ_λ *on the conjugacy class* C_μ *is*

$$\chi_\lambda(C_\mu) = \sum_{T \in RT(\lambda,\mu)} (-1)^T.$$

Proof. Fix a partition ν of length $\leq n$, let $\delta = (n-1, n-2, \ldots, 1, 0)$ as usual and let $\alpha = \nu + \delta$, i.e.,

$$\alpha_i = \nu_i + n - i.$$

Let a_α be the alternant of (9.31), and let ϵ_j be the sequence with a 1 in position j and 0s elsewhere. For $r \in \mathbb{N}$, we have

$$\begin{aligned}
a_\alpha p_r(x_1, \ldots, x_n) &= \sum_{w \in \mathfrak{S}_n} \varepsilon(w) w(\mathbf{x}^\alpha)(x_1^r + \cdots + x_n^r) \\
&= \sum_{w \in \mathfrak{S}_n} \varepsilon(w) x_{w(1)}^{\alpha_1} \cdots x_{w(n)}^{\alpha_n} (x_{w(1)}^r + \cdots + x_{w(n)}^r) \\
&= \sum_{w \in \mathfrak{S}_n} \varepsilon(w) \sum_{j=1}^{n} w(\mathbf{x}^{\alpha + r\epsilon_j}) \\
&= \sum_{j=1}^{n} \sum_{w \in \mathfrak{S}_n} \varepsilon(w) w(\mathbf{x}^{\alpha + r\epsilon_j}) \\
&= \sum_{j=1}^{n} a_{\alpha + r\epsilon_j}. \qquad (9.34)
\end{aligned}$$

If two entries of $\alpha + r\epsilon_j$ are equal, then $a_{\alpha + r\epsilon_j} = 0$. Otherwise, there is some $i \in [j]$ such that

$$\alpha_{i-1} > (\alpha + r\epsilon_j)_j > \alpha_i$$

or equivalently

$$\nu_{i-1} + n - (i-1) \;>\; \nu_j + n - j + r \;>\; \nu_i + n - i.$$

(If $i = 1$, just ignore the first inequality.) Therefore, sorting the parts of $\alpha + r\epsilon_j$ in decreasing order means moving the j^{th} part back to position i and pushing parts $i, i+1, \ldots, j-1$ up — that is, acting by a $(j-i+1)$-cycle, which has sign $(-1)^{j-i}$. That is, $a_{\alpha + r\epsilon_j} = (-1)^{j-i} a_{\lambda + \delta}$, where

$$\begin{aligned}
&\lambda + \delta = (\alpha_1, \ldots, \alpha_{i-1},\ \alpha_j + r, \alpha_i, \ldots, \alpha_{j-1},\ \alpha_{j+1}, \ldots, \alpha_n) \\
&\therefore \quad \lambda = (\nu_1, \ldots, \nu_{i-1},\ \nu_j + i - j + r, \nu_i + 1, \ldots, \nu_{j-1} + 1,\ \nu_{j+1}, \ldots, \nu_n).
\end{aligned}$$

Now a miracle occurs: by Lemma 9.14.1, these partitions λ are precisely the ones for which λ/ν is a ribbon of size r, spanning rows $i, \ldots, j$ and hence of height $j - i + 1$. Combining this observation with (9.34) we get

$$\begin{aligned} a_{\nu+\delta}\, p_r = a_\alpha\, p_r &= \sum_{j=1}^{n} a_{\alpha + r\epsilon_j} \\ &= \sum_{R,\lambda} (-1)^{\mathrm{ht}(R)+1} a_{\lambda+\delta} \end{aligned}$$

where the sum runs over ribbons R of size r that can be added to ν to obtain a partition λ. Dividing both sides by a_δ and applying Theorem 9.13.2 gives

$$s_\nu\, p_r = \sum_{R,\lambda} (-1)^{\mathrm{ht}(R)+1} s_\lambda. \tag{9.35}$$

(This is valid on the level of power series as well as for polynomials, since it remains valid under increasing the number of variables, so the coefficient of every monomial in the power series is equal on both sides.)

Now, let $\mu = (\mu_1, \ldots, \mu_k)$, so that $p_\mu = p_{\mu_1} \cdots p_{\mu_k}$. Iterating (9.35) gives

$$s_\nu\, p_\mu = \sum_{R_1,\ldots,R_k,\lambda} \prod_{i=1}^{k} (-1)^{\mathrm{ht}(R_i)+1} s_\lambda \tag{9.36}$$

where the sum runs over k-tuples of ribbons of lengths given by the parts of μ that can be added to ν to obtain λ. In particular, if $\nu = \emptyset$, then this is simply the statement that $T = (R_1, \ldots, R_k)$ is a ribbon tableau of shape λ and content μ, and the sign is $(-1)^T$, so we get

$$p_\mu = \sum_{\lambda} \sum_{T \in RT(\lambda,\mu)} (-1)^T s_\lambda \tag{9.37}$$

so that

$$\begin{aligned} \sum_{T \in RT(\lambda,\mu)} (-1)^T = \langle p_\mu, s_\lambda \rangle_\Lambda &= \langle \mathbf{ch}^{-1}(p_\mu),\, \mathbf{ch}^{-1}(s_\lambda) \rangle_{\mathfrak{S}_n} && \text{(since } \mathbf{ch}^{-1} \text{ is an isometry)} \\ &= \langle z_\mu 1_\mu,\, \chi_\lambda \rangle_{\mathfrak{S}_n} && \text{(by 1 and 8 of Thm. 9.11.2)} \\ &= \frac{1}{n!} \sum_{w \in \mathfrak{S}_n} z_\mu 1_\mu(w) \chi_\lambda(w) \\ &= \frac{z_\mu}{n!} \, |C_\mu| \, \chi_\lambda(C_\mu) \\ &= \chi_\lambda(C_\mu) \end{aligned}$$

completing the proof of the Murnaghan–Nakayama rule. □

As a first consequence, we can expand the Schur functions in the power-sum basis:

Corollary 9.14.3. *For all $\lambda \vdash n$ we have*

$$s_\lambda = \sum_\mu \chi_\lambda(C_\mu) \frac{p_\mu}{z_\mu} \qquad \textit{and} \qquad s_{\tilde{\lambda}} = \sum_\mu \chi_\lambda(C_\mu) \frac{\varepsilon_\mu p_\mu}{z_\mu}.$$

Proof. Write s_λ in the p-basis as $\sum_\mu b_{\lambda\mu} p_\mu$. Taking the Hall inner product of both sides with p_μ gives $\langle s_\lambda, p_\mu \rangle = b_{\lambda\mu} z_\mu$, or $b_{\lambda\mu} = z_\mu^{-1} \langle s_\lambda, p_\mu \rangle$, implying the first equality. Applying ω and invoking Corollaries 9.8.2 and 9.9.33 gives the second equality. □

An important special case of the Murnaghan–Nakayama rule is when $\mu = (1, 1, \ldots, 1)$, since then $\chi_\lambda(C_\mu) = \chi_\lambda(\mathrm{Id}_{\mathfrak{S}_n})$, is just the dimension of the irreducible character χ_λ. On the other hand, a ribbon tableau of content μ is nothing more or less than a standard tableau! So the Murnaghan–Nakayama Rule implies the following:

Corollary 9.14.4. $\dim \chi_\lambda = f^\lambda$, *the number of standard tableaux of shape* λ.

Thus it is of great interest to calculate f^λ; we will do that in the next section.

Another interesting observation is that the parameter μ can be taken to be a composition rather than a permutation: Murnaghan–Nakayama coefficient $\sum_\lambda \sum_{T \in RT(\lambda,\mu)} (-1)^T$ is invariant for any rearrangement of μ (which makes sense, since the cycle-type of a permutation does not carry any inherent order). w

Example 9.14.5. Let $n = 5$, $\lambda = (3, 2)$, and $\mu = (2, 2, 1)$. There are 5 standard tableaux of shape λ (see Problem 1.11), so the irreducible character σ_λ must be either χ_5 or χ_6 of Problem 8.6, so certainly $\sigma_\lambda(C_\mu) = 1$. We calculate $\sum_\lambda \sum_{T \in RT(\lambda,\rho)} (-1)^T$ for each permutation ρ of μ:

$\rho = 221$

T	$(-1)^T$
$\begin{array}{\|c\|c\|c\|}\hline 1&1&3\\\hline 2&2\\\cline{1-2}\end{array}$	$(-1)^{0+0+0} = 1$
$\begin{array}{\|c\|c\|c\|}\hline 1&2&2\\\hline 1&3\\\cline{1-2}\end{array}$	$(-1)^{1+0+0} = -1$
$\begin{array}{\|c\|c\|c\|}\hline 1&2&3\\\hline 1&2\\\cline{1-2}\end{array}$	$(-1)^{1+1+0} = 1$
Total	1

$\rho = 212$

T	$(-1)^T$
$\begin{array}{\|c\|c\|c\|}\hline 1&1&2\\\hline 3&3\\\cline{1-2}\end{array}$	$(-1)^{0+0+0} = 1$
Total	1

$\rho = 122$

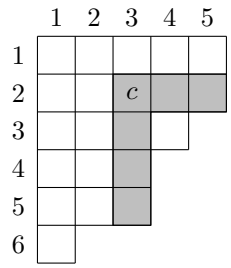
T	$(-1)^T$
$\begin{array}{\|c\|c\|c\|}\hline 1&2&2\\\hline 3&3\\\cline{1-2}\end{array}$	$(-1)^{0+0+0} = 1$
Total	1

◀

9.15 The Hook-Length Formula

Let $\lambda \vdash n$, let $\ell = \ell(\lambda)$, and let $\mathsf{SYT}(\lambda)$ the set of standard tableaux of shape λ, so $f^\lambda = |\mathsf{SYT}(\lambda)|$. In what follows, we label the rows and columns of a tableau starting at 1. If $c = (i, j)$ is the cell in the ith row and jth column of a tableau T, then $T(c)$ or $T(i, j)$ denotes the entry in that cell.

The **hook** $H(c)$ defined by a cell $c = (i, j)$ consists of itself together with all the cells due east or due south of it. The number of cells in the hook is the **hook length**, written $h(c)$ or $h(i, j)$. (In this section, the letter h always refers to hook lengths, never to the complete homogeneous symmetric function.) In the following example, $h(c) = h(2, 3) = 6$.



Explicitly, $h(i,j) = \lambda_i - (i-1) + \tilde{\lambda}_j - (j-1) - 1 = \lambda_i + \tilde{\lambda}_j - i - j + 1$ where $\tilde{\lambda}$ is the conjugate partition to λ (although we shall not need this formula in what follows).

Theorem 9.15.1 (Hook-Length Formula)**.** *Let $\lambda \vdash n$. Then the number f^λ of standard Young tableaux of shape λ equals $F(\lambda)$, where*

$$F(\lambda) \;=\; \frac{n!}{\prod_{c \in \lambda} h(c)}.$$

Example 9.15.2. For $\lambda = (5,3,3,1) \vdash 12$, the tableau of hook lengths is

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

so $f^\lambda = 12!/(8\cdot 6\cdot 5^2\cdot 4\cdot 3\cdot 2^3\cdot 1^3) = 4158$. As another example, if $\lambda = (n,n) \vdash 2n$, the hook lengths are $n+1, n, n-1, \ldots, 2$ (in the top row) and $n, n-1, n-2, \ldots, 1$ (in the bottom row). Therefore $f^\lambda = \frac{(2n)!}{(n+1)!\,n!} = \frac{1}{n+1}\binom{2n}{n}$, the nth Catalan number (as we already know). ◀

Here is how *not* to prove the hook-length formula. Consider the discrete probability space of all $n!$ fillings of the Ferrers diagram of λ with the numbers $1, \ldots, n$. Let S be the event that a uniformly chosen filling T is a standard tableau, and for each cell, let X_c be the event that $T(c)$ is the smallest number in the hook $H(c)$. Then $S = \bigcap_c X_c$, and $\Pr[X_c] = 1/h(c)$. We would like to conclude that $\Pr[S] = \prod_c 1/h(c)$, which would imply the hook-length formula. However, that inference would require that the events X_c are mutually independent, which they certainly are not! Still, this is a nice heuristic argument (attributed by Wikipedia to Knuth) that one can at least remember.

There are many proofs of the hook-length formula in the literature. This one is due to Greene, Nijenhuis and Wilf [GNW79].

Proof of Theorem 9.15.1. First, observe that for every $T \in \mathsf{SYT}(\lambda)$, the cell $c \in T$ containing the number $n = |\lambda|$ must be a corner of λ (i.e., the rightmost cell in its row and the bottom cell in its column). Deleting c produces a standard tableau of size $n-1$; we will call the resulting partition $\lambda - c$. This construction gives a collection of bijections

$$\{T \in \mathsf{SYT}(\lambda) \colon T(c) = n\} \to \mathsf{SYT}(\lambda - c)$$

for each corner c.

Now to the main argument. We will prove by induction on n that $f^\lambda = F(\lambda)$. The base case $n = 1$ is clear. For the inductive step, we wish to show that

$$F(\lambda) = \sum_{\text{corners } c} F(\lambda - c) \qquad \text{or equivalently} \qquad \sum_{\text{corners } c} \frac{F(\lambda - c)}{F(\lambda)} = 1 \tag{9.38}$$

since by the inductive hypothesis together with the bijections just described, the right-hand side of the first equation equals f^λ.

Let $c = (x, y)$ be a corner cell. Removing c decreases by 1 the sizes of the hooks $H(c')$ for cells c' strictly north or west of c, and leaves all other hook sizes unchanged. Therefore,

$$\begin{aligned}
\frac{F(\lambda - c)}{F(\lambda)} &= \frac{(n-1)!}{n!} \prod_{i=1}^{x-1} \frac{h(i,y)}{h(i,y)-1} \prod_{j=1}^{y-1} \frac{h(x,j)}{h(x,j)-1} \\
&= \frac{1}{n} \prod_{i=1}^{x-1} \left(1 + \frac{1}{h(i,y)-1}\right) \prod_{j=1}^{y-1} \left(1 + \frac{1}{h(x,j)-1}\right) \\
&= \frac{1}{n} \sum_{\substack{A \subseteq [x-1] \\ B \subseteq [y-1]}} \left(\prod_{i \in A} \frac{1}{h(i,y)-1}\right) \left(\prod_{j \in B} \frac{1}{h(x,j)-1}\right).
\end{aligned} \tag{9.39}$$

Consider the following random process (called a **hook walk**). First choose a cell (a_0, b_0) uniformly from λ. Then for each $t = 1, 2, \ldots$, move to a cell (a_t, b_t) chosen uniformly from all other cells in $H(a_{t-1}, b_{t-1})$. The process stops when it reaches a corner; let p_c be the probability of reaching a particular corner c. Evidently $\sum_c p_c = 1$. Our goal now becomes to show that

$$p_c = \frac{F(\lambda - c)}{F(\lambda)} \tag{9.40}$$

which will establish (9.38).

Consider a hook walk starting at $(a, b) = (a_1, b_1)$ and ending at $(a_m, b_m) = (x, y)$. Let $A = \{a_1, \ldots, a_m\}$ and $B = \{b_1, \ldots, b_m\}$ be the sets of rows and columns encountered (removing duplicates); call these sets the horizontal and vertical **projections** of W. Let

$$p(A, B \mid a, b)$$

denote the probability that a hook walk starting at (a, b) has projections A and B. We claim that

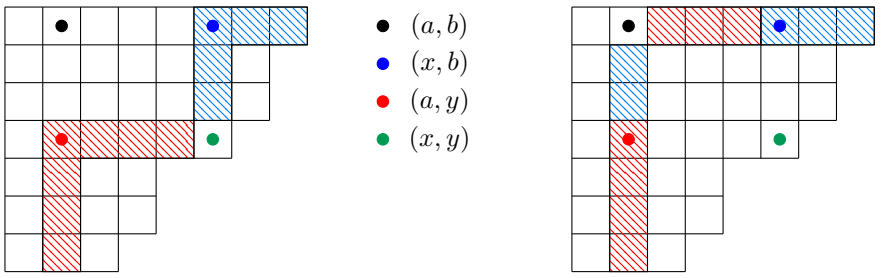
$$p(A, B \mid a, b) = \underbrace{\left(\prod_{i \in A \setminus x} \frac{1}{h(i,y)-1}\right) \left(\prod_{j \in B \setminus y} \frac{1}{h(x,j)-1}\right)}_{\Phi}. \tag{9.41}$$

We prove this by induction on m. If $m = 1$, then either $A = \{a\} = \{x\}$ and $B = \{b\} = \{y\}$, and the equation

reduces to $1 = 1$ (the RHS is the empty product), or else it reduces to $0 = 0$. If $m > 1$, then

$$\begin{aligned} p(A, B \mid a, b) &= \underbrace{\frac{p(A \setminus a_1, B \mid a_2, b_1)}{h(a,b) - 1}}_{\text{first move south to } (a_2, b_1)} + \underbrace{\frac{p(A, B \setminus b_1 \mid a_1, b_2)}{h(a,b) - 1}}_{\text{first move east to } (a_1, b_2)} \\ &= \frac{1}{h(a,b) - 1}\Big((h(a,y) - 1)\Phi + (h(x,b) - 1)\Phi\Big) \qquad \text{(by induction)} \\ &= \left(\frac{h(a,y) - 1 + h(x,b) - 1}{h(a,b) - 1}\right)\Phi. \end{aligned} \tag{9.42}$$

To see that the parenthesized expression in (9.42) is 1, consider the following diagram, with the hooks at (a, y) and (x, b) shaded in red and blue respectively, with the corner (x, y) omitted so that there are a total of $h(a, y) - 1 + h(x, b) - 1$ shaded cells. Pushing some red cells north and some blue cells to the left produces the hook at (a, b) with one cell omitted, as on the right.



$h(a, y) - 1 + h(x, b) - 1$ shaded cells

$h(a, b) - 1$ shaded cells

This proves (9.41). Now we compute p_c, the probability that a walk ends at a particular corner $c = (x, y)$. Equivalently, $x \in A$ and $y \in B$; equivalently, $A \subseteq [x]$ and $B \subseteq [y]$. Therefore, summing over all possible starting positions, we have

$$\begin{aligned} p_c &= \frac{1}{n} \sum_{\substack{(A,B,a,b): \\ A \subseteq [x],\ B \subseteq [y] \\ a = \min A,\ b = \min B \\ x = \max A,\ y = \max B}} p(A, B \mid a, b) \\ &= \frac{1}{n} \sum_{\substack{(A,B,a,b) \\ \text{as above}}} \left(\prod_{i \in A \setminus x} \frac{1}{h(i,y) - 1}\right)\left(\prod_{j \in B \setminus y} \frac{1}{h(x,j) - 1}\right) \qquad \text{(by (9.41))} \\ &= \frac{1}{n} \sum_{\substack{A \subseteq [x-1] \\ B \subseteq [y-1]}} \left(\prod_{i \in A} \frac{1}{h(i,y) - 1}\right)\left(\prod_{j \in B} \frac{1}{h(x,j) - 1}\right) \end{aligned}$$

which is precisely (9.39). This establishes (9.40) and completes the proof. □

9.16 The Littlewood–Richardson Rule

UNDER CONSTRUCTION

Recall that the Littlewood–Richardson coefficients $c^\lambda_{\mu\nu}$ are the structure coefficients for Λ as an algebra with vector space basis $\{s_\lambda : \lambda \in \mathsf{Par}\}$: that is,

$$s_\mu s_\nu = \sum_\lambda c^\lambda_{\mu\nu} s_\lambda.$$

We begin by proving an important interpretation of the Littlewood–Richardson coefficients in terms of skew Schur functions (see Definitions 9.7.6 and 9.7.7). The skew Schur function $s_{\lambda/\mu}$ is symmetric, so they too can be expanded in the Schur basis as

$$s_{\lambda/\mu} = \sum_\nu \tilde{c}_{\lambda/\mu,\nu}\, s_\nu$$

where $\tilde{c}_{\lambda/\mu,\nu} \in \mathbb{Z}$ for all λ, μ, ν. In fact these numbers are also Littlewood–Richardson coefficients, and they are symmetric in μ and ν (which is hardly obvious from the definition).

Proposition 9.16.1. *Let $\mathbf{x} = \{x_1, x_2, \dots\}$, $\mathbf{y} = \{y_1, y_2, \dots\}$ be two countably infinite sets of variables. Then*

$$s_\lambda(\mathbf{x}, \mathbf{y}) = \sum_{\mu \subseteq \lambda} s_\mu(\mathbf{x}) s_{\lambda/\mu}(\mathbf{y}).$$

Proof. Consider column-strict tableaux of shape λ with labels taken from the alphabet $1 < 2 < \cdots < 1' < 2' < \cdots$, and let the weight of such a tableau T be $\mathbf{x}^\alpha \mathbf{y}^\beta$, where α_i (resp., β_i) is the number of cells filled with i (resp., i'). Then the left-hand side is the generating function for all schools tableaux by weight. On the other hand, such a tableau consists of a CST of shape μ filled with $1, 2, \dots$ (for some $\mu \subseteq \lambda$) together with a CST of shape λ/μ filled with $1', 2', \dots$, so the RHS enumerates the same set of tableaux. □

Theorem 9.16.2. *For all partitions λ, μ, ν, we have*

$$\tilde{c}_{\lambda/\mu,\nu} = c^\lambda_{\mu,\nu} = c^\lambda_{\nu,\mu}.$$

Equivalently,

$$\langle s_\mu s_\nu,\ s_\lambda \rangle_\Lambda \ = \ \langle s_\nu,\ s_{\lambda/\mu} \rangle_\Lambda.$$

Proof. We need three countably infinite sets of variables $\mathbf{x}, \mathbf{y}, \mathbf{z}$ for this. Consider the "double Cauchy kernel"

$$\Omega(\mathbf{x}, \mathbf{z})\Omega(\mathbf{y}, \mathbf{z}) \ = \ \prod_{i,j} (1 - x_i z_j)^{-1} \prod_{i,j} (1 - y_i z_j)^{-1}.$$

On the one hand, expanding both factors in terms of Schur functions and then applying the definition of the Littlewood–Richardson coefficients to the $\mathbf{z}$ terms gives

$$\begin{aligned} \Omega(\mathbf{x}, \mathbf{z})\Omega(\mathbf{y}, \mathbf{z}) \ &= \ \left(\sum_\mu s_\mu(\mathbf{x}) s_\mu(\mathbf{z}) \right) \left(\sum_\nu s_\nu(\mathbf{y}) s_\nu(\mathbf{z}) \right) \ = \ \sum_{\mu,\nu} s_\mu(\mathbf{x}) s_\nu(\mathbf{y}) s_\mu(\mathbf{z}) s_\nu(\mathbf{z}) \\ &= \sum_{\mu,\nu} s_\mu(\mathbf{x}) s_\nu(\mathbf{y}) \sum_\lambda c^\lambda_{\mu,\nu} s_\lambda(\mathbf{z}). \end{aligned} \tag{9.43}$$

On the other hand, we also have (formally setting $s_{\lambda/\mu} = 0$ if $\mu \not\subseteq \lambda$)

$$\begin{aligned}\Omega(\mathbf{x},\mathbf{z})\Omega(\mathbf{y},\mathbf{z}) &= \sum_\lambda s_\lambda(\mathbf{x},\mathbf{y})s_\lambda(\mathbf{z}) = \sum_\lambda \sum_{\mu\subseteq\lambda} s_\mu(\mathbf{x})s_{\lambda/\mu}(\mathbf{y})s_\lambda(\mathbf{z}) \\ &= \sum_\lambda \sum_\mu s_\mu(\mathbf{x})s_\lambda(\mathbf{z}) \sum_\nu \tilde{c}_{\lambda/\mu,\nu}s_\nu(\mathbf{y}) \\ &= \sum_{\mu,\nu} s_\mu(\mathbf{x})s_\nu(\mathbf{y}) \sum_\lambda s_\lambda(\mathbf{z})\tilde{c}_{\lambda/\mu,\nu}. \end{aligned} \tag{9.44}$$

(The first equality is perhaps clearer in reverse; think about how to express the right-hand side as an infinite product over the variable sets $\mathbf{x} \cup \mathbf{y}$ and $\mathbf{z}$. The second equality uses Proposition 9.16.1.) Now the theorem follows from the equality of (9.43) and (9.44). □

There are a lot of combinatorial interpretations of the Littlewood–Richardson numbers. Here is one. A **ballot sequence** (or **Yamanouchi word**, or **lattice permutation**) is a sequence of positive integers such that each initial subsequence contains at least as many 1's as 2's, at least as many 2's as 3's, et cetera.

Theorem 9.16.3 (Littlewood–Richardson Rule)**.** *$c^\lambda_{\mu,\nu}$ equals the number of column-strict tableaux T of shape λ/μ, and content ν such that the word obtained by reading the entries of T row by row, right to left, top to bottom, is a ballot sequence.*

Include a proof. There are a lot of them but they tend to be hard.

Important special cases are the **Pieri rules**, which describe how to multiply by the Schur function corresponding to a single row or column (i.e., by an h or an e).

Theorem 9.16.4 (Pieri Rules)**.** *Let (k) denote the partition with a single row of length k, and let (1^k) denote the partition with a single column of length k. Then*

$$s_\mu s_{(k)} = s_\mu h_k = \sum_\lambda s_\lambda$$

where λ ranges over all partitions obtained from μ by adding k boxes, no more than one in each column; and

$$s_\mu s_{(1^k)} = s_\mu e_k = \sum_\lambda s_\lambda$$

where λ ranges over all partitions obtained from μ by adding k boxes, no more than one in each row.

Another important, even more special case is

$$s_\mu s_1 = \sum_\lambda s_\lambda$$

where λ ranges over all partitions obtained from μ by adding a single box. Via the Frobenius characteristic, this gives a "branching rule" for how the restriction of an irreducible character of $\mathfrak{S}_n$ splits into a sum of irreducibles when restricted:

$$\mathrm{Res}^{\mathfrak{S}_n}_{\mathfrak{S}_{n-1}}(\chi_\lambda) = \oplus_\mu \chi_\mu$$

where now μ ranges over all partitions obtained from λ by *deleting* a single box. **Details?**

9.17 Knuth equivalence and jeu de taquin

Definition 9.17.1. Let $\mathbf{b}, \mathbf{b}'$ be finite ordered lists of positive integers (or "words in the alphabet $\mathbb{N}_{>0}$"). We say that $\mathbf{b}, \mathbf{b}'$ are **Knuth equivalent**, written $\mathbf{b} \sim \mathbf{b}'$, if one can be obtained from the other by a sequence of transpositions as follows:

1. If $x \le y < z$, then $\cdots xzy \cdots \sim \cdots zxy \cdots$.
2. If $x < y \le z$, then $\cdots yxz \cdots \sim \cdots yzx \cdots$.

(Here the notation $\cdots xzy \cdots$ means a word that contains the letters x, z, y consecutively.)

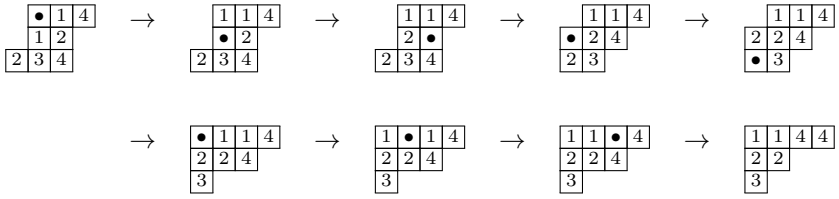
For example, $21221312 \sim 21223112$ by Rule 1, and $21223112 \sim 21221312$ by Rule 2 (applied in reverse).

This definition looks completely unmotivated at first, but hold that thought!

We now define an equivalence relation on column-strict skew tableaux, called **jeu de taquin**[10]. The rule is as follows:

That is, for each inner corner of T — that is, an empty cell that has numbers to the south and east, say x and y — then we can either slide x north into the empty cell (if $x \le y$) or slide y west into the empty cell (if $x > y$). It is not hard to see that any such slide (hence, any sequence of slides) preserves the property of column-strictness.

For example, the following is a sequence of jeu de taquin moves. The bullets • denote the inner corner that is being slid into.



(9.45)

If two skew tableaux T, T' can be obtained from each other by such slides (or by their reverses), we say that they are **jeu de taquin equivalent**, denoted $T \approx T'$. Note that any skew column-strict tableau T is jeu de taquin equivalent to an ordinary CST (called the **rectification** of T); see, e.g., the example (9.45) above. In fact, the rectification is unique; the order in which we choose inner corners does not matter.

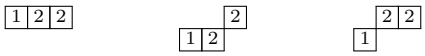
Definition 9.17.2. Let T be a column-strict skew tableau. The **row-reading word** of T, denoted $\mathrm{row}(T)$, is obtained by reading the rows left to right, bottom to top.

For example, the reading words of the skew tableaux in (9.45) are

$$2341214,\ 2342114,\ 2342114,\ 2324114,\ 3224114,\ 3224114,\ 3224114,\ 3224114,\ 3221144.$$

[10] French for "sliding game", roughly; it refers to the 15-square puzzle with sliding tiles that used to come standard on every Macintosh in about 1985.

If T is an ordinary (not skew) tableau, then it is determined by its row-reading word, since the "line breaks" occur exactly at the strict decreases of $\mathrm{row}(T)$. For skew tableaux, this is not the case. Note that some of the slides in (9.45) do not change the row reading word; as a simpler example, the following skew tableaux both have reading word 122:



On the other hand, it's not hard to se that rectifying the second or third tableau will yield the first; therefore, they are all jeu de taquin equivalent.

For a word $\mathbf{b}$ on the alphabet $\mathbb{N}_{>0}$, let $P(\mathbf{b})$ denote its insertion tableau under the RSK algorithm. (That is, construct a generalized permutation $\binom{\mathbf{q}}{\mathbf{b}}$ in which $\mathbf{q}$ is any word; run RSK; and remember only the tableau P, so that the choice of $\mathbf{q}$ does not matter.)

Theorem 9.17.3. *(Knuth–Schützenberger) For two words $\mathbf{b}, \mathbf{b}'$, the following are equivalent:*

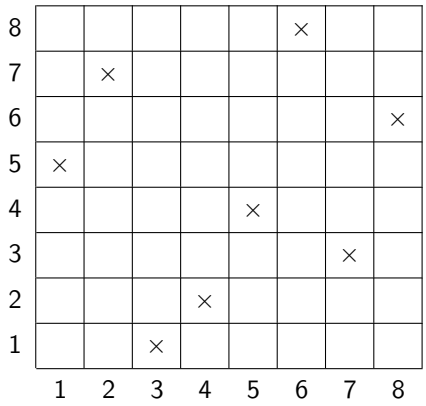
1. $P(\mathbf{b}) = P(\mathbf{b}')$.
2. $\mathbf{b} \sim \mathbf{b}'$.
3. $T \approx T'$, *for any (or all) column-strict skew tableaux T, T' with row-reading words $\mathbf{b}, \mathbf{b}'$ respectively.*

This is sometimes referred to (e.g., in [Ful97]) as the equivalence of "bumping" (the RSK algorithm as presented in Section 9.10) and "sliding" (jeu de taquin).

9.18 Yet another version of RSK

Fix $w \in \mathfrak{S}_n$. Start by drawing an $n \times n$ grid, numbering columns west to east and rows south to north. For each i, place an X in the i-th column and w_i-th row. We are now going to label each of the $(n+1) \times (n+1)$ intersections of the grid lines with a partition, such that the partitions either stay the same or get bigger as we move north and east. We start by labeling each intersection on the west and south sides with the empty partition $\emptyset$.

For instance, if $w = 57214836$, the grid is as follows.



For each box whose SW, SE and NW corners have been labeled λ, μ, ν respectively, label the NE corner ρ according to the following rules:

Rule 1: If $\lambda = \mu = \nu$ and the box doesn't contain an X, then set $\rho = \lambda$.

Rule 2: If $\lambda \subsetneq \mu = \nu$ and the box doesn't contain an X, then it must be the case that $\mu_i = \lambda_i + 1$ for some i. Obtain ρ from μ by incrementing μ_{i+1}.

Rule 3: If $\mu \neq \nu$, then set $\rho = \mu \vee \nu$ (where $\vee$ means the join in Young's lattice: i.e., take the componentwise maximum of the elements of μ and ν).

Rule X: If there is an X in the box, then it must be the case that $\lambda = \mu = \nu$. Obtain ρ from λ by incrementing λ_1.

Note that the underlined assertions need to be proved; this can be done by induction.

Example 9.18.1. Let $n = 8$ and $w = 57214836$. In Example 9.10.2, we found that $\mathsf{RSK}(w) = (P, Q)$, where

$$P = \begin{array}{|c|c|c|}\hline 1 & 3 & 6 \\ \hline 2 & 4 & 8 \\ \hline 5 & 7 \\ \cline{1-2}\end{array} \qquad \text{and} \qquad Q = \begin{array}{|c|c|c|}\hline 1 & 2 & 6 \\ \hline 3 & 5 & 8 \\ \hline 4 & 7 \\ \cline{1-2}\end{array}.$$

The following extremely impressive figure shows what happens when we run the alternate RSK algorithm on w. The partitions λ are shown in red. The numbers in parentheses indicate which rules were used.

0	1 (3)	2 (3)	21 (3)	211 (3)	221 (3)	321 ×	322 (3)	332 (2)
0	1 (3)	2 ×	21 (3)	211 (3)	221 (2)	221 (3)	222 (2)	322 (3)
0	1 (3)	1 (1)	11 (3)	111 (3)	211 (3)	211 (1)	221 (3)	321 ×
0	1 ×	1 (3)	11 (2)	111 (2)	211 (3)	211 (3)	221 (3)	221 (3)
0	0 (1)	0 (1)	1 (3)	11 (3)	21 ×	21 (3)	22 (2)	22 (3)
0	0 (1)	0 (1)	1 (3)	11 (3)	11 (1)	11 (1)	21 ×	21 (3)
0	0 (1)	0 (1)	1 ×	11 (2)	11 (3)	11 (3)	11 (3)	11 (3)
0	0 (1)	0 (1)	0 (1)	1 ×	1 (3)	1 (3)	1 (3)	1 (3)
0	0	0	0	0	0	0	0	0

◀

Observe that:

- Rule 1 is used exactly in those squares that have no X either due west or due south.
- For all squares s, $|\rho|$ is the number of X's in the rectangle whose northeast corner is s. In particular, the easternmost partition $\lambda_{(k)}$ in the kth row, and the northernmost partition $\mu_{(k)}$ in the kth column, both have size k.
- It follows that the sequences
$$\emptyset = \lambda_{(0)} \subseteq \lambda_{(1)} \subseteq \cdots \subseteq \lambda_{(n)},$$
$$\emptyset = \mu_{(0)} \subseteq \mu_{(1)} \subseteq \cdots \subseteq \mu_{(n)}$$
correspond to SYT's of the same shape (in this case 332).
- These SYT's are the P and Q of the RSK correspondence!

9.19 Quasisymmetric functions

Definition 9.19.1. A **quasisymmetric function** is a formal power series $F \in \mathbb{C}[[x_1, x_2, \dots]]$ with the following property: if $i_1 < \cdots < i_r$ and $j_1 < \cdots < j_r$ are two sets of indices in strictly increasing order and $\alpha_1, \dots, \alpha_r \in \mathbb{N}$, then
$$[x_{i_1}^{\alpha_1} \cdots x_{i_r}^{\alpha_r}]F = [x_{j_1}^{\alpha_1} \cdots x_{j_r}^{\alpha_r}]F$$
where $[\mu]F$ denotes the coefficient of μ in F.

Symmetric functions are automatically quasisymmetric, but not vice versa. For example,
$$\sum_{i<j} x_i^2 x_j$$
is quasisymmetric but not symmetric (in fact, it is not preserved by *any* permutation of the variables). On the other hand, the set of quasisymmetric functions forms a graded ring $QSym \subseteq \mathbb{C}[[\mathbf{x}]]$. We now describe a vector space basis for $QSym$.

A **composition** α is a sequence $(\alpha_1, \dots, \alpha_r)$ of positive integers, called its **parts**. Unlike a partition, we do not require that the parts be in weakly decreasing order. If $\alpha_1 + \cdots + \alpha_r = n$, we write $\alpha \models n$; the set of all compositions of n will be denoted $\mathrm{Comp}(n)$. Sorting the parts of a composition in decreasing order produces a partition of n, denoted by $\lambda(\alpha)$.

Compositions are much easier to count than partitions. Consider the set of partial sums
$$S(\alpha) = \{\alpha_1,\ \alpha_1 + \alpha_2,\ \dots,\ \alpha_1 + \cdots + \alpha_{r-1}\}.$$
The map $\alpha \mapsto S(\alpha)$ is a bijection from compositions of n to subsets of $[n-1]$; in particular, $|\mathrm{Comp}(n)| = 2^{n-1}$. We can define a partial order on $\mathrm{Comp}(n)$ via S by setting $\alpha \preceq \beta$ if $S(\alpha) \subseteq S(\beta)$; this is called **refinement**. The covering relations are merging two adjacent parts into one part.

The **monomial quasisymmetric function** of a composition $\alpha = (\alpha_1, \dots, \alpha_r) \models n$ is the power series
$$M_\alpha = \sum_{i_1 < \cdots < i_r} x_{i_1}^{\alpha_1} \cdots x_{i_r}^{\alpha_r} \in \mathbb{Z}[[x_1, x_2, \dots]]_n.$$

For example, the four monomial quasisymmetric functions of degree 4 are
$$M_3 = \sum_i x_i^3 = m_3, \quad M_{21} = \sum_{i<j} x_i^2 x_j, \quad M_{12} = \sum_{i<j} x_i x_j^2, \quad M_{111} = \sum_{i<j<k} x_i x_j x_k = m_{111}.$$

Just as for the monomial symmetric functions, every monomial appears in exactly one M_α, and Definition 9.19.1 says precisely that a power series f is quasisymmetric if all monomials appearing in the same M_α have the same coefficient in f. Therefore, the set $\{M_\alpha\}$ is a graded basis for $QSym$.

Example 9.19.2. Let $\mathcal{M}$ be a matroid on ground set E of size n. Consider weight functions $f : E \to \mathbb{N}_{>0}$; one of the definitions of a matroid (see the problem set) is that a smallest-weight basis of $\mathcal{M}$ can be chosen via the following greedy algorithm (list E in weakly increasing order by weight $e_1, \ldots, e_n$; initialize $B = \emptyset$; for $i = 1, \ldots, n$, if $B + e_i$ is independent, then replace B with $B + e_i$). The **Billera-Jia-Reiner invariant** of $\mathcal{M}$ is the formal power series

$$W(\mathcal{M}) = \sum_f x_{f(1)} x_{f(2)} \cdots x_{f(n)}$$

where the sum runs over all weight functions f for which there is a **unique** smallest-weight basis. The correctness of the greedy algorithm implies that $W(\mathcal{M})$ is quasisymmetric.

For example, let $E = \{e_1, e_2, e_3\}$ and $\mathcal{M} = U_2(3)$. The bases are e_1e_2, e_1e_3, and e_2e_3. Then E has a unique smallest-weight basis iff f has a unique maximum; it doesn't matter if the two smaller weights are equal or not. If the weights are all distinct then they can be assigned to E in $3! = 6$ ways; if the two smaller weights are equal then there are three choices for the heaviest element of E. Thus

$$W(U_2(3)) = \sum_{i<j<k} 6x_ix_jx_k + \sum_{i<j} 3x_ix_j^2 = 6M_{111} + 3M_{12}.$$

Questions: How are $W(\mathcal{M})$ and $W(\mathcal{M}^*)$ related? ◀

9.20 Exercises

Problem 9.1. Suppose $\lambda \vdash n$ and $\mu \vdash m$ are partitions. Then the product $m_\lambda m_\mu$ is a symmetric function of degree $m + n$, so there is a unique expression

$$m_\lambda m_\mu = \sum_{\nu \vdash m+n} a_\nu^{\lambda,\mu} m_\nu$$

where the $a_\nu^{\lambda,\mu}$ are scalars (the **structure coefficients of Λ in the monomial basis**). For example, one can check that

$$m_1m_1 = m_2 + 2m_11, \qquad m_1m_2 = m_3 + m_{21}, \qquad m_1m_{11} = m_{21} + 3m_{111},$$

etc. What can be said about these numbers $a_\nu^{\lambda,\mu}$ in general?

Problem 9.2. In analogy to Corollary 9.4.4, prove that the transition matrix between the bases $\{h_\lambda\}$ and $\{m_\mu\}$ is symmetric.

Problem 9.3. Prove assertion (a) of Proposition 9.9.5.

Problem 9.4. More generally, for two graded bases $\{u_\lambda\}$, $\{v_\mu\}$ of Λ, show how to get the values of $\langle u_\lambda, v_\mu \rangle_\Lambda$ by expanding the Cauchy kernel.

Problem 9.5. Let $\lambda \vdash n$. Verify that $|C_\lambda| = n!/z_\lambda$, where z_λ is defined as in (9.16).

Problem 9.6. Give a purely combinatorial proof that $\exp \log(1 + x) = 1 + x$. In other words, expand the composition $\exp \log x$ as a formal power series, using the definitions of exp and log in (9.19), and compute the coefficient of x^k for each k. Hint: Interpret the coefficients as counting permutations.

Problem 9.7. Supply the proofs for the identities (9.18), i.e.,

$$\Omega^* = \sum_\lambda e_\lambda(\mathbf{x}) m_\lambda(\mathbf{y}) = \sum_\lambda \varepsilon_\lambda \frac{p_\lambda(\mathbf{x}) p_\lambda(\mathbf{y})}{z_\lambda}.$$

Problem 9.8. Prove part (6) of Theorem 9.11.2.

Problem 9.9. Confirm that the Murnaghan–Nakayama rule correctly predicts the values of the trivial, sign, and standard characters on $\mathfrak{S}_n$.

Problem 9.10. Fill in the proofs of the underlined assertions in Rule 2 and Rule X for the alternate RSK algorithm in Section 9.18.

Problem 9.11. For this problem, you will probably want to use one of the alternate RSK algorithms from Sections 9.17 and 9.18.

(a) For $w \in \mathfrak{S}_n$, let $(P(w), Q(w))$ be the pair of tableaux produced by the RSK algorithm from w. Denote by w^* the reversal of w in one-line notation (for instance, if $w = 57214836$ then $w^* = 63841275$). Prove that $P(w^*) = P(w)^T$ (where ^{T} means transpose).

(b) *(Open problem)* For which permutations does $Q(w^*) = Q(w)$? Computation indicates that the number of such permutations is

$$\begin{cases} \dfrac{2^{(n-1)/2}(n-1)!}{((n-1)/2)!^2} & \text{if } n \text{ is odd,} \\ 0 & \text{if } n \text{ is even,} \end{cases}$$

but I don't know a combinatorial (or even an algebraic) reason.

(c) *(Open problem)* For which permutations does $Q(w^*) = Q(w)^T$? I have no idea what the answer is. The sequence $(q_1, q_2, \dots) = (1, 2, 2, 12, 24, 136, 344, 2872, 7108, \dots)$, where $q_n = \#\{w \in \mathfrak{S}_n : Q(w^*) = Q(w)^T\}$, does not seem to appear in the Online Encyclopedia of Integer Sequences.

Problem 9.12. Let $G = (V, E)$ be a finite simple graph with vertex set V. Let $C(G)$ denote the set of **proper colorings** of G: functions $\kappa : V \to \mathbb{N}_{>0}$ such that $\kappa(v) \neq \kappa(w)$ whenever v, w are adjacent in G. Define a formal power series in indeterminates $x_1, x_2, \dots$, by

$$X_G = \sum_{\kappa \in C(G)} \underbrace{\prod_{v \in V} x_{\kappa(v)}}_{\mathbf{x}^\kappa}.$$

(a) Show that X_G is a symmetric function (this is not too hard). It is known as the **chromatic symmetric function**, and was introduced by Stanley [Sta95];

(b) Determine X_G for (i) K_n; (ii) $\overline{K_n}$ (i.e., the graph with n vertices and no edges); (iii) the four simple graphs on 3 vertices; (iv) the two trees on 4 vertices.

(c) Explain how to recover the chromatic polynomial $p_G(k)$ (see Example 2.3.5) from X_G. Does $p_G(k)$ determine X_G?

(d) For a set $A \subseteq E$, let $\lambda(A)$ denote the partition whose parts are the sizes of the components of the subgraph $G|_A$ induced by A (so $\lambda \vdash |V(G)|$ and $\ell(\lambda)$ is the number of components). Prove [Sta95, Thm. 2.5] that the expansion of X_G in the power-sum basis is

$$X_G = \sum_{A \subseteq E} (-1)^{|A|} p_{\lambda(A)}.$$

(Hint: Use inclusion/exclusion.)

(e) (Unsolved) Do there exist two non-isomorphic trees T, U with $X_T = X_U$? (This problem is the biggest bête noire of the author of these notes, having troubled his sleep since approximately 2005.)

Chapter 10

Combinatorial Hopf Theory

For many combinatorial structures, there is a natural way of taking apart one object into two, or combining two objects into one.

- Let $G = (V, E)$ be a (simple, undirected) graph. For any $W \subseteq V$, we can break G into the two pieces $G|_W$ and $G|_{V \setminus W}$. On the other hand, given two graphs, we can form their disjoint union $G \uplus H$.
- Let M be a matroid on ground set E. For any $A \subseteq E$, we can break M into the restriction $M|_A$ (equivalently, the deletion of $E \setminus A$) and the contraction M/A. Two matroids can be combined into one by taking the direct sum.
- Let P be a ranked poset. For any $x \in P$, we can extract the intervals $[\hat{0}, x]$ and $[x, \hat{1}]$. (Of course, we don't get every element of the poset this way.) Meanwhile, two graded posets P, Q can be combined into one poset in many ways, such as Cartesian product (see Definition 1.1.13).
- Let $\alpha = (\alpha_1, \ldots, \alpha_\ell) \models n$. For $0 \leq k \leq \ell$, we can break α up into two sub-compositions $\alpha_{(k)} = (\alpha_1, \ldots, \alpha_k)$, $\alpha^{(k)} = (\alpha_{k+1}, \ldots, \alpha_\ell)$. Of course, two compositions can be combined by concatenating them.

In all these operations, there are lots of ways to split, but only one way to combine. Moreover, all the operations are graded with respect to natural size functions on the objects: for instance, matroid direct sum is additive on size of ground set and on rank.

Splitting	**Combining**
$\lvert V(G\lvert_W)\rvert + \lvert V(G\lvert_{V\setminus W})\rvert = \lvert V(G)\rvert$	$\lvert V(G \uplus H)\rvert = \lvert V(G)\rvert + \lvert V(H)\rvert$
$\lvert E(M\lvert_A)\rvert + \lvert E(M/A)\rvert = \lvert E(M)$	$\lvert E(M \oplus M')\rvert = \lvert E(M)\rvert + E(M')\rvert$
$r([\hat{0}, x]) + r([x, \hat{1}]) = r(P)$	$r(P \oplus Q) = r(P) + r(Q)$
$\lvert\alpha_{(k)}\rvert + \lvert\alpha^{(k)}\rvert = \lvert\alpha\rvert$	$r(\alpha\beta) = r(\alpha) + r(\beta)$

10.1 Hopf algebras

A **Hopf algebra** is a vector space $\mathcal{H}$ (over $\mathbb{C}$, say) with two additional operations, a **product** $\mu : \mathcal{H} \otimes \mathcal{H} \to \mathcal{H}$ (which represents combining) and a **coproduct** $\Delta : \mathcal{H} \to \mathcal{H} \otimes \mathcal{H}$ which represents splitting. These operations

are respectively **associative** and **coassociative**, and they are compatible in a certain way. Technically, all this data defines the slightly weaker structure of a **bialgebra**; a Hopf algebra is a bialgebra with an additional map $\mathsf{S} : \mathcal{H} \to \mathcal{H}$, called the **antipode**. Most bialgebras that arise in combinatorics have a unique antipode and thus a unique Hopf structure.

What is a $\mathbb{C}$-algebra? It is a $\mathbb{C}$-vector space A equipped with a ring structure. Its multiplication can be thought of as a $\mathbb{C}$-bilinear map

$$\mu : A \otimes A \to A$$

that is associative, i.e., $\mu(\mu(a,b),c) = \mu(a,\mu(b,c))$. Associativity can be expressed as the commutativity of the diagram

$$\begin{array}{ccc} A \otimes A \otimes A & \xrightarrow{\mu \otimes \mathrm{Id}} & A \otimes A \\ \downarrow{\scriptstyle \mathrm{Id} \otimes \mu} & & \downarrow{\scriptstyle \mu} \\ A \otimes A & \xrightarrow{\mu} & A \end{array} \qquad \begin{array}{ccc} a \otimes b \otimes c & \longmapsto & ab \otimes c \\ \downarrow & & \downarrow \\ a \otimes bc & \longmapsto & abc \end{array} \tag{10.1}$$

where I denotes the identity map. (Diagrams like this rely on the reader to interpret notation such as $\mu \otimes I$ as the only thing it could be possibly be; in this case, "apply μ to the first two tensor factors and tensor what you get with [I applied to] the third tensor factor".)

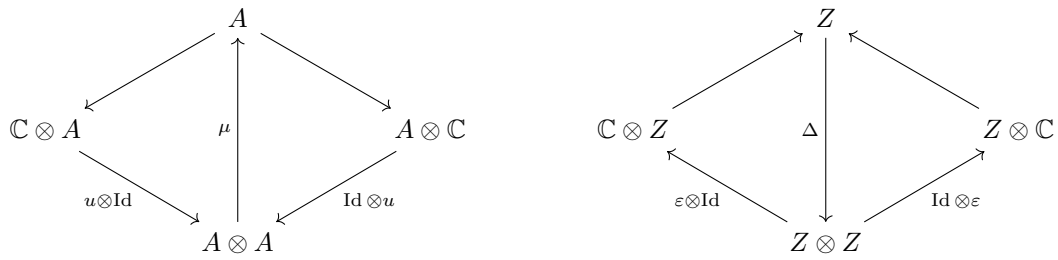
What then is a $\mathbb{C}$-**coalgebra**? It is a $\mathbb{C}$-vector space Z equipped with a $\mathbb{C}$-linear **comultiplication** map

$$\Delta : Z \to Z \otimes Z$$

that is **coassociative**, a condition defined by reversing the arrows in the previous diagram:

$$\begin{array}{ccc} Z \otimes Z \otimes Z & \xleftarrow{\Delta \otimes \mathrm{Id}} & Z \otimes Z \\ \uparrow{\scriptstyle \mathrm{Id} \otimes \Delta} & & \uparrow{\scriptstyle \Delta} \\ Z \otimes Z & \xleftarrow{\Delta} & Z \end{array} \tag{10.2}$$

Just as an algebra has a unit, a coalgebra has a **counit**. To say what this is, let us diagramify the defining property of the multiplicative unit 1_A in an algebra A: it is the image of $1_{\mathbb{C}}$ under a map $u : \mathbb{C} \to A$ such that the diagram on the left commutes (where the top diagonal maps take $a \in A$ to $1 \otimes a$ or $a \otimes 1$). Thus a counit of a coalgebra is a map $\varepsilon : Z \to \mathbb{C}$ such that the diagram on the right commutes (where the top diagonal maps are projections).



(10.3)

A **bialgebra** is a vector space B that has both a multiplication and a comultiplication, and such that multiplication is a coalgebra morphism and comultiplication is an algebra morphism. Both of these conditions

are expressible by commutativity of the diagram

$$\begin{array}{ccc} B \otimes B & \xrightarrow{\Delta\otimes\Delta} & B\otimes B\otimes B\otimes B \\ \downarrow{\scriptstyle \mu} & & \downarrow{\scriptstyle \mu_{13}\otimes\mu_{24}} \\ B & \xrightarrow{\Delta} & B\otimes B \end{array} \tag{10.4}$$

where $\mu_{13} \otimes \mu_{24}$ means the map that sends $a \otimes b \otimes c \otimes d$ to $ac \otimes bd$ (the subscripts refer to the positions of the tensor factors).

Comultiplication takes some getting used to. As explained above, in combinatorial settings, one should generally think of multiplication as putting two objects together, and comultiplication as taking an object apart into two subobjects. A unit is a trivial object (putting it together with another object has no effect), and the counit is the linear functional that picks off the coefficient of the unit.

Example 10.1.1 (The polynomial Hopf algebra)**.** A simple example of a Hopf algebra is the polynomial ring $\mathbb{C}[x]$. It is an algebra in the usual way, and can be made into a coalgebra by the counit $\varepsilon(f(x)) = f(0)$ (equivalently, mapping every polynomial to its constant term) and the coproduct $\Delta(x) = 1 \otimes x + x \otimes 1$. Checking the bialgebra axioms is left as an exercise. ◀

Example 10.1.2 (The graph Hopf algebra)**.** For $n \geq 0$, let $\mathcal{G}_n$ be the set of formal $\mathbb{C}$-linear combinations of unlabeled simple graphs on n vertices (or if you prefer, of isomorphism classes $[G]$ of simple graphs G, but it is easier to drop the brackets), and let $\mathcal{G} = \bigoplus_{n\geq 0} \mathcal{G}_n$. Thus $\mathcal{G}$ is a graded vector space, which we make into a $\mathbb{C}$-algebra by defining $\mu(G \otimes H) = G \cup H$, where $\cup$ denotes union under the assumption $V(G) \cap V(H) = \emptyset$.The unit is the unique graph K_0 with no vertices (or, technically, the map $u : \mathbb{C} \to \mathcal{G}_0$ sending $c \in \mathbb{C}$ to cK_0). Comultiplication in $\mathcal{G}$ is defined by

$$\Delta(G) = \sum_{A,B:\ V(G)=A\cup B} G|_A \otimes G|_B.$$

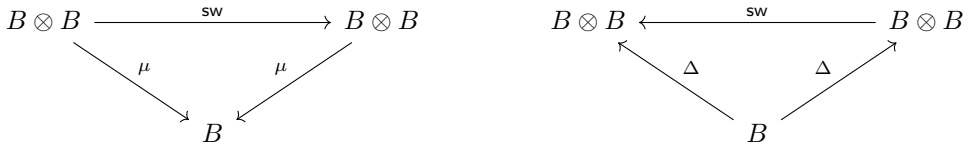
As an illustration of how the compatibility condition (10.4) works, we will check it for $\mathcal{G}$. To avoid "overfull hbox" errors, set $\tilde{\mu} = \mu_{13} \otimes \mu_{24}$. Then

$$\begin{aligned} \tilde{\mu}(\Delta \otimes \Delta(G_1 \otimes G_2)) &= \tilde{\mu}\left(\left(\sum_{A_1\cup B_1=V(G_1)} G_1|_{A_1} \otimes G_1|_{B_1}\right) \otimes \left(\sum_{A_2\cup B_2=V(G_2)} G_2|_{A_2} \otimes G_2|_{B_2}\right)\right) \\ &= \tilde{\mu}\left(\sum_{\substack{A_1\cup B_1=V(G_1)\\ A_2\cup B_2=V(G_2)}} G_1|_{A_1} \otimes G_1|_{B_1} \otimes G_2|_{A_2} \otimes G_2|_{B_2}\right) \\ &= \sum_{\substack{A_1\cup B_1=V(G_1)\\ A_2\cup B_2=V(G_2)}} (G_1|_{A_1} \cup G_2|_{A_2}) \otimes (G_1|_{B_1} \cup G_2|_{B_2}) \\ &= \sum_{A\cup B=V(G_1\cup G_2)} (G_1 \cup G_2)|_A \otimes (G_1 \cup G_2)|_B \\ &= \Delta(\mu(G_1 \otimes G_2)). \end{aligned}$$

Comultiplication in $\mathcal{G}$ is in fact **cocommutative**[1]. Let sw be the "switching map" that sends $a\otimes b$ to $b\otimes a$; then commutativity and cocommutativity of multiplication and comultiplication on a bialgebra B are expressed

[1]There are those who call this "mmutative".

by the diagrams



So cocommutativity means that $\Delta(G)$ is symmetric under switching; for the graph algebra this is clear because A and B are interchangeable in the definition. ◀

Example 10.1.3 (Rota's Hopf algebra of posets)**.** For $n \geq 0$, let $\mathcal{P}_n$ be the vector space of formal $\mathbb{C}$-linear combinations of isomorphism classes $[P]$ of finite graded posets P of rank n. Thus $\mathcal{P}_0$ and $\mathcal{P}_1$ are one-dimensional (generated by the chains of lengths 0 and 1), but $\dim \mathcal{P}_n = \infty$ for $n \geq 2$. We make $\mathcal{P} = \bigoplus_n \mathcal{P}_n$ into a graded $\mathbb{C}$-algebra by defining $\mu([P] \otimes [Q]) = [P \times Q]$, where $\times$ denotes Cartesian product; thus $u(1) = \bullet$. Comultiplication is defined by

$$\Delta[P] = \sum_{x \in P} [\hat{\mathbf{0}}, x] \otimes [x, \hat{\mathbf{1}}].$$

Coassociativity is checked by the following calculation, which should remind you of the proof of associativity of convolution in the incidence algebra of a poset (Prop. 2.1.2):

$$\begin{aligned}
\Delta \otimes I(\Delta(P)) &= \Delta \otimes I \left(\sum_{x \in P} [\hat{\mathbf{0}}, x] \otimes [x, \hat{\mathbf{1}}] \right) \\
&= \sum_{x \in P} \Delta([\hat{\mathbf{0}}, x]) \otimes [x, \hat{\mathbf{1}}] \\
&= \sum_{x \in P} \left(\sum_{y \in [\hat{\mathbf{0}}, x]} [\hat{\mathbf{0}}, y] \otimes [y, x] \right) \otimes [x, \hat{\mathbf{1}}] \\
&= \sum_{x \leq y \in P} [\hat{\mathbf{0}}, y] \otimes [y, x] \otimes [x, \hat{\mathbf{1}}] \\
&= \sum_{y \in P} [\hat{\mathbf{0}}, y] \otimes \left(\sum_{x \in [y, \hat{\mathbf{1}}]} [y, x] \otimes [x, \hat{\mathbf{1}}] \right) \\
&= \sum_{y \in P} [\hat{\mathbf{0}}, y] \otimes \Delta([y, \hat{\mathbf{1}}]) \;=\; I \otimes \Delta(\Delta(P)).
\end{aligned}$$

This Hopf algebra is commutative, but not cocommutative; the switching map does not fix $\Delta(P)$ unless P is self-dual. ◀

Example 10.1.4 (The Hopf algebra of matroids)**.** For $n \geq 0$, let $\mathcal{M}_n$ be the vector space of formal $\mathbb{C}$-linear combinations of isomorphism classes $[M]$ of finite matroids M on n elements. Here $\dim \mathcal{P}_0 = 1$ and $\dim \mathcal{P}_n < \infty$ for every n. We make $\mathcal{M} = \bigoplus_n \mathcal{M}_n$ into a graded $\mathbb{C}$-algebra by defining $\mu([P] \otimes [Q]) = [P \otimes Q]$. The trivial matroid (with empty ground set) is the multiplicative identity. Note that multiplication is commutative. Letting E denote the ground set of M, we define comultiplication by

$$\Delta[M] = \sum_{A \subseteq E} M|_A \otimes M/A.$$

Coassociativity is essentially a consequence of the compatibility of deletion and contraction (Prop. 3.8.2). Note that the coproduct is not cocommutative. ◀

This is a good place to introduce what is known as **Sweedler notation**. Often, it is highly awkward to notate all the summands in a coproduct, particularly if we are trying to prove general facts about Hopf algebra. The Sweedler notation for a coproduct is

$$\Delta(h) = \sum h_1 \otimes h_2$$

which should be read as "the coproduct of h is a sum of a bunch of tensors, each of which has a first element and a second element." This notation looks dreadfully abusive at first, but in fact it is incredibly convenient, is unambiguous if used properly, and one soon discovers that any other way of doing things would be worse (imagine having to conjure an index set out of thin air and deal with a lot of double subscripts just to write down a coproduct). Sweedler notation iterates well; for example, we could write

$$\Delta^2(h) = (\mathrm{Id} \otimes \Delta)(\Delta(h)) = (\Delta \otimes \mathrm{Id})(\Delta(h)) = \sum h_1 \otimes h_2 \otimes h_3$$

(cf. (10.2), which gives the second equality).

Example 10.1.5. The ring Λ of symmetric functions is a coalgebra in the following way. Recall that Λ is a subring of the ring of formal power series $\mathbb{C}[[\mathbf{x}]] = \mathbb{C}[[x_1, x_2, \dots]]$. First, the counit is just the map that takes a formal power series to its constant term. To define the coproduct of $F \in \Lambda$, we first apply the "Hilbert Hotel substitution": replace $x_1, x_2, x_3, x_4, \dots$ with $x_1, y_1, x_2, y_2, \dots$ to obtain a power series $F(\mathbf{x}, \mathbf{y}) \in \mathbb{C}[[\mathbf{x}, \mathbf{y}]] = \mathbb{C}[[\mathbf{x}]] \otimes \mathbb{C}[[\mathbf{y}]]$. This power series is symmetric in each of the variable sets $\mathbf{x}$ and $\mathbf{y}$, i.e.,

$$\Lambda(\mathbf{x}, \mathbf{y}) \subseteq \Lambda(\mathbf{x}) \otimes \Lambda(\mathbf{y}).$$

So every symmetric function $F(\mathbf{x}, \mathbf{y})$ can be written (uniquely) in the form $\sum F_1(\mathbf{x})F_2(\mathbf{y})$ (in Sweedler notation). We then define $\Delta(F) = \sum F_1 \otimes F_2$.

For example, clearly $\Delta(c) = c = c \otimes 1 = 1 \otimes c$ for any scalar c. Moreover, for every k, we have

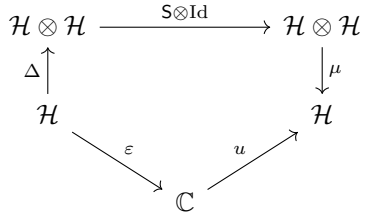
$$h_k(\mathbf{x}, \mathbf{y}) = \sum_{j=0}^{k} h_j(\mathbf{x}) h_{k-j}(\mathbf{y}), \qquad e_k(\mathbf{x}, \mathbf{y}) = \sum_{j=0}^{k} e_j(\mathbf{x}) e_{k-j}(\mathbf{y})$$

and therefore

$$\Delta(h_k) = \sum_{j=0}^{k} h_j \otimes h_{k-j}, \qquad \Delta(e_k) = \sum_{j=0}^{k} e_j \otimes e_{k-j}.$$

◀

Definition 10.1.6. A Hopf algebra is a bialgebra $\mathcal{H}$ with a **antipode** $\mathsf{S} : \mathcal{H} \to \mathcal{H}$, which satisfies the commutative diagram



(10.5)

In other words, to calculate the antipode of something, comultiply it to get $\Delta g = \sum g_1 \otimes g_2$. Now hit every first tensor factor with S and then multiply it out again to obtain $\sum \mathsf{S}(g_1) \cdot g_2$. If you started with the unit

then this should be 1, while if you started with any other homogeneous object then you get 0. This enables calculating the antipode recursively. For example, in QSym:

$$\begin{aligned}\mu(\mathsf{S}\otimes I(\Delta 1)) = \mu(\mathsf{S}\otimes I(1\otimes 1)) &= \mu(\mathsf{S}(1)\otimes 1) = \mathsf{S}(1)\\ u(\varepsilon(1)) &= 1\\ \mathsf{S}(1) &= 1\end{aligned}$$

$$\begin{aligned}\mu((\mathsf{S}\otimes I)(\Delta h_1)) = \mu((\mathsf{S}\otimes I)(h_1\otimes 1 + 1\otimes h_1)) &= \mu(\mathsf{S}(h_1)\otimes 1 + \mathsf{S}(1)\otimes h_1) = \mathsf{S}(h_1) + h_1\\ u(\varepsilon(h_1)) &= 0\\ \mathsf{S}(h_1) &= -h_1\\ &\cdots\end{aligned}$$

Lemma 10.1.7 (Humpert, Prop 1.4.4)**.** *Let B be a bialgebra that is graded and **connected**, i.e., the 0th graded piece has dimension 1 as a vector space. Let $n > 0$ and let $h \in \mathcal{H}_n$. Then*

$$\Delta(h) = h\otimes 1 + \sum h_1\otimes h_2 + 1\otimes h$$

where the Sweedler-notation sum contains only elements of degrees strictly between 0 and n.

Proof. Refer to the diagrams for the unit and counit (10.3). In particular, the right-hand triangle gives $\sum h_1\otimes\varepsilon(h_2) = h$. So certainly one of the summands must have $h_1\in\mathcal{H}_n$, but then $h_2\in\mathcal{H}_0$. Since $\mathcal{H}_0\cong\mathbb{C}$ we may as well group all those summands together; they must sum to $h\otimes 1$. Meanwhile, the left-hand triangle says that grouping together all the summands of bidegree $0, n$ gives $1\otimes h$. □

Proposition 10.1.8. *Let B be a connected and graded bialgebra. Then the commutative diagram (10.5) defines a unique antipode $\mathsf{S}: B\to B$, and thus B can be made into a Hopf algebra in a unique way.*

Combinatorics features lots of graded connected bialgebras (such as all those we have seen so far), so this proposition gives us a Hopf algebra structure "for free".

There is a general recipe for the antipode, known as *Takeuchi's formula* [Tak71]. Let $\pi:\mathcal{H}\to\mathcal{H}$ be the map that kills $\mathcal{H}_0$ and fixes each positive graded piece pointwise. Then

$$\mathsf{S} = u\varepsilon + \sum_{k\geq 1}(-1)^k\mu^{k-1}\pi^{\otimes k}\Delta^{k-1}, \tag{10.6}$$

i.e.,

$$\mathsf{S}(h) = u(\varepsilon(h)) - \pi(h) + \sum\pi(h_1)\pi(h_2) - \sum\pi(h_1)\pi(h_2)\pi(h_3) + \cdots$$

However, there is a lot of cancellation in this sum, making it impractical for looking at specific Hopf algebras. Therefore, one of the first things one wants in studying a particular Hopf algebra is to find a cleaner formula for the antipode. An excellent example is the Hopf algebra of symmetric functions, in which the antipode is the involution ω interchanging e_λ and h_λ (Proposition 9.5.3). The proof is left as an exercise (Problem 10.3).

10.2 Characters

A **character** on a Hopf algebra $\mathcal{H}$ is a $\mathbb{C}$-linear map $\zeta : \mathcal{H} \to \mathbb{C}$ that is multiplicative, i.e., $\zeta(1_{\mathcal{H}}) = 1_{\mathbb{C}}$ and $\zeta(h \cdot h') = \zeta(h)\zeta(h')$. For example, if $\mathcal{H}$ is the graph Hopf algebra, then we can define a character by

$$\zeta(G) = \begin{cases} 1 & \text{if } G \text{ has no edges,} \\ 0 & \text{if } G \text{ has one or more edges,} \end{cases} \tag{10.7}$$

for a graph G, and then extending by linearity to all of $\mathcal{G}$. This map is multiplicative (because $G \cdot H$ has an edge iff either G or H does); it also looks kind of like a silly map. However, the reason this is interesting is that characters can be multiplied together. The multiplication is called **convolution product**, defined as follows: if $h \in \mathcal{H}$ and $\Delta(h) = \sum h_1 \otimes h_2$ in Sweedler notation, then

$$(\zeta * \eta)(h) = \sum \zeta(h_1)\eta(h_2).$$

One can check that convolution is associative; the calculation resembles checking that the incidence algebra of a poset is an algebra. The counit ε is a two-sided identity for convolution, i.e., $\zeta * \varepsilon = \varepsilon * \zeta = \zeta$ for all characters ζ. Moreover, the definition (10.5) of the antipode implies that

$$\zeta * (\zeta \circ \mathsf{S}) = \varepsilon$$

(check this too). Therefore, the set of all characters forms a group.

Why would you want to convolve characters? Consider the graph Hopf algebra with the character ζ, and let $k \in \mathbb{N}$. The kth convolution power of ζ is given by

$$\begin{aligned} \zeta^k(G) = \underbrace{\zeta * \cdots * \zeta}_{k \text{ times}}(G) &= \sum_{V(G)=V_1 \cup \cdots \cup V_k} \zeta(G|_{V_1}) \cdots \zeta(G|_{V_k}) \\ &= \sum_{V(G)=V_1 \cup \cdots \cup V_k} \begin{cases} 1 & \text{if } V_1, \ldots, V_k \text{ are all cocliques,} \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

(recall that a **coclique** is a set of vertices of which no two are adjacent). In other words, $\zeta^n(G)$ counts the number of functions $f : V \to [k]$ so that $f(x) \neq f(y)$ whenever x, y are adjacent. But such a thing is precisely a proper k-coloring! I.e.,

$$\zeta^n(G) = p(G; k)$$

where p is the chromatic polynomial (see Section 4.4). This turns out to be true as a polynomial identity in k — for instance, $\zeta^{-1}(G)$ is the number of acyclic orientations. One can even view the Tutte polynomial $T(G; x, y)$ as a character $\tau_{x,y}(G)$ with parameters x, y; it turns out that $\tau^k_{x,y}(G)$ is itself a Tutte polynomial evaluation — see Brandon Humpert's Ph.D. thesis [Hum11].

A **combinatorial Hopf algebra**, or CHA, is a pair $(\mathcal{H}, \zeta)$, where $\mathcal{H}$ is a graded connected Hopf algebra and ζ is a character. A **morphism** of CHA's is a map $(\mathcal{H}, \zeta) \xrightarrow{\Phi} (\mathcal{H}', \zeta')$ that is an algebra and coalgebra morphism and satisfies $\zeta' \circ \Phi = \Phi \circ \zeta'$.

Example 10.2.1. The **binomial Hopf algebra** is the ring of polynomials $\mathbb{C}[x]$, equipped with the coproduct generated by $\Delta(x) = x \otimes 1 + 1 \otimes x$. To justify the name, note that

$$\Delta(x^n) = \Delta(x)^n = (x \otimes 1 + 1 \otimes x)^n = \sum_{k=0}^{n} \binom{n}{k} x^k \otimes x^{n-k}.$$

This is extended linearly, so that $\Delta(f(x)) = f(\Delta(x))$ for any polynomial f. The counit is $\varepsilon(f) = f(0)$, and the antipode is given by $\mathsf{S}(x^k) = (-1)^k x^k$ (check this). We make it into a CHA by endowing it with the character $\varepsilon_1(f) = f(1)$.

For any CHA $(\mathcal{H}, \zeta)$, there is then a canonical morphism

$$P_\zeta : (\mathcal{H}, \zeta) \to (\mathbb{C}[x], \varepsilon_1)$$

which maps $h \in \mathcal{H}$ to the unique polynomial $P_(\zeta, h)(x)$ satisfying

$$P_{\zeta,h}(x) = \zeta^k(h) \qquad \forall k \in \mathbb{Z}.$$

For example, if $\mathcal{H}$ is the graph algebra and ζ the characteristic function of edgeless graphs (10.7), then P_ζ is the chromatic polynomial. ◀

Example 10.2.2. The ring $QSym$ of quasisymmetric functions can be made into a Hopf algebra as follows. Let $\alpha = (\alpha_1, \ldots, \alpha_k)$ be a composition; then

$$\Delta M_\alpha = \sum_{j=0}^{k} M_{(\alpha_1,\ldots,\alpha_j)} \otimes M_{(\alpha_{j+1},\ldots,\alpha_k)}.$$

One can check (Problem 10.3) that the Hopf algebra of symmetric functions described in Example 10.1.5 is a Hopf subalgebra of $QSym$; that is, this coproduct restricts to the one defined earlier on Λ. We then endow $QSym$ with the character ζ_Q defined on the level of power series by $\zeta_Q(x_1) = 1$ and $\zeta_Q(x_j) = 0$ for $j \geq 2$; equivalently,

$$\zeta_Q(M_\alpha) = \begin{cases} 1 & \text{if } \alpha \text{ has at most one part,} \\ 0 \text{ otherwise.} \end{cases}$$

One of the main theorems about CHAs, due to Aguiar, Bergeron and Sottile [ABS06], is that $(QSym, \zeta_Q)$ is a **terminal object in the category of CHAs**, i.e., every CHA $(\mathcal{H}, \zeta)$ admits a canonical morphism to $(QSym, \zeta)$. For the graph algebra, this morphism is the chromatic symmetric function; for the matroid algebra, it is the Billera-Jia-Reiner invariant. ◀

10.3 Hopf monoids

Hopf monoids are a more recent area of research. One exhaustive reference is the book by Aguiar and Mahajan [AM10]; more accessible introductions (and the main sources for these notes) include Klivans' talk slides [Kli] and the preprint by Aguiar and Ardila [AA17]. One of the ideas behind Hopf monoids is to work with labeled rather than unlabeled objects.

So, a **Hopf monoid** H consists of the following data.

First, we need a set $\mathsf{H}[I]$ for every finite set I. One should think of $\mathsf{H}[I]$ as the vector space spanned by combinatorial objects of a certain ilk, with I as the labeling set. (For example, graphs with vertices I, matroids with ground set I, linear orderings of I, polyhedra in $\mathbb{R}^I$, etc.) Every bijection $\pi : I \to I'$ should induce a linear isomorphism $\mathsf{H}[\pi] : \mathsf{H}[I] \to \mathsf{H}[I']$, which should be thought of as relabeling, and the association of $\mathsf{H}[\pi]$ with π is functorial[2]. A functor H with these properties is called a **vector species**. Moreover, we require that $\dim \mathsf{H}[\emptyset] = 1$, and we identify a particular nonzero element of $\mathsf{H}[\emptyset]$ as the "trivial object".

[2]This is a fancy way of saying that it obeys some completely natural identities: $\mathsf{H}[\mathrm{Id}_I] = \mathrm{Id}_{\mathsf{H}[I]}$ and $\mathsf{H}[\pi \circ \sigma] = \mathsf{H}[\pi] \circ \mathsf{H}[\sigma]$. Don't worry too much about it.

Then, we need to have multiplication and comultiplication maps for every decomposition $I = A \sqcup B$:

$$\mathsf{H}[A] \otimes \mathsf{H}[B] \xrightarrow{\mu_{A,B}} \mathsf{H}[I] \qquad \text{and} \qquad \mathsf{H}[I] \xrightarrow{\Delta_{A,B}} \mathsf{H}[A] \otimes \mathsf{H}[B]. \tag{10.8}$$

These are subject to a whole lot of conditions. The most important of these are labeled versions of associativity, coassociativity, and compatibility:

$$\begin{array}{ccc} \mathsf{H}[I] \otimes \mathsf{H}[J] \otimes \mathsf{H}[K] & \xrightarrow{\mu_{I,J} \otimes \mathrm{Id}_K} & \mathsf{H}[I] \otimes \mathsf{H}[J \sqcup K] \\ \downarrow{\scriptstyle \mathrm{Id}_I \otimes \mu_{J,K}} & & \downarrow{\scriptstyle \mathrm{Id}_I \otimes \mu_{I, J \sqcup K}} \\ \mathsf{H}[I \sqcup J] \otimes \mathsf{H}[K] & \xrightarrow{\mu_{I \sqcup J, K}} & \mathsf{H}[I \sqcup J \sqcup K] \end{array} \qquad \text{(associativity)}, \tag{10.9}$$

$$\begin{array}{ccc} \mathsf{H}[I] \otimes \mathsf{H}[J] \otimes \mathsf{H}[K] & \xleftarrow{\Delta_{I,J} \otimes \mathrm{Id}_K} & \mathsf{H}[I] \otimes \mathsf{H}[J \sqcup K] \\ \uparrow{\scriptstyle \mathrm{Id}_I \otimes \Delta_{J,K}} & & \uparrow{\scriptstyle \Delta_{I, J \sqcup K}} \\ \mathsf{H}[I \sqcup J] \otimes \mathsf{H}[K] & \xleftarrow{\Delta_{I \sqcup J, K}} & \mathsf{H}[I \sqcup J \sqcup K] \end{array} \qquad \text{(coassociativity)}, \tag{10.10}$$

$$\begin{array}{ccc} \mathsf{H}[I \sqcup J] \otimes \mathsf{H}[K \sqcup L] & \xrightarrow{\Delta_{I,J} \otimes \Delta_{K,L}} & \mathsf{H}[I] \otimes \mathsf{H}[J] \otimes \mathsf{H}[K] \otimes \mathsf{H}[L] \\ \downarrow{\scriptstyle \mu_{I \sqcup J, K \sqcup L}} & & \downarrow{\scriptstyle (\mu_{I,K} \otimes \mu_{J,L}) \circ \tau} \\ \mathsf{H}[I \sqcup J \sqcup K \sqcup L] & \xrightarrow{\Delta_{I \sqcup K, J \sqcup L}} & \mathsf{H}[I \sqcup K] \otimes \mathsf{H}[J \sqcup L] \end{array} \qquad \text{(compatibility)}, \tag{10.11}$$

where τ interchanges the second and third tensor factors.

Note that instead of defining a single coproduct as the sum over all possible decompositions A, B (as in the Hopf algebra setup), we are keeping the different decompositions separate.

In many cases, the operations can be defined on the level of individual combinatorial objects. In other words, we start with a **set species** h — a collection of *sets* $\mathsf{h}[I]$ indexed by finite sets I, subject to the conditions that any bijection $I \to I'$ naturally induces a bijection $\mathsf{h}[I] \to \mathsf{h}[I']$, define multiplication and comultiplication operations

$$\mathsf{h}[A] \times \mathsf{h}[B] \xrightarrow{\mu_{A,B}} \mathsf{h}[I] \qquad \text{and} \qquad \mathsf{h}[I] \xrightarrow{\Delta_{A,B}} \mathsf{h}[A] \times \mathsf{h}[B]$$

(in contrast to eqrefvector-species-product-coproduct, these are Cartesian products of sets rather than tensor products of vector spaces), then define a vector species H by setting $\mathsf{H}[I] = \Bbbk\mathsf{h}[I]$, and define multiplication and comultiplication on H by linear extension. Such a Hopf monoid is called **linearized**. This is certainly a very natural kind of Hopf monoid, but not all the Hopf monoids we care about come from a set species in this way.

Example 10.3.1. Let $\boldsymbol{\ell}[I]$ denote the set of linear orders on a finite set I, which we can think of as bijections $w : [n] \to I$ (and represent by the sequence $w(1), \ldots, w(n)$). Given a decomposition $I = A \sqcup B$, the most obvious way to define product and coproduct on the set species $\boldsymbol{\ell}$ is by concatenation and restriction. For instance, if $A = \{a, b, c\}$ and $B = \{p, q, r, s\}$, then

$$\mu_{A,B}(bac, prsq) = bacprsq, \qquad \Delta_{A,B}(arscqbp) = (cab, rsqp).$$

Linearizing this setup produces the **Hopf monoid of linear orders** $\mathsf{L} = \Bbbk\boldsymbol{\ell}$. ◀

Example 10.3.2. Let $\mathsf{m}[I]$ denote the set of matroids with ground set I, with product and coproduct defined setwise by

$$\mu(M_1, M_2) = M_1 \oplus M_2, \qquad \Delta_{A,B}(M) = (M|_A,\ M/A).$$

The linearized Hopf monoid $\mathsf{M} = \Bbbk\mathsf{m}$ is a labeled analogue of the matroid Hopf algebra $\mathcal{M}$ described in Example 10.1.4. ◀

Multiplication and comultiplication can be iterated. For any set composition A (i.e., an ordered list $A = A_1|\dots|A_n$ whose disjoint union is I), there are maps

$$\bigotimes_{i=1}^{n} \mathsf{H}[A_i] \xrightarrow{\mu_A} \mathsf{H}[I] \qquad \text{and} \qquad \mathsf{H}[I] \xrightarrow{\Delta_A} \bigotimes_{i=1}^{n} \mathsf{H}[A_i]$$

that are well defined by associativity and coassociativity. (For set species, replace tensor products with Cartesian products.) For example, if $A = (I, J, K)$ then we can define μ_A by either traveling south then east, or east then south, in (10.9) — we get the same answer in both cases.

The **antipode** in a Hopf monoid H is the following collection of maps $\mathsf{S}_I : \mathsf{H}[I] \to \mathsf{H}[I]$ given by the *Takeuchi formula*: for $x \in \mathsf{H}[I]$,

$$\mathsf{S}(x) = \mathsf{S}_I(x) = \begin{cases} x & \text{if } I = \emptyset, \\ \sum_{A \models I} (-1)^n \mu_A(\Delta_A(x)) & \text{if } I \neq \emptyset. \end{cases} \tag{10.12}$$

Here $A \models I$ means that A runs over all set compositions of I *with nonempty parts* (in particular, there are only finitely many summands). As in the Hopf algebra setting, this formula typically has massive cancellation, so in order to study a particular Hopf monoid it is desirable to find a cancellation-free formula.

Example 10.3.3. Let us calculate some antipodes in L. The trivial ordering on $\emptyset$ is trivially fixed by S, while for a singleton set $I = \{a\}$ we have $\mathsf{S}(a) = -a$ (the Takeuchi formula has only one term, corresponding to the set partition of I with one block). For $ab \in \mathsf{L}[\{a, b\}]$ we have

$$\begin{aligned} \mathsf{S}(ab) &= -\mu_{12}(\Delta_{12}(ab)) + \mu_{1|2}(\Delta_{1|2}(ab)) + \mu_{2|1}(\Delta_{2|1}(ab)) \\ &= -ab + (a)(b) + (b)(a) \\ &= -ab + ab + ba = ba, \end{aligned}$$

while for $abc \in \mathsf{L}[I]$ the antipode is calculated by the following table:

A	$\lvert A\rvert$	$\Delta_A(I)$	$(-1)^{\lvert A\rvert}\mu_A(\Delta_A(I))$				
123	1	abc	$-abc$				
$1, 23$	2	a, bc	$+abc$				
$2, 13$	2	b, ac		$+bac$			
$3, 12$	2	c, ab				$+cab$	
$12, 3$	2	ab, c	$+abc$				
$13, 2$	2	ac, b		$+acb$			
$23, 1$	2	bc, a			$+bca$		
$1, 2, 3$	3	a, b, c	$-abc$				
$1, 3, 2$	3	a, c, b		$-acb$			
$2, 1, 3$	3	b, a, c			$-bac$		
$2, 3, 1$	3	b, c, a				$-bca$	
$3, 1, 2$	3	c, a, b					$-cab$
$3, 2, 1$	3	c, b, a					$-cba$
		Total					$-cba$

It is starting to look suspiciously as though $\mathsf{S}_I = (-1)^I \mathsf{rev}$, where rev denotes the map that reverses ordering. In fact this is the case (proof left as an exercise). ◀

Material to be written: duality, L^*, generalized permutahedra and the Aguilar-Ardila antipode calculation, ...

10.4 Exercises

Problem 10.1. Show that in a Hopf algebra one has $\Delta \circ \mu = \mathrm{Id}$. (This is known as "Enrique's lemma" at KU.) An immediate corollary is that product and coproduct are injective and surjective, respectively.

Problem 10.2. Confirm that the polynomial Hopf algebra (Example 10.1.1) satisfies (10.2) and (10.4), and determine its antipode.

Problem 10.3. Confirm that the symmetric functions Λ form a Hopf subalgebra of the quasi-symmetric functions $QSym$, as asserted in Example 10.2.2. Then show that the antipode on Λ is the involution ω interchanging e_λ and h_λ (Proposition 9.5.3).

Problem 10.4. Let $E(M)$ denote the ground set of a matroid M, and call $|E(M)|$ the *order* of M. Let $\mathcal{M}_n$ be the vector space of formal $\mathbb{C}$-linear combinations of isomorphism classes $[M]$ of matroids M of order n. Let $\mathcal{M} = \bigoplus_{n \geq 0} \mathcal{M}_n$. Define a graded multiplication on $\mathcal{M}$ by $[M][M'] = [M \oplus M']$ and a graded comultiplication by

$$\Delta[M] = \sum_{A \subseteq E(M)} [M|_A] \otimes [M/A]$$

where $M|_A$ and M/A denote restriction and contraction respectively. Check that these maps make $\mathcal{M}$ into a graded bialgebra, and therefore into a Hopf algebra by Proposition 10.1.8.

Problem 10.5. Prove that the Billera–Jia–Reiner invariant defines a Hopf algebra morphism $\mathcal{M} \to QSym$.

Problem 10.6. Prove that the antipode in L is indeed given by $\mathsf{S}_I = (-1)^I \mathsf{rev}$, as in Example 10.3.3.

Chapter 11

More Topics

11.1 The Max-Flow/Min-Cut Theorem

The main theorem of this section is the Max-Flow/Min-Cut Theorem of Ford and Fulkerson. Strictly speaking, it probably belongs to graph theory or combinatorial optimization rather than algebraic combinatorics, but it is a wonderful theorem and has applications to posets and algebraic graph theory, so I can't resist including it.

Definition 11.1.1. A **network** N consists of a directed graph (V, E), two distinct vertices $s, t \in V$ (called the **source** and **sink** respectively), and a **capacity function** $c : E \to \mathbb{R}_{\geq 0}$.

Throughout this section, we will fix the symbols V, E, s, t, and c for these purposes. We will assume that the network has no edges into the source or out of the sink.

A network is supposed to model the flow of "stuff"—data, traffic, liquid, electrical current, etc.—from s to t. The capacity of an edge is the maximum amount of stuff that can flow through it (or perhaps the amount of stuff per unit time). This is a general model that can be specialized to describe cuts, connectivity, matchings and other things in directed and undirected graphs. This interpretation is why we exclude edges into s or out of t; we will see later why this assumption is in fact justified.

If $c(e) \in \mathbb{N}$ for all $e \in E$, we say the network is **integral**. In what follows, we will only consider integral networks.

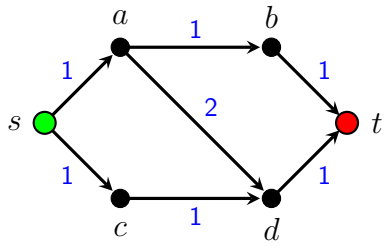


Figure 11.1: A network with source s, sink t, and capacity function c.

Definition 11.1.2. A **flow** on N is a function $f : E \to \mathbb{N}$ that satisfies the **capacity constraints**

$$0 \leq f(e) \leq c(e) \qquad \forall e \in E \tag{11.1}$$

and the **conservation constraints**

$$f^-(v) = f^+(v) \qquad \forall v \in V \setminus \{s, t\} \tag{11.2}$$

where

$$f^-(v) = \sum_{e=\overrightarrow{uv}} f(e), \qquad f^+(v) = \sum_{e=\overrightarrow{vw}} f(e).$$

The **value** of the flow is

$$|f| := f^-(t) = f^+(s)$$

(the equality of the second and third expressions follows from the conservation constraints, together with the observation $f^+(t) = f^-(s) = 0$).

The number $f(e)$ represents the amount of stuff flowing through e. That amount is bounded by the capacity of that edge, hence the constraints (11.1). Meanwhile, the conservation constraints say that stuff cannot accumulate at any internal vertex of the network, nor can it appear out of nowhere.

The **max-flow problem** is to find a flow of maximum value. The dual problem is the **min-cut problem**, which we now describe.

Definition 11.1.3. Let N be a network. Let $S, T \subseteq V$ with $S \cup T = V$, $S \cap T = \emptyset$, $s \in S$, and $t \in T$. The corresponding **cut** is

$$[S, T] \;=\; \{\overrightarrow{xy} \in E \colon x \in S, y \in T\}$$

and the **capacity** of that cut is

$$c(S, T) = \sum_{e \in [S,T]} c(e).$$

A cut can be thought of as a bottleneck through which all stuff must pass. For example, in the network of Figure 11.1, we could take $S = \{s, a, c\}$, $T = \{b, d, t\}$, so that $[S, T] = \{\overrightarrow{ab}, \overrightarrow{ad}, \overrightarrow{cd}\}$, and $c(S, T) = 1+2+1 = 4$.

The **min-cut** problem is to find a cut of minimum capacity. This problem is certainly feasible, since there are only finitely many cuts and each one has finite capacity.

For $A \subseteq V$, define $f^-(A) = \sum_{e \in [\bar{A}, A]} f(e)$, $f^+(A) = \sum_{e \in [A, \bar{A}]} f(e)$.

Proposition 11.1.4. *Let f be a flow, and let $A \subseteq V$. Then:*

$$f^+(A) - f^-(A) \;=\; \sum_{v \in A} (f^+(v) - f^-(v)). \tag{11.3a}$$

In particular, if $[S, T]$ is a cut, then

$$f^+(S) - f^-(S) = f^-(T) - f^+(T) \;=\; |f|, \tag{11.3b}$$

$$|f| \leq\; c(S, T). \tag{11.3c}$$

The proof (which requires little more than careful bookkeeping) is left as an exercise.

The inequality (11.3c) is known as **weak duality**; it says that the maximum value of a flow is less than or equal to the minimum capacity of a cut. (*Strong* duality would say that equality holds.)

Suppose that there is a path P from s to t in which no edge is being used to its full capacity. Then we can increase the flow along every edge on that path, and thereby increase the value of the flow by the same amount. As a simple example, we could start with the zero flow f_0 on the network of Figure 11.1 and increase flow by 1 on each edge of the path $sadt$; see Figure 11.2.

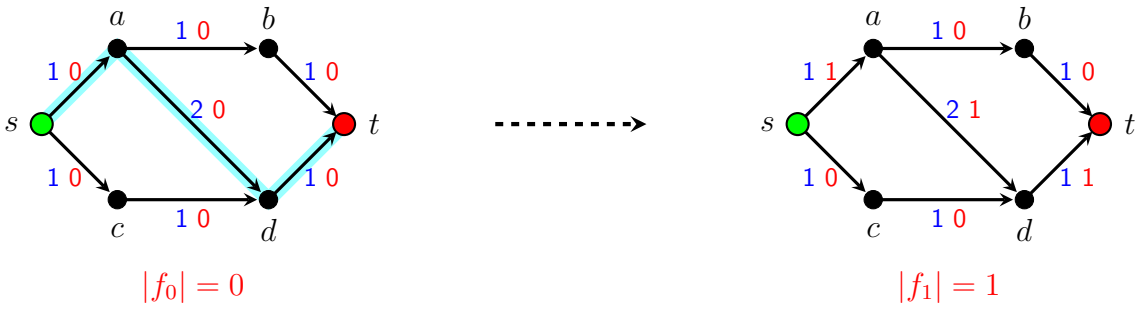


Figure 11.2: Increasing flow in the obvious way.

The problem is that there can exist flows that cannot be increased in this elementary way — but nonetheless are not maximum. The flow f_1 of Figure 11.2 is an example. In every path from s to t, there is some edge e with $f(e) = c(e)$. However, it easy to construct a flow of value 2:

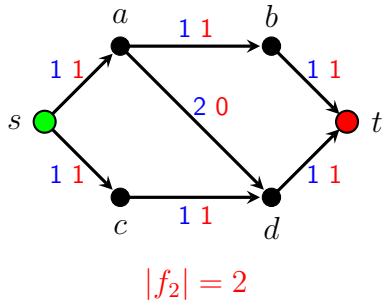


Figure 11.3: A better flow that cannot be obtained from f_1 in the obvious way.

Fortunately, there is a more general way to increase the value of a flow. The key idea is that flow along an edge $\overrightarrow{xy}$ can be regarded as *negative* flow from y to x. Accordingly, all we need is a path from s to t in which each edge e is either pointed forward and has $f(e) < c(e)$, or is pointed backward and has $f(e) > 0$. Then, increasing flow on the forward edges and decreasing flow on the backward edges will increase the value of the flow. This is called an **augmenting path** for f.

The **Ford-Fulkerson Algorithm** is a systematic way to construct a maximum flow by looking for augmenting paths. The wonderful feature of the algorithm is that if a flow f has no augmenting path, the algorithm will automatically find a cut of capacity equal to $|f|$ — thus certifying immediately that the flow is maximum and that the cut is minimum.

Input: An integral network N.
Initialization: Set f to the zero flow, i.e., $f(e) = 0$ for all edges e.

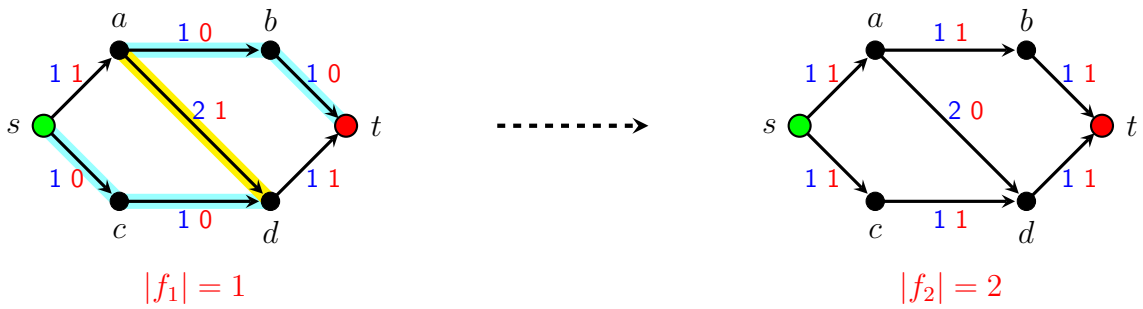


Figure 11.4: Exploiting the augmenting path $scdabt$ for f_1. The flow is increased by 1 on each of the "forward" edges sc, cd, ab, bt and decreased by 1 on the "backward" edge da to obtain the improved flow f_2.

1. If possible, find an *augmenting path*, i.e., a sequence of edges and vertices
$$P: \quad x_0 = s,\ e_1,\ x_1,\ e_2,\ x_2,\ \ldots,\ x_{n-1},\ e_n,\ x_n = t$$
such that the x_i are distinct and for every i, $i = 0, \ldots, n-1$, either
   - $e_i = \overrightarrow{x_{i-1}x_i} \in E$, and $f(e_i) < c(e_i)$ ("e_i is a forward edge"); or
   - $e_i = \overrightarrow{x_i x_{i-1}} \in E$, and $f(e_i) > 0$ ("e_i is a backward edge").
2. For each i, define the *tolerance* $\tau(e_i)$ to be $c(e_i) - f(e_i)$ if e_i is forward, or $f(e_i)$ if e_i is backward. (Note that $\tau(e_i) > 0$.) Define $\tau(P) = \min \tau(e_i)$.
3. Define $\tilde{f} : E \to \mathbb{N}$ by $\tilde{f}(e) = f(e) + \tau(P)$ if e appears forward in P; $\tilde{f}(e) = f(e) - \tau(P)$ if e appears backward in P; and $\tilde{f}(e) = f(e)$ if $e \notin P$. Then it is easy to verify $\tilde{f}$ satisfies the capacity and conservation constraints, and that $|\tilde{f}| = |f| + \tau(P)$.
4. Repeat steps 1–3 until no augmenting path can be found.

By integrality and induction, all tolerances are integers and all flows are integer-valued. In particular, each iteration of the loop increases the value of the best known flow by 1. Since the value of every flow is bounded by the minimum capacity of a cut (by weak duality), the algorithm is guaranteed to terminate in a finite number of steps. (By the way, Step 1 of the algorithm can be accomplished efficiently by a slight modification of, say, breadth-first search.)

The next step is to prove that this algorithm actually works. That is, when it terminates, it will have computed a flow of maximum possible value.

Proposition 11.1.5. *Suppose that f is a flow that has no augmenting path. Let*
$$S = \{v \in V : \textit{there is an augmenting path from } s \textit{ to } v\}, \qquad T = V \setminus S.$$
Then $s \in S$, $t \in T$, and $c(S,T) = |f|$. In particular, f is a maximum flow and $[S,T]$ is a minimum cut.

Proof. Note that $t \notin S$ precisely because f has no augmenting path. Applying (11.3b) gives
$$|f| \;=\; f^+(S) - f^-(S) \;=\; \sum_{e \in [S,\bar{S}]} f(e) - \sum_{e \in [\bar{S},S]} f(e) \;=\; \sum_{e \in [S,\bar{S}]} f(e).$$
But $f(e) = c(e)$ for every $e \in [S,T]$ (otherwise S would be bigger than what it actually is), so this last quantity is just $c(S,T)$. The final assertion follows by weak duality. □

We have proven:

Theorem 11.1.6 (Max-Flow/Min-Cut Theorem for Integral Networks ("MFMC"))**.** *For every integral network N, the maximum value of a flow equals the minimum value of a cut.*

In light of this, we will call the optimum of both the max-flow and min-cut problems the **value** of N, written $|N|$. In fact MFMC holds for non-integral networks as well, although the Ford-Fulkerson algorithm may not work in that case (the flow value might converge to $|N|$ without ever reaching it.)

Definition 11.1.7. Let N be a network. A flow f in N is *acyclic* if, for every directed cycle C in N (i.e., every set of edges $x_1 \to x_2 \to \cdots \to x_n \to x_1$), there is some $e \in C$ for which $f(e) = 0$. The flow f is *partitionable* if there is a collection of s,t-paths $P_1, \ldots, P_{|f|}$ such that for every $e \in E$,

$$f(e) = \#\{i \colon e \in P_i\}.$$

(Here "s,t-path" means "path from s to t".) In this sense f can be regarded as the "sum" of the paths P_i, each one contributing a unit of flow.

Proposition 11.1.8. *Let N be a network. Then:*

1. *For every flow in N, there exists an acyclic flow with the same value. In particular, N admits an acyclic flow with $|f| = |N|$.*
2. *Every acyclic integral flow is partitionable.*

Proof. Suppose that some directed cycle C has positive flow on every edge. Let $k = \min\{f(e) \colon e \in C\}$. Define $\tilde{f} : E \to \mathbb{N}$ by

$$\tilde{f}(e) = \begin{cases} f(e) - k & \text{if } e \in C, \\ f(e) & \text{if } e \notin C. \end{cases}$$

Then it is easy to check that $\tilde{f}$ is a flow, and that $|\tilde{f}| = |f|$. If we repeat this process, it must eventually stop (because the positive quantity $\sum_{e \in E} f(e)$ decreases with each iteration), which means that the resulting flow is acyclic. This proves (1).

Given a nonzero acyclic flow f, find an s,t-path P_1 along which all flow is positive. Decrement the flow on each edge of P_1; doing this will also decrement $|f|$. Now repeat this for an s,t-path P_2, etc. When the resulting flow is zero, we will have partitioned f into a collection of s,t-paths of cardinality $|f|$. □

Remark 11.1.9. This discussion justifies our earlier assumption that there are no edges into the source or out of the sink, since every acyclic flow must be zero on all such edges. Therefore, deleting those edges from a network does not change the value of its maximum flow.

This result has many applications in graph theory: Menger's theorems, the König-Egerváry theorem, etc.

11.2 Min-max theorems on posets

The basic result in this area is Dilworth's Theorem, which resembles the Max-Flow/Min-Cut Theorem (and can indeed be derived from it; see the exercises).

Definition 11.2.1. A **chain cover** of a poset P is a collection of chains whose union is P. The minimum size of a chain cover is called the **width** of P.

Let $m(P)$ denote the maximum size of an antichain in P.

Theorem 11.2.2 (Dilworth's Theorem)**.** *Let P be a finite poset. Then*

$$\text{width}(P) \;=\; m(P).$$

Proof. The "$\geq$" direction is clear, because if A is an antichain, then no chain can meet A more than once, so P cannot be covered by fewer than $|A|$ chains.

For the more difficult "$\leq$" direction, we induct on $n = |P|$. The result is trivial if $n = 1$ or $n = 2$.

Let Y be the set of all minimal elements of P, and let Z be the set of all maximal elements. Note that Y and Z are both antichains. First, suppose that no set other than Y or Z is a maximum[1] antichain; dualizing if necessary, we may assume $|Y| = m(P)$. Let $y \in Y$ and $z \in Z$ with $y \leq z$. Let $P' = P \setminus \{y, z\}$'; then $m(P') = |Y| - 1$. By induction, $\text{width}(P') \leq |Y| - 1$, and taking a chain cover of P' and tossing in the chain $\{y, z\}$ gives a chain cover of P of size $|Y|$.

Now, suppose that A is a maximum antichain other than Y or Z as a subset. Define

$$P^+ = \{x \in P \colon x \geq a \text{ for some } a \in A\},$$
$$P^- = \{x \in P \colon x \leq a \text{ for some } a \in A\}.$$

Then

- $P^+, P^- \neq A$ (otherwise A equals Z or Y).
- $P^+ \cup P^- = P$ (otherwise A is contained in some larger antichain).
- $P^+ \cap P^- = A$ (otherwise A isn't an antichain).

So P^+ and P^- are posets smaller than P, each of which contains A as a maximum antichain. By induction, each $P^\pm$ has a chain cover of size $|A|$. So for each $a \in A$, there is a chain $C_a^+ \subseteq P^+$ and a chain $C_a^- \subseteq P^-$ with $a \in C_a^+ \cap C_a^-$, and

$$\left\{C_a^+ \cap C_a^- \colon a \in A\right\}$$

is a chain cover of P of size $|A|$. □

If we switch "chain" and "antichain", then Dilworth's theorem remains true and becomes a much easier result.

Proposition 11.2.3 (Mirsky's Theorem)**.** *In any finite poset, the minimum size of an antichain cover equals the maximum size of an chain.*

Proof. For the $\geq$ direction, if C is a chain and $\mathcal{A}$ is an antichain cover, then no antichain in $\mathcal{A}$ can contain more than one element of C, so $|\mathcal{A}| \geq |C|$. On the other hand, let

$$A_i = \{x \in P \colon \text{ the longest chain headed by } x \text{ has length } i\};$$

then $\{A_i\}$ is an antichain cover whose cardinality equals the length of the longest chain in P. □

There is a marvelous common generalization of Dilworth's and Mirsky's Theorems due to Curtis Greene and Daniel Kleitman [GK76, Gre76]. An excellent source on this topic, including multiple proofs, is the survey article [BF01] by Thomas Britz and Sergey Fomin.

[1] I.e., a chain of size $m(P)$ — not merely a chain that is maximal with respect to inclusion, which might have smaller cardinality.

Theorem 11.2.4 (Greene-Kleitman). *Let P be a finite poset. Define two sequences of positive integers*

$$\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_\ell), \qquad \mu = (\mu_1, \mu_2, \ldots, \mu_m)$$

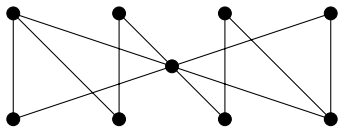
by

$$\lambda_1 + \cdots + \lambda_k = \max\big\{|C_1 \cup \cdots \cup C_k| : C_i \subseteq P \text{ chains}\big\},$$
$$\mu_1 + \cdots + \mu_k = \max\big\{|A_1 \cup \cdots \cup A_k| : A_i \subseteq P \text{ disjoint antichains}\big\}.$$

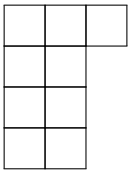
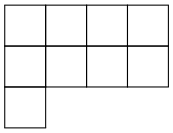
Then:

1. *λ and μ are both* partitions *of $|P|$, i.e., weakly decreasing sequences whose sum is $|P|$.*
2. *λ and μ are* conjugates *(written $\mu = \tilde{\lambda}$): the row lengths of λ are the column lengths in μ, and vice versa.*

Note that Dilworth's Theorem is just the special case $\mu_1 = \ell$. As an example, the poset with Hasse diagram



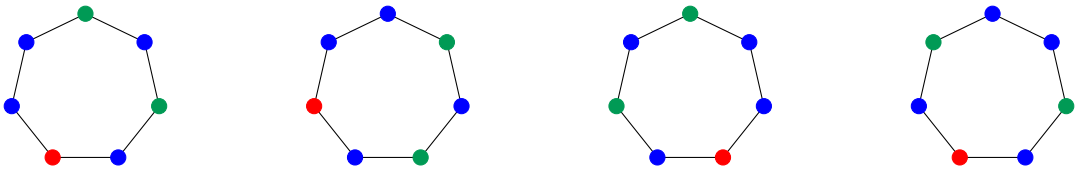
has

$\lambda = (3,2,2,2) =$  and $\mu = (4,4,1) =$  $= \tilde{\lambda}$.

11.3 Group actions and Polyá theory

How many different necklaces can you make with four blue, two green, and one red bead?

It depends what "different" means. The second necklace can be obtained from the first by rotation, and the third by reflection, but the fourth one is honestly different from the first two.



If we just wanted to count the number of ways to permute four blue, two green, and one red beads, the answer would be the multinomial coefficient

$$\binom{7}{4,2,1} = \frac{7!}{4!\,2!\,1!} = 105.$$

However, what we are really trying to count is *orbits under a group action*.

Let G be a group and X a set. An **action** of G on X is a group homomorphism $\alpha : G \to \mathfrak{S}_X$, the group of permutations of X.

Equivalently, an action can also be regarded as a map $G \times X \to X$, sending (g, x) to gx, such that

- $\mathrm{Id}_G\, x = x$ for every $x \in X$ (where Id_G denotes the identity element of G);
- $g(hx) = (gh)x$ for every $g, h \in G$ and $x \in X$.

The *orbit* of $x \in X$ is the set

$$O_x = \{gx \colon g \in G\} \subseteq X$$

and its *stabilizer* is

$$S_x = \{g \in G \colon gx = x\} \subseteq G,$$

which is a subgroup of G.

To go back to the necklace problem, we now see that "same" really means "in the same orbit". In this case, X is the set of all 105 necklaces, and the group acting on them is the dihedral group D_7 (the group of symmetries of a regular heptagon). The number we are looking for is the number of orbits of D_7.

Lemma 11.3.1. *Let $x \in X$. Then $|O_x||S_x| = |G|$.*

Proof. The element gx depends only on which coset of S_x contains g, so $|O_x|$ is the number of cosets, which is $|G|/|S_x|$. □

Proposition 11.3.2. *[Burnside's Theorem] The number of orbits of the action of G on X equals the average number of fixed points:*

$$\frac{1}{|G|} \sum_{g \in G} \#\{x \in X \colon gx = x\}$$

Proof. For a sentence P, let $\chi(P) = 1$ if P is true, or 0 if P is false (the "Garsia chi function"). Then

$$\begin{aligned} \text{Number of orbits} &= \sum_{x \in X} \frac{1}{|O_x|} = \frac{1}{|G|} \sum_{x \in X} |S_x| \\ &= \frac{1}{|G|} \sum_{x \in X} \sum_{g \in G} \chi(gx = x) \\ &= \frac{1}{|G|} \sum_{g \in G} \sum_{x \in X} \chi(gx = x) = \frac{1}{|G|} \sum_{g \in G} \#\{x \in X \colon gx = x\}. \end{aligned}$$

□

Typically, it is easier to count fixed points than to count orbits directly.

Example 11.3.3. We can apply this technique to the necklace example above.

- The identity of D_7 has 105 fixed points.
- Each of the seven reflections in D_7 has three fixed points (the single bead lying on the reflection line must be red, and then the two green beads must be equally distant from it, one on each side).
- Each of the six nontrivial rotations has no fixed points.

Therefore, the number of orbits is

$$\frac{105 + 7 \cdot 3}{|D_7|} = \frac{126}{14} = 9,$$

which is much more pleasant than trying to count them directly. ◀

Example 11.3.4. Suppose we wanted to find the number of orbits of 7-bead necklaces with 3 colors, without specifying how many times each color is to be used.

- The identity element of D_7 has $3^7 = 2187$ fixed points.
- Each reflection fixes one bead, which can have any color. There are then three pairs of beads flipped, and we can specify the color of each pair. Therefore, there are $3^4 = 81$ fixed points.
- Each rotation acts by a 7-cycle on the beads, so it has only three fixed points (all the beads must have the same color).

Therefore, the number of orbits is

$$\frac{2187 + 7 \cdot 81 + 6 \cdot 3}{14} = 198.$$

More generally, the number of inequivalent 7-bead necklaces with k colors allowed is

$$\frac{k^7 + 7k^4 + 6k}{14}. \tag{11.4}$$

◀

As this example indicates, it is helpful to look at the cycle structure of the elements of G, or more precisely on their images $\alpha(g) \in \mathfrak{S}_X$.

Proposition 11.3.5. *Let X be a finite set, and let $\alpha : G \to \mathfrak{S}_X$ be a group action. Color the elements of X with k colors, so that G also acts on the colorings.*

1. For $g \in G$, the number of fixed points of the action of g is $k^{\ell}(g)$, where $\ell(g)$ is the number of cycles in the disjoint-cycle representation of $\alpha(g)$.

2. Therefore,

$$\#\textit{equivalence classes of colorings} = \frac{1}{|G|} \sum_{g \in G} k^{\ell(g)}. \tag{11.5}$$

Let's rephrase Example 11.3.4 in this notation. The identity has cycle-type 1111111 (so $\ell = 7$); each of the six reflections has cycle-type 2221 (so $\ell = 4$); and each of the seven rotations has cycle-type 7 (so $\ell = 1$). Thus (11.4) is an example of the general formula (11.5).

Example 11.3.6. How many ways are there to k-color the vertices of a tetrahedron, up to moving the tetrahedron around in space?

Here X is the set of four vertices, and the group G acting on X is the alternating group on four elements. This is the subgroup of $\mathfrak{S}_4$ that contains the identity, of cycle-type 1111; the eight permutations of cycle-type 31; and the three permutations of cycle-type 22. Therefore, the number of colorings is

$$\frac{k^4 + 11k^2}{12}.$$

◀

11.4 Grassmannians

A standard reference for everything in this and the following section is Fulton [Ful97].

One motivation for the combinatorics of partitions and tableaux comes from classical enumerative geometric questions like this:

> **The Four-Lines Problem:** *Let there be given four lines L_1, L_2, L_3, L_4 in $\mathbb{R}^3$ in general position. How many lines M meet each of L_1, L_2, L_3, L_4 nontrivially?*

To a combinatorialist, "general position" means "all pairs of lines are skew, and their direction vectors are as linearly independent as possible — that is, the matroid they represent is $U_3(4)$." To a probabilist, it means "choose the lines randomly according to some reasonable measure on the space of all lines." So, what does the space of all lines look like?

In general, if V is a vector space over a field $\Bbbk$ (which we will henceforth take to be $\mathbb{R}$ or $\mathbb{C}$), and $0 \leq k \leq \dim V$, then the space of all k-dimensional vector subspaces of V is called the **Grassmannian** $\mathsf{Gr}(k, V)$. (Warning: this notation varies considerably from source to source.) As we will see, $\mathsf{Gr}(k, V)$ has many nice properties:

- It is a smooth manifold of dimension $k(n-k)$ over $\Bbbk$.
- It can be decomposed into pieces, called **Schubert cells**, each of which is naturally diffeomorphic to $\Bbbk^j$, for some appropriate j.
- The Schubert cells correspond to the interval $Y_{k,n} := [\emptyset, k^{n-k}]$ in Young's lattice. (Here $\emptyset$ means the empty partition and k^{n-k} means the partition with $n-k$ parts, all of size k, so that the Ferrers diagram is a rectangle.) That is, for each partition λ there is a corresponding Schubert cell Ω_λ of dimension $|\lambda|$ (the number of boxes in the Ferrers diagram).
- How these cells fit together topologically is described by $Y_{k,n}$ in the following sense: the closure of Ω_λ is given by the formula
$$\overline{\Omega_\lambda} = \bigcup_{\mu \leq \lambda} \Omega_\mu$$
where $\leq$ is the usual partial order on Young's lattice (i.e., containment of Ferrers diagrams).
- When $\Bbbk = \mathbb{C}$, the Poincaré polynomial of $\mathsf{Gr}(k, \mathbb{C}^n)$, i.e., the Hilbert series of the cohomology ring of $\mathsf{Gr}(k, \mathbb{C}^n)$,[2] is the rank-generating function for the graded poset $Y_{k,n}$, namely, the q-binomial coefficient $\begin{bmatrix} n \\ k \end{bmatrix}_q$ (see Problem 2.8(c)).

To accomplish all this, we need some way to describe points of the Grassmannian. For as long as possible, we won't worry about the ground field.

Let $W \in \mathsf{Gr}(k, \Bbbk^n)$; that is, W is a k-dimensional subspace of $V = \Bbbk^n$. We can describe W as the column

[2] If these terms don't make sense, here is a sketch of what you need to know. The cohomology ring $H^*(X) = H^*(X; \mathbb{Q})$ of a space X is just some ring that is a topological invariant of X. If X is a reasonably civilized space — say, a compact finite-dimensional real or complex manifold, or a finite simplicial complex — then $H^*(X)$ is a graded ring $H^0(X) \oplus H^1(X) \oplus \cdots \oplus H^d(X)$, where $d = \dim X$, and each graded piece $H^i(X)$ is a finite-dimensional $\mathbb{Q}$-vector space. The Poincaré polynomial records the dimensions of these vector spaces as a generating function:
$$\mathrm{Poin}(X, q) = \sum_{i=0}^{d} \dim_{\mathbb{Q}} H^i(X)\, q^i.$$
For lots of spaces, this polynomial has a nice combinatorial formula. For instance, take $X = \mathbb{R}P^d$ (real projective d-space). It turns out that $H^*(X) \cong \mathbb{Q}[z]/(z^{n+1})$. Each graded piece $H^i(X)$, for $0 \leq i \leq d$, is a 1-dimensional $\mathbb{Q}$-vector space (generated by the monomial x^i), and $\mathrm{Poin}(X, q) = 1 + q + q^2 + \cdots + q^d = (1 - q^{d+1})/(1 - q)$. In general, if X is a compact orientable manifold, then Poincaré duality implies (among other things) that $\mathrm{Poin}(X, q)$ is a palindrome.

space of a $n \times k$ matrix M of full rank:

$$M = \begin{bmatrix} m_{11} & \cdots & m_{1k} \\ \vdots & & \vdots \\ m_{n1} & \cdots & m_{nk} \end{bmatrix}.$$

However, the Grassmannian is not simply the space Z of all such matrices, because many different matrices can have the same column space. Specifically, any invertible column operation on M leaves its column space unchanged. On the other hand, every matrix whose column space is W can be obtained from M by some sequence of invertible column operations; that is, by multiplying on the right by some invertible $k \times k$ matrix. Accordingly, it makes sense to write

$$\mathsf{Gr}(k, \Bbbk^n) = Z/GL_k(\Bbbk). \tag{11.6}$$

That is, the k-dimensional subspaces of $\Bbbk^n$ can be identified with the orbits of Z under the action of the general linear group $GL_k(\Bbbk)$. In fact, as one should expect from (11.6),

$$\dim \mathsf{Gr}(k, \Bbbk^n) = \dim Z - \dim GL_k(\Bbbk) = nk - k^2 = k(n-k)$$

where "dim" means dimension as a manifold over $\Bbbk$; note that $\dim Z = nk$ because Z is a dense open subset of $\Bbbk^{n \times k}$. (Technically, this dimension calculation does not follow from (11.6) alone; you need to know that the action of $GL_k(\Bbbk)$ on Z is suitably well-behaved. Nevertheless, we will soon be able to calculate the dimension of $\mathsf{Gr}(k, \Bbbk^n)$ more directly.)

We now want to find a canonical representative for each $GL_k(\Bbbk)$-orbit. In other words, given $W \in \mathsf{Gr}(k, \Bbbk^n)$, we want the "nicest" matrix whose column space is W. How about the reduced column-echelon form? Basic linear algebra says that we can pick any matrix with column space W and perform Gauss-Jordan elimination on its columns, ending up with a uniquely determined matrix $M = M(W)$ with the following properties:

- $\operatorname{colspace} M = W$.
- The top nonzero entry of each column of M (the **pivot** in that column) is 1.
- Let p_i be the row in which the ith column has its pivot. Then $1 \le p_1 < p_2 < \cdots < p_k \le n$.
- Every entry below a pivot of M is 0, as is every entry to the right of a pivot.
- The remaining entries of M (i.e., other than the pivots and the 0s just described) can be anything whatsoever, depending on what W was in the first place.

For example, if $n = 4$ and $k = 2$, then M will have one of the following six forms:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 \\ 0 & \star \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 \\ 0 & \star \\ 0 & \star \\ 0 & 1 \end{bmatrix} \quad \begin{bmatrix} \star & \star \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \quad \begin{bmatrix} \star & \star \\ 1 & 0 \\ 0 & \star \\ 0 & 1 \end{bmatrix} \quad \begin{bmatrix} \star & \star \\ \star & \star \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{11.7}$$

Note that there is only one subspace W for which M ends up with the first form. At the other extreme, if the ground field $\Bbbk$ is infinite and you choose the space W uniformly at random (for basically any sensible measure on $\mathsf{Gr}(k, \Bbbk^n)$), then you will almost always end up with a matrix M of the last form.

Definition 11.4.1. Let $0 \le k \le n$ and let $\mathbf{p} = \{p_1 < \cdots < p_k\} \in \binom{[n]}{k}$ (i.e., $p_1, \ldots, p_k$ are distinct elements of $[n]$, ordered least to greatest). The **Schubert cell** $\Omega_{\mathbf{p}}$ is the set of all elements $W \in \mathsf{Gr}(k, \Bbbk^n)$ such that, for every i, the ith column of $M(W)$ has its pivot in row p_i.

Theorem 11.4.2. *1. Each $W \in \mathsf{Gr}(k, \Bbbk^n)$ belongs to exactly one Schubert cell; that is, $\mathsf{Gr}(k, \Bbbk^n)$ is the disjoint union of the subspaces $\Omega_{\mathbf{p}}$, for $\mathbf{p} \in \binom{[n]}{k}$.*

2. *For every* $\mathbf{p} \in \binom{[n]}{k}$*, there is a homeomorphism* $\Omega_{\mathbf{p}} \cong \mathbb{k}^{|\mathbf{p}|}$*, where* $|\mathbf{p}| = (p_1 - 1) + (p_2 - 2) + \cdots + (p_k - k) = p_1 + p_2 + \cdots + p_k - \binom{k+1}{2}$.
3. *Define a partial order on* $\binom{[n]}{k}$ *as follows: for* $\mathbf{p} = \{p_1 < \cdots < p_k\}$ *and* $\mathbf{q} = \{q_1 < \cdots < q_k\}$*, set* $\mathbf{p} \geq \mathbf{q}$ *if* $p_i \geq q_i$ *for every* i*. Then*
$$\mathbf{p} \geq \mathbf{q} \implies \overline{\Omega_{\mathbf{p}}} \supseteq \Omega_{\mathbf{q}}. \tag{11.8}$$
4. *The poset* $\binom{[n]}{k}$ *is isomorphic to the interval* $Y_{k,n}$ *in Young's lattice.*
5. $\mathsf{Gr}(k, \mathbb{k}^n)$ *is a compactification of the Schubert cell* $\Omega_{(n-k+1,n-k+2,\ldots,n)} \cong \mathbb{k}^{k(n-k)}$*. In particular,* $\dim_{\mathbb{k}} \mathsf{Gr}(k, \mathbb{k}^n) = k(n-k)$.

The cell closures $\overline{\Omega_{\mathbf{p}}}$ are called **Schubert varieties**.

Proof. (1) is immediate from the definition.

For (2), the map $\Omega_{\mathbf{p}} \to \mathbb{k}^{|\mathbf{p}|}$ is given by reading off the $\star$s in the reduced column-echelon form of $M(W)$. (For instance, let $n = 4$ and $k = 2$. Then the matrix representations in (11.7) give explicit diffeomorphisms of the Schubert cells of $\mathsf{Gr}(k, \mathbb{k}^n)$ to $\mathbb{k}^0$, $\mathbb{k}^1$, $\mathbb{k}^2$, $\mathbb{k}^2$, $\mathbb{k}^3$, $\mathbb{k}^4$ respectively.) The number of $\star$s in the i-th column is $p_i - i$ ($p_i - 1$ entries above the pivot, minus $i - 1$ entries to the right of previous pivots), so the total number of $\star$s is $|\mathbf{p}|$.

For (3): This is best illustrated by an example. Consider the second matrix in (11.7):
$$M = \begin{bmatrix} 1 & 0 \\ 0 & z \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$
where I have replaced the entry labeled $\star$ by a parameter z. Here's the trick: Multiply the second column of this matrix by the scalar $1/z$. Doing this doesn't change the column span, i.e.,
$$\operatorname{colspace} M = \operatorname{colspace} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1/z \\ 0 & 0 \end{bmatrix}.$$
Now you can see that
$$\lim_{|z|\to\infty} \operatorname{colspace} M = \operatorname{colspace} \lim_{|z|\to\infty} M = \operatorname{colspace} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$
which is the first matrix in (11.7). Therefore, the Schubert cell $\Omega_{1,2}$ is in the closure of the Schubert cell $\Omega_{1,3}$. In general, decrementing a single element of $\mathbf{p}$ corresponds to taking a limit of column spans in this way, so the covering relations in the poset $\binom{[n]}{k}$ give containment relations of the form (11.8).

Assertion (4) is purely combinatorial. The elements of $Y_{k,n}$ are partitions $\lambda = (\lambda_1, \ldots, \lambda_k)$ such that $n - k \geq \lambda_1 > \cdots > \lambda_k \geq 0$. The desired poset isomorphism is $\mathbf{p} \mapsto \lambda_{\mathbf{p}} = (p_k - k, p_{k-1} - (k-1), \ldots, p_1 - 1)$. For

example, starting with (11.7)

Matrix	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & \star \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & \star \\ 0 & \star \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} \star & \star \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} \star & \star \\ 1 & 0 \\ 0 & \star \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} \star & \star \\ \star & \star \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$
$\mathbf{p}$	12	13	14	23	24	34
$\lambda_{\mathbf{p}}$	$\emptyset$					

(5) now follows because $\mathbf{p} = (n-k+1, n-k+2, \ldots, n)$ is the unique maximal element of $\binom{[n]}{k}$, and an easy calculation shows that $|\mathbf{p}| = k(n-k)$. □

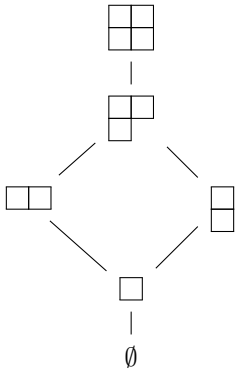
This theorem amounts to a description of $\mathsf{Gr}(k, \Bbbk^n)$ as a cell complex. (If you have not heard the term "cell complex" before, now you know what it means: a topological space that is the disjoint union of cells — that is, of homeomorphic copies of vector spaces — such that the closure of every cell is itself a union of cells.) Furthermore, the poset isomorphism with $Y_{k,n}$ says that for every i, the number of cells of $\mathsf{Gr}(k, \Bbbk^n)$ of dimension i is precisely the number of Ferrers diagrams with i blocks that fit inside the rectangle k^{n-k}. Combinatorially, we may write this equality as follows:

$$\sum_i (\#\text{ Schubert cells of dimension } i)\ q^i \ = \ \sum_i \#\{\lambda \subseteq k^{n-k}\}\ q^i \ = \ \begin{bmatrix} n \\ k \end{bmatrix}_q .$$

Example 11.4.3. If $k = 1$, then $\mathsf{Gr}(1, \Bbbk^n)$ is the space of lines through the origin in $\Bbbk^n$; that is, projective space $\Bbbk P^{n-1}$. As a cell complex, this has one cell of every dimension. For instance, the projective plane is the union of three cells of dimensions 2, 1, and 0, i.e., a plane, a line and a point. In the standard geometric picture, the 1-cell and 0-cell together form the "line at infinity". Meanwhile, the interval $Y_{k,n}$ is a chain of rank $n-1$. Its rank-generating function is $1 + q + q^2 + \cdots + q^{n-1}$. (For $\Bbbk = \mathbb{C}$, double the dimensions of all the cells, and substitute q^2 for q.) ◀

Remark 11.4.4. If $\Bbbk = \mathbb{C}$, then $\mathsf{Gr}(k, \mathbb{C}^n)$ is a cell complex with no odd-dimensional cells (because, topologically, the dimension of cells is measured over $\mathbb{R}$). Therefore, readers who know some algebraic topology (see, e.g., [Hat02, §2.2]) may observe that the cellular boundary maps are all zero (because each one has either zero domain or zero range), so the cellular homology groups are exactly the chain groups. That is, the Poincaré series of $\mathsf{Gr}(k, \mathbb{C}^n)$ is exactly the generating function for the dimensions of the cells. On the other hand, If $\Bbbk = \mathbb{R}$, then the boundary maps need not be zero, and the homology can be more complicated. Indeed, $\mathsf{Gr}(1, \mathbb{R}^n) = \mathbb{R}P^{n-1}$ has torsion homology in odd dimensions.

Example 11.4.5. Let $n = 4$ and $k = 2$. Here is $Y_{k,n}$:



These six partitions correspond to the six matrix-types in (11.7). The rank-generating function is

$$\begin{bmatrix}4\\2\end{bmatrix}_q = \frac{(1-q^4)(1-q^3)}{(1-q^2)(1-q)} = 1+q+2q^2+q^3+q^4.$$

◀

Remark 11.4.6. What does all this have to do with enumerative geometry questions such as the Four-Lines Problem? The answer (modulo technical details) is that the cohomology ring $H^*(X)$ encodes intersections of subvarieties[3] of X: for every subvariety $Z \subseteq \mathsf{Gr}(k, \mathbb{k}^n)$ of codimension i, there is a corresponding element $[Z] \in H^i(X)$ (the "cohomology class of Z") such that $[Z \cup Z'] = [Z] + [Z']$ and $[Z \cap Z'] = [Z][Z']$. These equalities hold only if Z and Z' are in general position with respect to each other (which has to be defined precisely), but the consequence is that the Four-Lines Problem reduces to a computation in $H^*(\mathsf{Gr}(k, \mathbb{k}^n))$: find the cohomology class $[Z]$ of the subvariety

$$Z = \{W \in \mathsf{Gr}(2, \mathbb{C}^4) \colon W \text{ meets some plane in } \mathbb{C}^4 \text{ nontrivially}\}$$

and compare $[Z]^4$ to the cohomology class $[\bullet]$ of a point. In fact, $[Z]^4 = 2[\bullet]$; this says that the answer to the Four-Lines Problem is **two**, which is hardly obvious! To carry out this calculation, one needs to calculate an explicit presentation of the ring $H^*(\mathsf{Gr}(k, \mathbb{k}^n))$ as a quotient of a polynomial ring (which requires the machinery of line bundles and Chern classes, but that's another story) and then figure out how to express the cohomology classes of Schubert cells with respect to that presentation. This is the theory of **Schubert polynomials**.

11.5 Flag varieties

There is a corresponding theory for the **flag variety**, which is the set $F\ell(n)$ of nested chains of vector spaces

$$F_\bullet = (0 = F_0 \subseteq F_1 \subseteq \cdots \subseteq F_n = \mathbb{k}^n)$$

or equivalently saturated chains in the (infinite) lattice $L_n(\mathbb{k})$. The flag variety is in fact a smooth manifold over $\mathbb{k}$ of dimension $\binom{n}{2}$. Like the Grassmannian, it has a decomposition into Schubert cells X_w, which are indexed by permutations $w \in \mathfrak{S}_n$ rather than partitions, as we now explain.

For every flag $F_\bullet$, we can find a vector space basis $\{v_1, \ldots, v_n\}$ for $\mathbb{k}^n$ such that $F_k = \mathbb{k}\langle v_1, \ldots, v_k\rangle$ for all k, and represent $F_\bullet$ by the invertible matrix $M \in G = GL(n, \mathbb{k})$ whose columns are $v_1, \ldots, v_n$. OTOH, any ordered basis of the form

$$v_1' = b_{11}v_1, \quad v_2' = b_{12}v_1 + b_{22}v_2, \quad \ldots, v_n' = b_{1n}v_1 + b_{2n}v_2 + \cdots + b_{nn}v_n,$$

where $b_{kk} \neq 0$ for all k, defines the same flag. That is, a flag is a coset of B in G, where B is the subgroup of invertible upper-triangular matrices (the **Borel subgroup**). Thus the flag variety can be (and often is) regarded as the quotient G/B. This immediately implies that it is an irreducible algebraic variety (as G is irreducible, and any image of an irreducible variety is irreducible). Moreover, it is smooth (e.g., because every point looks like every other point, and so either all points are smooth or all points are singular and the latter is impossible) and its dimension is $(n-1) + (n-2) + \cdots + 0 = \binom{n}{2}$.

As in the case of the Grassmannian, there is a canonical representative for each coset of B, obtained by Gaussian elimination, and reading off its pivot entries gives a decomposition

$$F\ell(n) = \coprod_{w \in \mathfrak{S}_n} X_w.$$

[3] If you are more comfortable with differential geometry than algebraic geometry, feel free to think "submanifold" instead of "subvariety".

Here the dimension of a Schubert cell X_w is the number of **inversions** of w, i.e.,

$$\dim X_w = \mathrm{inv}(w) = \#\{(i,j) : \ 1 \le i < j \le n \text{ and } w(i) > w(j)\}.$$

Recall that this is the rank function of the Bruhat and weak Bruhat orders on $\mathfrak{S}_n$. In fact, the (strong) Bruhat order is the cell-closure partial order (analogous to (11.8)). It follows that the Poincaré polynomial of $F\ell(n)$ is the rank-generating function of Bruhat order, namely

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

More strongly, it can be shown that the cohomology ring $H^*(F\ell(n);\mathbb{Z})$ is the quotient of $\mathbb{Z}[x_1,\ldots,x_n]$ by the ideal generated by symmetric functions.

The Schubert varieties in $F\ell(n)$ are

$$\overline{X_w} = \bigcup_{v \in \mathfrak{S}_n :\ v \le w} X_v$$

where $\le$ means (strong) Bruhat order (see Ex. 1.2.13). These are much-studied objects in combinatorics; for example, determining which Schubert varieties are singular turns out to to be a combinatorial question involving the theory of *pattern avoidance*. Even more generally, instead of $\mathfrak{S}_n$, start with any finite Coxeter group G (roughly, a group generated by elements of order two — think of them as reflections). Then G has a combinatorially well-defined partial order also called the Bruhat order, and one can construct a G-analogue of the flag variety: that is, a smooth manifold whose structure as a cell complex is given by Bruhat order on G.

We now describe the calculation of the cohomology ring of $F\ell(n)$ using Chern classes. This is *not* intended to be self-contained, and many facts will be presented as black boxes. For the full story, refer to, e.g., [BT82].

Definition 11.5.1. Let B and F be topological spaces. A **bundle with base B and fiber F** is a space $\mathcal{E}$ together with a map $\pi : \mathcal{E} \to B$ such that

1. If $b \in B$, then $\pi^{-1}(b) \cong F$; and, more strongly,
2. Every $b \in B$ has an open neighborhood U of b such that $V := \pi^{-1}(U) \cong U \times F$, and $\pi|_V$ is just projection on the first coordinate.

Think of a bundle as a family of copies of F parameterized by B and varying continuously. The simplest example of a bundle is a Cartesian product $B \times F$ with $\pi(b,f) = b$; this is called a **trivial bundle**. Very often the fiber is a vector space of dimension d, when we call the bundle a **vector bundle of rank d**; when $d = 1$ the bundle is a **line bundle**.

Frequently we require all these spaces to lie in a more structured category than that of topological spaces, and we require the projection map to be a morphism in that category (e.g., manifolds with diffeomorphisms, or varieties with algebraic maps).

Example 11.5.2. An example of a nontrivial bundle is a Möbius strip M, where $B = S^1$ is the central circle and $F = [0,1]$ is a line segment. Indeed, a Möbius strip looks like a bunch of line segments parameterized by a circle, and if U is any small interval in S^1 then the part of the bundle lying over U is just $U \times [0,1]$. However, the global structure of M is not the same as the cylinder $S^1 \times I$. ◀

Example 11.5.3. Another important example is the **tautological bundle** on projective space $\mathbb{P}^{d-1}_{\Bbbk} = \mathsf{Gr}(1, \Bbbk^d)$. Recall that this is the space of lines ℓ through the origin in $\Bbbk^d$. The tautological bundle[4] $\mathcal{T}$ is the line bundle defined by $\mathcal{T}_\ell = \ell$. That is, the fiber over a line is just the set of points on that line. ◀

[4]The standard symbol for the tautological bundle is actually $\mathcal{O}(-1)$; let's not get into why.

Let $\Bbbk$ be either $\mathbb{R}$ or $\mathbb{C}$, and let us work in the category of closed compact manifolds over $\Bbbk$. A **vector bundle of rank d** is a bundle whose fiber is $\Bbbk^d$. (For example, the tautological bundle is a vector bundle of rank 1.) Standard operations on vector spaces (direct sum, tensor product, dual, etc.) carry over to vector bundles, defined fiberwise.

Let $\mathcal{E}$ be a rank-d vector bundle over M. Its **projectivization** $\mathbb{P}(\mathcal{E})$ is the bundle with fiber $\mathbb{P}^{d-1}_{\Bbbk}$ defined by

$$\mathbb{P}(\mathcal{E})_m = \mathbb{P}(\mathcal{E}_m).$$

That is, a point in $\mathbb{P}(\mathcal{E})$ is given by a point $m \in M$ and a line ℓ through the origin in $\mathcal{E}_m \cong \Bbbk^d$. In turn, $\mathbb{P}(\mathcal{E})$ has a tautological line bundle $\mathscr{L} = \mathscr{L}(\mathcal{E})$ whose fiber over (ℓ, m) is ℓ.

Associated with the bundle $\mathcal{E}$ are certain **Chern classes** $c_i(\mathcal{E}) \in H^{2i}(M)$ for every i, which measure "how twisty $\mathcal{E}$ is." (The 2 happens because we are talking about a complex manifold.) I will not define these classes precisely (see [BT82]), but instead will treat them as a black box that lets us calculate cohomology. The Chern classes have the following properties:

1. $c_0(\mathcal{E}) = 1$ by convention.
2. $c_i(\mathcal{E}) = 0$ for $i > \operatorname{rank} \mathcal{E}$.
3. If $\mathcal{E}$ is trivial then $c_i(\mathcal{E}) = 0$ for $i > 0$.
4. If $0 \to \mathcal{E}' \to \mathcal{E} \to \mathcal{E}'' \to 0$ is an exact sequence of M-bundles, then $c(\mathcal{E}) = c(\mathcal{E}')c(\mathcal{E}'')$, where $c(\mathcal{E}) = \sum_i c_i(\mathcal{E})$ (the "total Chern class").
5. For a line bundle L, $c_1(L^*) = -c_1(L)$.

Here is the main formula, which expresses the cohomology ring of a bundle as a module over the cohomology of its base.

$$H^*(\mathbb{P}(\mathcal{E}); \mathbb{Z}) = H^*(M; \mathbb{Z})[x] / \langle x^d + c_1(\mathcal{E})x^{d-1} + \cdots + c_{d-1}(\mathcal{E})x + c_d(\mathcal{E}) \rangle \tag{11.9}$$

where $x = c_1(\mathscr{L})$.

Example 11.5.4 (Projective space). $\mathbb{P}^{d-1}\mathbb{C}$ is the projectivization of the trivial rank-d bundle over $M = \{\bullet\}$. Of course $H^*(M; \mathbb{Z}) = \mathbb{Z}$, so $H^*(\mathbb{P}^{d-1}\mathbb{C}; \mathbb{Z}) = \mathbb{Z}[x]/\langle x^d \rangle$. ◀

Example 11.5.5 (The flag variety $F\ell(3)$). Let $M = \mathbb{P}^2 = \mathsf{Gr}(1, \mathbb{C}^3)$. Define a bundle $\mathcal{E}^2$ by

$$\mathcal{E}^2_\ell = \mathbb{C}^3/\ell.$$

Then $\mathcal{E}^2$ has rank 2, and $\mathbb{P}(\mathcal{E}^2)$ is just the flag variety $F\ell(3)$, because specifying a line in $\mathbb{C}^3/\ell$ is the same thing as specifying a plane in $\mathbb{C}^3$ containing ℓ. Let $\mathscr{L} = \mathscr{L}(\mathcal{E}^2)$. For each $\ell \in M$ we have an exact sequence $0 \to \ell \to \mathbb{C}^3 \to \mathbb{C}^3/\ell \to 0$, which gives rise to a short exact sequence of bundles

$$0 \to \mathcal{O} \to \mathbb{C}^3 \to \mathcal{E}^2 \to 0$$

where $\mathcal{O}$ is the tautological bundle on M, with $c_1(\mathcal{O}) = x$ (the generator of $H^*(M)$). The rules for Chern classes then so the rules for Chern classes tell us that

$$(1 + x)(1 + c_1(\mathcal{E}^2) + c_2(\mathcal{E}^2)) = 1$$

and extracting the graded pieces we get

$$x + c_1(\mathcal{E}^2) = 0, \qquad xc_1(\mathcal{E}^2) + c_2(\mathcal{E}^2) = 0$$

so $c_1(\mathcal{E}^2) = -x$ and $c_2(\mathcal{E}^2) = -xc_1(\mathcal{E}^2) = x^2$. Now (11.9) tells us that

$$H^*(F\ell(3)) = H^*(\mathbb{P}^2)[y] / \langle y^2 - xy + x^2 \rangle = \mathbb{Q}[x, y] / \langle x^3, y^2 - xy + x^2 \rangle.$$

In fact this ring is isomorphic to

$$\mathbb{Q}[a,b,c]/\langle a+b+c,\ ab+ac+bc,\ abc\rangle.$$

(For the isomorphism, set $a = x$, $b = -y$, $c = -x + y$.) ◀

Example 11.5.6 (General flag varieties.). $F\ell(n)$ can be constructed as an iterated bundle:

$X_0 = \{\bullet\}$. Let $\mathcal{E}_0$ be the (trivial) rank-n bundle over X_0.

$X_1 = \mathbb{P}(\mathcal{E}_0)$. Let $\mathcal{E}_1$ be the rank-$(n-1)$ bundle whose fiber over a line E_1 is $\mathbb{C}^n/E_1$.

$X_2 = \mathbb{P}(\mathcal{E}_1)$. This is the partial flag variety of flags $E_\bullet :\ 0 = E_0 \subseteq E_1 \subseteq E_2$. Let $\mathcal{E}_2$ be the rank-$(n-2)$ bundle whose fiber over $E_\bullet$ is $\mathbb{C}^n/E_2$.

$X_3 = \mathbb{P}(\mathcal{E}_2)$. And so forth.

We end up with generators $x_1, \ldots, x_n$, one for the tautological bundle of each $\mathcal{E}_i$. The relations turn out to be the symmetric functions on them. That is.

$$H^*(F\ell(n)) \cong \mathbb{Q}[x_1, \ldots, x_n]/\langle e_1, e_2, \ldots, e_n\rangle$$

where e_k is the kth elementary symmetric function, i.e.,

$$e_k = \sum_{1 \le i_1 < \cdots < i_k \le n} x_{i_1} \cdots x_{i_k}.$$

◀

The Poincare polynomial of the flag variety (i.e., the Hilbert series of its cohomology ring) can be worked out explicitly. Modulo the elementary symmetric functions, every polynomial can be written as a sum of monomials of the form

$$x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n}$$

where $a_i < i$ for all i. Therefore,

$$\mathrm{Poin}(F\ell(n), q) \ = \ \sum_k q^k \dim_{\mathbb{Q}} H^{2i}(F\ell(n)) \ = \ (1)(1+q)(1+q+q^2)\cdots(1+q+\cdots+q^{n-1}) = [q]_n!.$$

This expression has a lovely combinatorial interpretation:

$$[q]_n! = \sum_{w \in \mathfrak{S}_n} q^{\mathrm{inv}(w)}$$

where $\mathfrak{S}_n$ is the symmetric group on n letters and $\mathrm{inv}(w)$ is the number of inversions:

$$\mathrm{inv}(w) = \#\{(i,j) :\ 1 \le i < j \le n,\ w(i) > w(j)\}.$$

In fact the flag variety has a natural cell decomposition into **Schubert cells**. Given any flag

$$E_\bullet :\ 0 = E_0 \subseteq E_1 \subseteq \cdots \subseteq E_n = \mathbb{C}^n$$

construct a $n \times n$ matrix $[v_1|\cdots|v_n]$ in which the first k columns are a basis of $\mathcal{E}_k$, for every k. We can canonicalize the matrix as follows:

- Scale the first column so that its bottom nonzero entry is 1. Say this occurs in row w_1.
- Add an appropriate multiple of v_1 to each of $v_2, \ldots, v_n$ so as to kill off the entry in row w_1. Note that this does not change the flag.
- Scale the second column so that its bottom nonzero entry is 1. Say this occurs in row w_2. Note that $w_2 \neq w_1$.
- Add an appropriate multiple of v_2 to each of $v_3, \ldots, v_n$ so as to kill off the entry in row w_1.
- Repeat.

(Here we are really using the description

$$F\ell(n) = GL_n/B$$

where B is the Borel subgroup of upper-triangular invertible matrices. The column operations that we have done correspond to choosing a canonical element of each coset of B in GL_n.)

We end up with a matrix that includes a "pivot" 1 in each row and column, with zeroes below and to the right of every 1. The pivots define a permutation $w \in \mathfrak{S}_n$. For example, if $w = 4132$ then the matrix will have the form

$$\begin{bmatrix} * & 1 & 0 & 0 \\ * & 0 & * & 1 \\ * & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

The set X°_{3142} of all matrices of this type is a subspace of $F\ell(4)$ that is in fact isomorphic to $\mathbb{C}^3$ — the stars are affine coordinates. Thus we obtain a decomposition into **Schubert cells**

$$F\ell(n) = \coprod_{w \in \mathfrak{S}_n} X^\circ_w$$

and moreover the stars correspond precisely to inversions of w. This gives the Poincaré polynomial.

The closure of a Schubert cell is called a **Schubert variety**. The cohomology classes of Schubert varieties are also a vector space basis for $H^*(F\ell(n))$, and there is a whole theory of how to translate between the "algebraic" basis (coming from line bundles) and the "geometric" basis (Schubert varieties).

11.6 Exercises

Max-flow/min-cut and min-max theorems on posets

Problem 11.1. Prove Proposition 11.1.4.

Problem 11.2. Let $G(V, E)$ be a graph. A **matching** on G is a collection of edges no two of which share an endpoint. A **vertex cover** is a set of vertices that include at least one endpoint of each edge of G. Let $\mu(G)$ denote the size of a maximum matching, and let $\beta(G)$ denote the size of a minimum vertex cover.

(a) (Warmup) Show that $\mu(G) \leq \beta(G)$ for every graph G. Exhibit a graph for which the inequality is strict.
(b) The *König-Egerváry Theorem* asserts that $\mu(G) = \beta(G)$ whenever G is bipartite, i.e., the vertices of G can be partitioned as $X \cup Y$ so that every edge has one endpoint in each of X, Y. Derive the König-Egerváry Theorem as a consequence of the Max-Flow/Min-Cut Theorem.
(c) Prove that the König-Egerváry Theorem and Dilworth's Theorem imply each other.

Polyá theory

Problem 11.3. Let $n \geq 2$ and for $\sigma \in \mathfrak{S}_n$, let $f(\sigma)$ denote the number of fixed points. Prove that for every $k \geq 1$, the number $\frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} f(\sigma)^k$ is an integer.

Grassmannians and flag varieties

Appendix: Catalan Numbers

The *Catalan numbers* are the sequence $C_0, C_1, \ldots$, defined by

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

The Catalan numbers are ubiquitous in combinatorics. A famous exercise in volume 2 of Stanley's *Enumerative Combinatorics* [Sta99, Problem 6.19] lists 66 combinatorial interpretations of the Catalan numbers and asks the reader to come up with $\binom{66}{2}$ bijections between them. That was in 1999; more recently, Stanley wrote an entire monograph [Sta15] with 214 interpretations. Here we will just review the basics.

A **Dyck path of size n** is a path from $(0,0)$ to $(2n,0)$ in $\mathbb{R}^2$ consisting of n up-steps and n down-steps that stays (weakly) above the x-axis.

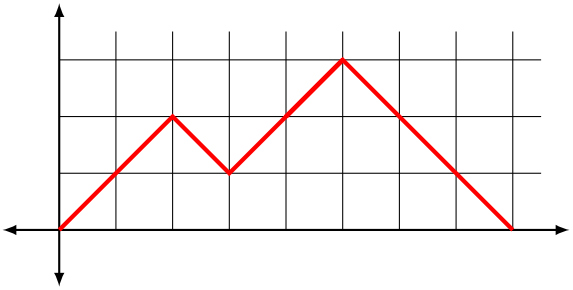


Figure 11.5: A Dyck path of size 4.

We can denote Dyck paths efficiently by a list of U's and D's; the path P shown above is UUDUUDDD. Each up-step can be thought of as a left parenthesis, and each down-step as a right parenthesis, so we could also write $P = (()(()))$. The requirement of staying above the x-axis then says that each right parenthesis must close a previous left parenthesis.

Proposition 11.6.1. *The number of Dyck paths of size n is the Catalan number C_n.*

Sketch of proof. The proof is an illustration of the Sheep Principle ("in order to count the sheep in a flock, count the legs and divide by four"). Consider the family L of all lattice paths from $(0,0)$ to $(2n+1,-1)$ consisting of n up-steps and $n+1$ down-steps (with no restrictions); evidently $|L| = \binom{2n+1}{n}$.

Consider the action of the cyclic group $\mathbb{Z}_{2n+1}$ on L by cyclic rotation. First, the orbits all have size $2n+1$. (There is no way that a nontrivial element of $\mathbb{Z}_{2n+1}$ can fix the locations of the up-steps, essentially because $\gcd(2n+1, n) = 1$ — details left to the reader.) Second, each orbit contains exactly one *augmented Dyck path*,

i.e., a Dyck path followed by a down-step. (Of all the lowest points in a path, find the leftmost one and call it z. Rotate so that the last step is the down-step that lands at z.)

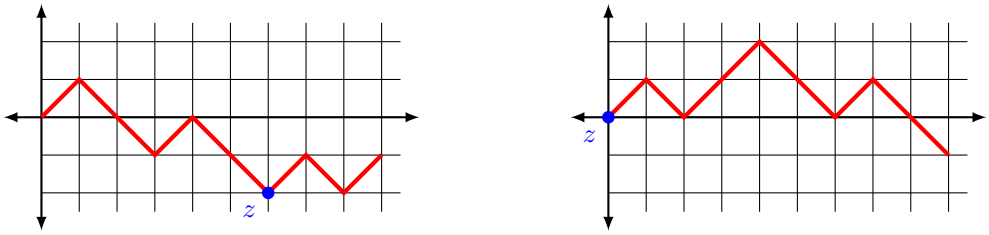


Figure 11.6: Rotating the lattice path UDDUDD|UDU to obtain the augmented Dyck path UDU|UDDUDD.

Every (augmented) Dyck path arises in this way, so we have a bijection. The orbits are sheep and each sheep has $2n+1$ legs, so the number of Dyck paths is

$$\frac{1}{2n+1}\binom{2n+1}{n} = \frac{(2n+1)!}{(2n+1)\,(n+1)!\,n!} = \frac{(2n)!}{(n+1)!\,n!} = \frac{(2n)!}{(n+1)\,n!\,n!} = \frac{1}{n+1}\binom{2n}{n}. \qquad \square$$

To show that a class of combinatorial objects is enumerated by the Catalan numbers, one can now find a bijection to Dyck paths. A few of the most commonly encountered interpretations of C_n are:

- Triangulations of a convex $(n+2)$-gon into n triangles using $n-1$ diagonals.
- Binary trees with n vertices. ("Binary" means that each vertex has at most 2 children.)
- Plane trees with n vertices. ("Plane" means that each set of siblings comes with a left-to-right order.)

Others will be encountered in the course of these notes. For details, see [Sta99] or [Sta15] Another core feature of the Catalan numbers is that they satisfy the following recurrence:

$$C_n = C_{n-1} + \sum_{k=1}^{n-1} C_{k-1}C_{n-k} \qquad \text{for } n \geq 1. \tag{11.10}$$

This equation can be checked by a banal induction argument, but it is also worthwhile seeing the combinatorial reason for it. Call a Dyck path of size n *primitive* if it stays strictly above the x-axis for $0 < x < 2n$. If a path P is primitive, then it is of the form UP'D for some Dyck path P' of size $n-1$ (not necessarily primitive); this accounts for the C_{n-1} term in the Catalan recurrence. Otherwise, let $(2k,0)$ be the smallest positive x-intercept, so that $1 \leq k \leq n-1$. The part of the path from $(0,0)$ to $(2k,0)$ is a primitive Dyck path of size k, and the part from $(2k,0)$ to $(2n,0)$ is a Dyck path of size $n-k$, not necessarily primitive.

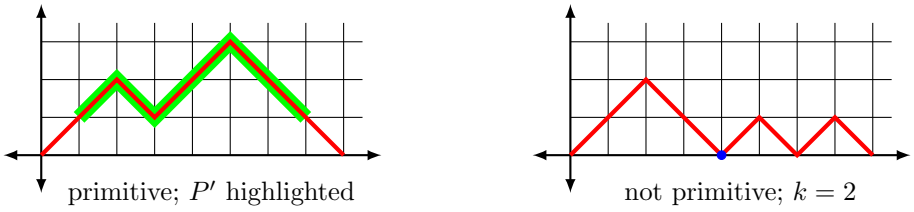


Figure 11.7: Primitive and non-primitive Dyck paths.

Notational Index

Basics

◀	End of an example	
$[n]$	$\{1, \ldots, n\}$	
$\mathbb{N}$	nonnegative integers $0, 1, 2, \ldots$	`\mathbb{N}`
$\mathbb{N}_{>0}$	positive integers $1, 2, \ldots$	`\mathbb{P}`
2^S	power set of a set S (or the associated poset)	
$\lvert S\rvert$ or $\#S$	cardinality of set S	
$\cupdot$	disjoint union	`\cupdot` (requires header.tex)
$\triangle$	symmetric difference $A \triangle B = (A \cup B) \setminus (A \cap B)$	`\triangle`
$\mathfrak{S}_n$	symmetric group on n letters	`\mathfrak{S}_n`
$\binom{S}{k}$	set of k-element subsets of a set S	`\binom{S}{k}`
C_n	Catalan numbers	
$\Bbbk\langle \mathbf{v}_1, \ldots, \mathbf{v}_n \rangle$	$\Bbbk$-vector space with basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$	`\fld\langle ... \rangle`

Posets

$\gtrdot, \lessdot$	"covers", "is covered by"	`\gtrdot`, `\lessdot`
$\hat{\mathbf{0}}, \hat{\mathbf{1}}$	unique min and max elements of a poset	`\hat{\mathbf{0}}`
$[x, y]$, $[x, y]_P$	interval in a poset	
P^*	dual poset to P	
Π_n	lattice of all set partitions of $[n]$	
$K(G)$	connectivity lattice of a graph G	
Y	Young's lattice of integer partitions	
$\lambda \vdash n$	λ is a partition of n	`\vdash`
$\tilde{\lambda}$	conjugate of a partition λ	

Lattices

$\wedge, \vee$	meet, join	`\wedge`, `\vee`
Bool_n	Boolean lattice of rank n	`\mathscr{B}_n`
D_n	lattice of divisors of an integer n	
$\mathbb{F}_q$	finite field of order q	`\mathbb{F}_q`
$L_n(q)$	lattice of vector subspaces of $\mathbb{F}_q^n$	
$J(P)$	lattice of order ideals of a poset P	
$\mathrm{Irr}(L)$	poset of join-irreducible elements in a lattice L	
N_5	nonmodular, nonranked 5-element lattice	
M_5	modular, ranked, nondistributive 5-element lattice	
$L(E)$	geometric lattice represented by vectors E	
$L^{\mathrm{aff}}(E)$	geometric lattice represented by affine points E	

Poset Algebra

$\mathrm{Int}(P)$	set of intervals of poset P	
$I(P)$	incidence algebra of P	
$f * g$	convolution product in $I(P)$	
δ	Kronecker delta function (identity of $I(P)$)	
ζ, μ	zeta and Möbius functions in $I(P)$	
$\chi_P(x)$	characteristic polynomial of poset P	
$\mathrm{Möb}(L)$	Möbius algebra of a lattice L	

Matroids and the Tutte Polynomial

$A - e$, $A + e$	abbreviations for $A \setminus \{e\}$, $A \cup \{e\}$	
$\bar{A}$	closure operator applied to A	
$M(G)$	graphic matroid of a graph G	
$\mathscr{I}$	matroid independence system	`\mathscr{I}` (uses `mathrsfs` package)
$\mathscr{B}$	matroid basis system	`\mathscr{B}`
$\mathscr{C}$	matroid circuit system	`\mathscr{C}`
M^*	dual of a matroid M	
$M \oplus M'$	direct sum of matroids	`\oplus`
$M\backslash e$, M/e	matroid deletion and contraction	
$U_k(n)$	uniform matroid of rank k on set of size n	
T_M, $T_M(x, y)$	Tutte polynomial of M	
$p_G(k)$	chromatic polynomial of a graph G	
$C(e, B)$	fundamental circuit of e w/r/t basis B	
$C^*(e, B)$	fundamental cocircuit of e w/r/t basis B	

Hyperplane Arrangements

Bool_n	Boolean arrangement	
Br_n	braid arrangement	
$\mathcal{A}_G$	arrangement associated with a graph G	
$L(\mathcal{A})$	intersection poset of arrangement $\mathcal{A}$	
$\mathrm{ess}(\mathcal{A})$	essentialization of $\mathcal{A}$	
$r(\mathcal{A})$	number of regions of a real arrangement $\mathcal{A}$	
$b(\mathcal{A})$	number of relatively bounded regions of $\mathcal{A}$	
$\mathcal{A}_x$, $\mathcal{A}^x$	See Eqn. 5.6	
Shi_n	Shi arrangement	

Simplicial Complexes and Polytopes

$\langle \cdots \rangle$	simplicial complex generated by a set of faces	`\langle ... \rangle`
$\lvert\Delta\rvert$	(standard) geometric realization of Δ	
$\Delta(P)$	order complex of a poset P	
$\Bbbk[\Delta]$	Stanley-Reisner (face) ring of Δ over a field $\Bbbk$	`\fld[\Delta]`
$C_k(\Delta, R)$	simplicial chain groups over a ring R	
$\tilde{H}_k(\Delta, R)$	reduced simplicial homology groups	
$\mathbb{S}^d$	d-dimensional sphere	`\mathbb{S}`
P^*	dual of a polytope P	

Representation Theory

Id	identity element of a group	
D_n	dihedral group of symmetries of a regular n-gon	
ρ_{triv}, ρ_{reg}	trivial and regular representations	
ρ_{sign}, ρ_{def}	sign and defining representations of $\mathfrak{S}_n$	
χ_ρ or χ	character of representation ρ	
$C\ell(G)$	space of class functions of G	
$\langle \rho, \chi \rangle_G$	inner product on $C\ell(G)$ (see Thm. 8.6.5)	
$\mathrm{Hom}_{\mathbb{C}}(V, W)$	$\mathbb{C}$-linear maps $V \to W$	
$\mathrm{Hom}_G(V, W)$	G-equivariant $\mathbb{C}$-linear maps $V \to W$	
V^G	space of invariants of a G-action on V	
$\mathrm{Ch}(G)$	group of one-dimensional characters of G	
$[a, b]$	commutator: $aba^{-1}b^{-}1$	
$[G, G]$	commutator subgroup of G	
$\mathfrak{A}_n$	alternating group on n letters	`\mathfrak{A}_n`
$\mathsf{Par}(n)$	partitions of n	
C_λ	conjugacy class of permutations with cycle-type λ	
$\lambda < \mu$	lexicographic (total) order on partitions	
$\lambda \lhd \mu$, $\lambda \unlhd \mu$	dominance (partial) order on partitions	`\lhd`, `\unlhd`
$\mathrm{sh}(T)$	shape of a tabloid T	
(ρ_μ, V_μ)	tabloid representation of shape μ	
χ_μ	character of tabloid representation	
$K_{\lambda,\mu}$	Kostka numbers	
$\mathrm{Res}^G_H \rho$, $\mathrm{Res}^G_H \chi$	restricted representation/character	
$\mathrm{Ind}^G_H \rho$, $\mathrm{Ind}^G_H \chi$	induced representation/character	

Symmetric Functions

$\mathbf{x}^\alpha$	monomial in variables $\mathbf{x}$ with exponent vector α	
$R[[\mathbf{x}]]$	ring of formal power series in $\mathbf{x}$ with coefficients in R	
$[\mathbf{x}^\alpha]F$	coefficient of monomial $\mathbf{x}^\alpha$ in power series F	
m_λ	monomial symmetric function	
e_λ	elementary symmetric function	
h_λ	(complete) homogeneous symmetric function	
p_λ	power-sum symmetric function	
Λ_d, $\Lambda_{R,d}(\mathbf{x})$	R-module of degree-d symmetric functions in $\mathbf{x}$	
Λ, $\Lambda_R(\mathbf{x})$	R-algebra of symmetric functions in $\mathbf{x}$	
ω	involutory automorphism $\Lambda \to \Lambda$ swapping e's and h's	(not to be confused with w!)
$\mathsf{CST}(\lambda)$	set of column-strict tableaux of shape λ	
s_λ	Schur symmetric function	
Ω, Ω^*	Cauchy kernel and dual Cauchy kernel	
z_λ	size of centralizer of a partition of shape λ (see (9.16))	
ε_λ	sign of a partition of shape λ (see (9.16))	
$\mathsf{SYT}(\lambda)$	set of standard tableaux of shape λ	
f^λ	number of standard tableaux of shape λ	
$T \leftarrow x$	row-insertion (§9.10)	`T\gets x`
$\mathbf{ch}$	Frobenius characteristic (§9.11)	
$c^\lambda_{\mu,\nu}$	Littlewood–Richardson coefficients	
$h(x)$	length of the hook with corner x	

Combinatorial Algebraic Varieties

$\mathsf{Gr}(k, V)$	Grassmannian of k-dimensional subspaces of V
Ω_λ	Schubert cell in a Grassmannian
$F\ell(n)$	(complete) flag variety in dimension n
X_w	Schubert cell in a flag variety

Hopf Algebras and Monoids

μ	product
Δ	coproduct
u	unit
ε	counit
S	antipode

Bibliography

[AA17] Marcelo Aguiar and Federico Ardila, *Hopf monoids and generalized permutahedra*, preprint, arXiv:1709.07504, 2017. 138, 228

[ABS06] Marcelo Aguiar, Nantel Bergeron, and Frank Sottile, *Combinatorial Hopf algebras and generalized Dehn-Sommerville relations*, Compos. Math. **142** (2006), no. 1, 1–30. MR 2196760 228

[AM10] Marcelo Aguiar and Swapneel Mahajan, *Monoidal functors, species and Hopf algebras*, CRM Monograph Series, vol. 29, American Mathematical Society, Providence, RI, 2010, With forewords by Kenneth Brown and Stephen Chase and André Joyal. MR 2724388 228

[Ath96] Christos A. Athanasiadis, *Characteristic polynomials of subspace arrangements and finite fields*, Adv. Math. **122** (1996), no. 2, 193–233. MR 1409420 (97k:52012) 101, 102

[BB05] Anders Björner and Francesco Brenti, *Combinatorics of Coxeter Groups*, Graduate Texts in Mathematics, vol. 231, Springer, New York, 2005. MR 2133266 (2006d:05001) 17, 30

[BF01] Thomas Britz and Sergey Fomin, *Finite posets and Ferrers shapes*, Adv. Math. **158** (2001), no. 1, 86–127. MR 1814900 237

[BH93] Winfried Bruns and Jürgen Herzog, *Cohen-Macaulay Rings*, Cambridge Studies in Advanced Mathematics, vol. 39, Cambridge University Press, Cambridge, 1993. MR 1251956 (95h:13020) 117, 122

[Bjö94] Anders Björner, *Subspace arrangements*, First European Congress of Mathematics, Vol. I (Paris, 1992), Progr. Math., vol. 119, Birkhäuser, Basel, 1994, pp. 321–370. MR 1341828 108

[BLVS+99] Anders Björner, Michel Las Vergnas, Bernd Sturmfels, Neil White, and Günter M. Ziegler, *Oriented matroids*, second ed., Encyclopedia of Mathematics and its Applications, vol. 46, Cambridge University Press, Cambridge, 1999. MR 1744046 111

[BM71] H. Bruggesser and P. Mani, *Shellable decompositions of cells and spheres*, Math. Scand. **29** (1971), 197–205 (1972). MR 0328944 (48 #7286) 135

[BNP98] Marilena Barnabei, Giorgio Nicoletti, and Luigi Pezzoli, *Matroids on partially ordered sets*, Adv. in Appl. Math. **21** (1998), no. 1, 78–112. MR 1623325 71

[BO92] Thomas Brylawski and James Oxley, *The Tutte polynomial and its applications*, Matroid applications, Encyclopedia Math. Appl., vol. 40, Cambridge Univ. Press, Cambridge, 1992, pp. 123–225. MR 1165543 (93k:05060) 88

[Bol98] Béla Bollobás, *Modern Graph Theory*, Graduate Texts in Mathematics, vol. 184, Springer-Verlag, New York, 1998. MR 1633290 (99h:05001) 80

[BR15] Matthias Beck and Sinai Robins, *Computing the continuous discretely*, second ed., Undergraduate Texts in Mathematics, Springer, New York, 2015., Full text available free from https://matthbeck.github.io/ccd.html. MR 3410115 86, 138

[Bri73] Egbert Brieskorn, *Sur les groupes de tresses [d'après V. I. Arnol'd]*, Séminaire Bourbaki, 24ème année (1971/1972), Exp. No. 401, Springer, Berlin, 1973, pp. 21–44. Lecture Notes in Math., Vol. 317. 108

[BT82] Raoul Bott and Loring W. Tu, *Differential Forms in Algebraic Topology*, Graduate Texts in Mathematics, vol. 82, Springer-Verlag, New York-Berlin, 1982. MR 658304 (83i:57016) 246, 247

[CR70] Henry H. Crapo and Gian-Carlo Rota, *On the Foundations of Combinatorial Theory: Combinatorial Geometries*, preliminary ed., The M.I.T. Press, Cambridge, Mass.-London, 1970. MR 0290980 101

[Dir61] G.A. Dirac, *On rigid circuit graphs*, Abh. Math. Sem. Univ. Hamburg **25** (1961), 71–76. MR 0130190 (24 #A57) 107

[Edm70] Jack Edmonds, *Submodular functions, matroids, and certain polyhedra*, Combinatorial Structures and their Applications (Proc. Calgary Internat. Conf., Calgary, Alta., 1969), Gordon and Breach, New York, 1970, pp. 69–87. MR 0270945 138

[FS05] Eva Maria Feichtner and Bernd Sturmfels, *Matroid polytopes, nested sets and Bergman fans*, Port. Math. (N.S.) **62** (2005), no. 4, 437–468. MR 2191630 138

[Ful97] William Fulton, *Young Tableaux*, London Mathematical Society Student Texts, vol. 35, Cambridge University Press, Cambridge, 1997. 197, 216, 241

[GGMS87] I. M. Gel'fand, R. M. Goresky, R. D. MacPherson, and V. V. Serganova, *Combinatorial geometries, convex polyhedra, and Schubert cells*, Adv. in Math. **63** (1987), no. 3, 301–316. MR 877789 138

[GK76] Curtis Greene and Daniel J. Kleitman, *The structure of Sperner k-families*, J. Combin. Theory Ser. A **20** (1976), no. 1, 41–68. 237

[GNW79] Curtis Greene, Albert Nijenhuis, and Herbert S. Wilf, *A probabilistic proof of a formula for the number of Young tableaux of a given shape*, Adv. in Math. **31** (1979), no. 1, 104–109. MR 521470 210

[Gre76] Curtis Greene, *Some partitions associated with a partially ordered set*, J. Combin. Theory Ser. A **20** (1976), no. 1, 69–79. 237

[Grü03] Branko Grünbaum, *Convex Polytopes*, second ed., Graduate Texts in Mathematics, vol. 221, Springer-Verlag, New York, 2003, Prepared and with a preface by Volker Kaibel, Victor Klee and Günter M. Ziegler. 63, 129

[GSS93] Jack Graver, Brigitte Servatius, and Herman Servatius, *Combinatorial Rigidity*, Graduate Studies in Mathematics, vol. 2, American Mathematical Society, Providence, RI, 1993. MR 1251062 51

[Hat02] Allen Hatcher, *Algebraic Topology*, Cambridge University Press, Cambridge, 2002. MR 1867354 (2002k:55001) 117, 119, 121, 244

[HS15] Joshua Hallam and Bruce Sagan, *Factoring the characteristic polynomial of a lattice*, J. Combin. Theory Ser. A **136** (2015), 39–63. MR 3383266 107

[Hum90] James E. Humphreys, *Reflection Groups and Coxeter Groups*, Cambridge Studies in Advanced Mathematics, vol. 29, Cambridge University Press, Cambridge, 1990. MR 1066460 (92h:20002) 17

[Hum11] Brandon Humpert, *Polynomials Associated with Graph Coloring and Orientations*, Ph.D. thesis, University of Kansas, 2011. 227

[Kli] Caroline J. Klivans, *A quasisymmetric function for generalized permutahedra*, Talk slides, http://www.dam.brown.edu/people/cklivans/birs.pdf. 228

[LS00] Shu-Chung Liu and Bruce E. Sagan, *Left-modular elements of lattices*, J. Combin. Theory Ser. A **91** (2000), no. 1-2, 369–385. MR 1780030 49

[MR05] Jeremy L. Martin and Victor Reiner, *Cyclotomic and simplicial matroids*, Israel J. Math. **150** (2005), 229–240. MR 2255809 73

[MS05] Ezra Miller and Bernd Sturmfels, *Combinatorial Commutative Algebra*, Graduate Texts in Mathematics, vol. 227, Springer-Verlag, New York, 2005. MR 2110098 (2006d:13001) 117

[OS80] Peter Orlik and Louis Solomon, *Combinatorics and topology of complements of hyperplanes*, Invent. Math. **56** (1980), no. 2, 167–189. 108

[OT92] Peter Orlik and Hiroaki Terao, *Arrangements of Hyperplanes*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 300, Springer-Verlag, Berlin, 1992. 89

[Oxl92] James G. Oxley, *Matroid Theory*, Oxford Science Publications, The Clarendon Press Oxford University Press, New York, 1992. 51, 64, 72

[Per16] Xander Perrott, *Existence of projective planes*, arXiv:1603.05333 (retrieved 8/26/18), 2016. 63

[Pos09] Alexander Postnikov, *Permutohedra, associahedra, and beyond*, Int. Math. Res. Not. (2009), no. 6, 1026–1106. MR 2487491 138

[PRW08] Alex Postnikov, Victor Reiner, and Lauren Williams, *Faces of generalized permutohedra*, Doc. Math. **13** (2008), 207–273. MR 2520477 111, 138

[Rei] Victor Reiner, *Lectures on matroids and oriented matroids*, Lecture notes for Algebraic Combinatorics in Europe (ACE) Summer School in Vienna, July 2005; available at http://www-users.math.umn.edu/∼reiner/Talks/Vienna05/Lectures.pdf. 111, 115

[RGZ97] Jürgen Richter-Gebert and Günter M. Ziegler, *Oriented matroids*, Handbook of discrete and computational geometry, CRC Press Ser. Discrete Math. Appl., CRC, Boca Raton, FL, 1997, pp. 111–132. MR 1730162 111, 112

[Ryb11] G.L. Rybnikov, *On the fundamental group of the complement of a complex hyperplane arrangement*, Funktsional. Anal. i Prilozhen. **45** (2011), no. 2, 71–85. 108

[S+14] W. A. Stein et al., *Sage Mathematics Software (Version 6.4.1)*, The Sage Development Team, 2014, `http://www.sagemath.org`. 90

[Sag01] Bruce E. Sagan, *The Symmetric Group*, second ed., Graduate Texts in Mathematics, vol. 203, Springer-Verlag, New York, 2001. 186

[Sch13] Alexander Schrijver, *A Course in Combinatorial Optimization*, Available online at `http://homepages.cwi.nl/∼lex/files/dict.pdf` (retrieved 1/21/15), 2013. 129

[Sta95] Richard P. Stanley, *A symmetric function generalization of the chromatic polynomial of a graph*, Adv. Math. **111** (1995), no. 1, 166–194. MR 1317387 220

[Sta96] ———, *Combinatorics and Commutative Algebra*, second ed., Progress in Mathematics, vol. 41, Birkhäuser Boston Inc., Boston, MA, 1996. MR 1453579 (98h:05001) 117

[Sta99] ———, *Enumerative Combinatorics. Vol. 2*, Cambridge Studies in Advanced Mathematics, vol. 62, Cambridge University Press, Cambridge, 1999, With a foreword by Gian-Carlo Rota and appendix 1 by Sergey Fomin. MR 1676282 (2000k:05026) 177, 183, 191, 197, 203, 206, 251, 252

[Sta07] ———, *An introduction to hyperplane arrangements*, Geometric combinatorics, IAS/Park City Math. Ser., vol. 13, Amer. Math. Soc., Providence, RI, 2007, Also available online: http://www-math.mit.edu/~rstan/arrangements/arr.html, pp. 389–496. 89, 103, 106, 116

[Sta12] ———, *Enumerative Combinatorics. Volume 1*, second ed., Cambridge Studies in Advanced Mathematics, vol. 49, Cambridge University Press, Cambridge, 2012. MR 2868112 39, 138

[Sta15] ———, *Catalan Numbers*, Cambridge University Press, New York, 2015. MR 3467982 251, 252

[Tak71] Mitsuhiro Takeuchi, *Free Hopf algebras generated by coalgebras*, J. Math. Soc. Japan **23** (1971), 561–582. MR 0292876 226

[Tut54] W. T. Tutte, *A contribution to the theory of chromatic polynomials*, Canad. J. Math. **6** (1954), 80–91. MR 61366 82

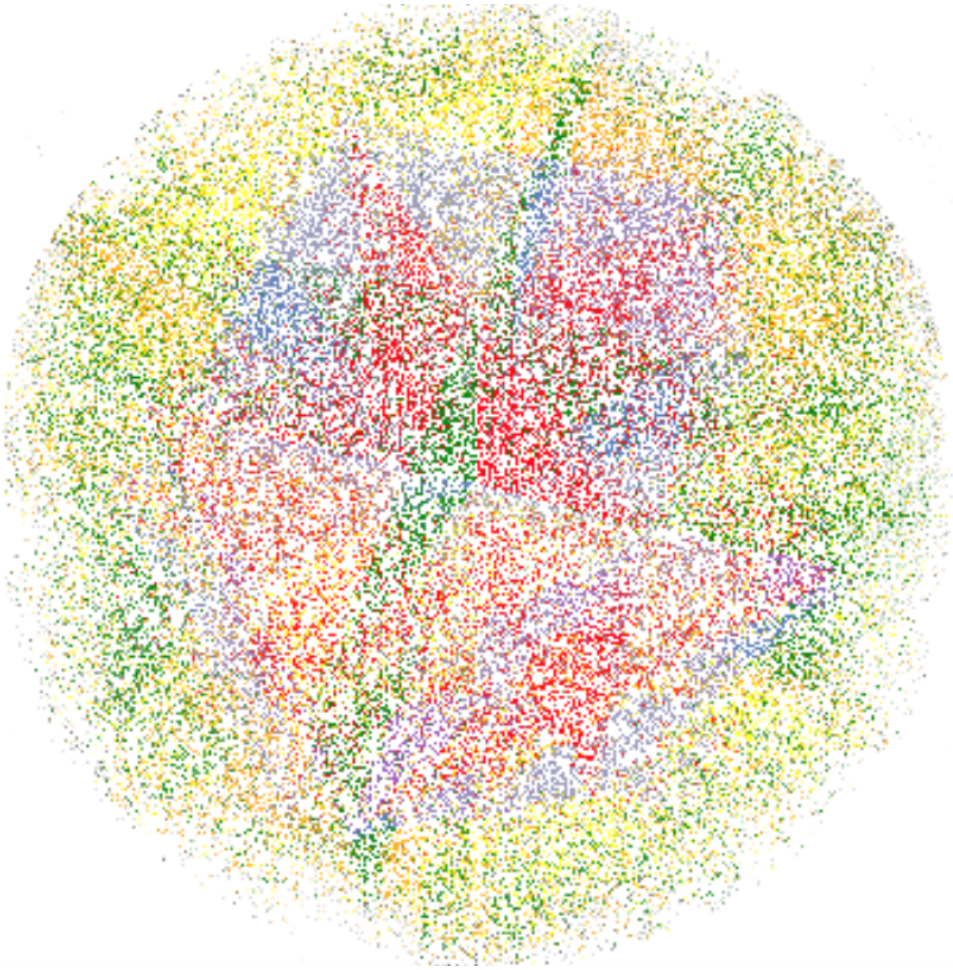
[Wes96] Douglas B. West, *Introduction to Graph Theory*, Prentice Hall, Inc., Upper Saddle River, NJ, 1996. MR 1367739 (96i:05001) 107

[Whi32a] Hassler Whitney, *The coloring of graphs*, Ann. of Math. (2) **33** (1932), no. 4, 688–718. MR 1503085 98

[Whi32b] ———, *A logical expansion in mathematics*, Bull. Amer. Math. Soc. **38** (1932), no. 8, 572–579. MR 1562461 98

[Zas75] Thomas Zaslavsky, *Facing up to arrangements: face-count formulas for partitions of space by hyperplanes*, Mem. Amer. Math. Soc. **1** (1975), no. issue 1, 154, vii+102. MR 0357135 95

[Zie95] Günter M. Ziegler, *Lectures on polytopes*, Graduate Texts in Mathematics, vol. 152, Springer-Verlag, New York, 1995. MR 1311028 129, 131, 137



A "pointillist" picture of the essentialized braid arrangement $\mathrm{ess}(\mathsf{Br}_4)$, produced by a computer glitch.