Lecture Notes on Algebraic Combinatorics

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

Copyright ©2010–2018 by Jeremy L. Martin (last updated November 5, 2018). Licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License.
# Contents

1 Posets ......................................... 6  
  1.1 The basics .................................. 6  
  1.2 Operations on posets ......................... 8  
  1.3 Ranked posets ................................ 9  
  1.4 Lattices .................................... 10  
  1.5 Exercises .................................. 16  

2 The Structure of Lattices ......................... 17  
  2.1 Distributive lattices ......................... 17  
  2.2 Modular lattices ............................. 20  
  2.3 Semimodular lattices ......................... 23  
  2.4 Geometric lattices ........................... 26  
  2.5 Exercises .................................. 28  

3 Poset Algebra .................................. 31  
  3.1 The incidence algebra of a poset ............... 31  
  3.2 The Möbius function ......................... 34  
  3.3 Möbius inversion ............................. 37  
  3.4 The characteristic polynomial ................... 41  
  3.5 Möbius functions of lattices ................. 41  
  3.6 Exercises .................................. 45  

4 Matroids ....................................... 48  
  4.1 Closure operators ............................. 48  
  4.2 Matroids and geometric lattices ............... 49  
  4.3 Graphic matroids ............................. 52  
  4.4 Matroid independence, basis and circuit systems 53  
  4.5 Representability and regularity ............... 58  
  4.6 Direct sum ................................ 62  
  4.7 Duality .................................... 64  
  4.8 Deletion and contraction ....................... 66  
  4.9 Exercises .................................. 68  

5 The Tutte Polynomial .......................... 70  
  5.1 The two definitions of the Tutte polynomial .... 70  
  5.2 Recipes .................................... 75  
  5.3 Basis activities ............................. 76  
  5.4 The characteristic and chromatic polynomials .... 78  
  5.5 Acyclic orientations ......................... 80  
  5.6 The Tutte polynomial and linear codes ........... 82
## 6 Hyperplane Arrangements

- **6.1 Basic definitions**
- **6.2 Counting regions: examples**
- **6.3 Zaslavsky’s theorem**
- **6.4 The finite field method**
- **6.5 Supersolvable lattices and arrangements**
- **6.6 Arrangements over \( \mathbb{C} \)**
- **6.7 Exercises**

## 7 Simplicial complexes

- **7.1 Basic definitions and terminology**
- **7.2 Simplicial homology**
- **7.3 Stanley-Reisner theory**
- **7.4 Shellable and Cohen-Macaulay simplicial complexes**
- **7.5 Exercises**

## 8 Polytopes

- **8.1 The basics**
- **8.2 Shelling simplicial polytopes**
- **8.3 Ehrhart theory (contributed by Margaret Bayer)**
- **8.4 Exercises**

## 9 Group Representations

- **9.1 Basic definitions**
- **9.2 Isomorphisms and homomorphisms**
- **9.3 Irreducibility, indecomposability and Maschke’s theorem**
- **9.4 Characters**
- **9.5 New characters from old**
- **9.6 The inner product on class functions; Schur’s Lemma**
- **9.7 The fundamental theorem of character theory for finite groups**
- **9.8 Computing character tables**
- **9.9 One-dimensional characters**
- **9.10 Restriction, induction, and Frobenius reciprocity**
- **9.11 Characters of the symmetric group**
- **9.12 Exercises**

## 10 Symmetric Functions

- **10.1 Prelude: Symmetric polynomials**
- **10.2 Formal power series**
- **10.3 Symmetric functions**
- **10.4 Elementary symmetric functions**
- **10.5 Complete homogeneous symmetric functions**
- **10.6 Power-sum symmetric functions**
- **10.7 Schur functions**
- **10.8 The Cauchy kernel and the Hall inner product**
- **10.9 The RSK Correspondence**
- **10.10 Knuth equivalence and jeu de taquin**
- **10.11 Yet another version of RSK**
- **10.12 The Frobenius characteristic**
- **10.13 Skew tableaux and the Littlewood-Richardson Rule**
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, Dylan Beck, Lucas Chaffee, Geoffrey Critzer, Mark Denker, Joseph Doolittle, Ken Duna, Josh Fenton, Logan Godkin, Bennet Goeckner, Darij Grinberg (especially!), Brent Holmes, Arturo Jaramillo, Alex Lazar, Nick Packauskas, Abraham Pascoe, Smita Praharaj, John Portin, Billy Sanders, Tony Se, and Amanda Wilkens. Marge Bayer contributed the material on Ehrhart theory in §8.3.
Chapter 1

Posets

1.1 The basics

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 \( x \prec y \), if \( x < y \) and there exists no \( z \) such that \( x < z < y \). Two posets \( P, Q \) are isomorphic if there is a bijection \( \phi : P \rightarrow Q \) that is order-preserving; that is, \( x \leq y \) in \( P \) iff \( \phi(x) \leq \phi(y) \) in \( Q \).

We'll 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'll say what “chains” and “intervals” are soon.)

Definition 1.1.2. A poset \( L \) is a lattice if every pair \( x, y \in L \) (i) has a unique largest common lower bound, called their meet and written \( x \wedge y \); (ii) has a unique smallest common upper bound, called their join and written \( x \vee y \). That is, for all \( z \in L \),

\[
\begin{align*}
z \leq x \text{ and } z \leq y \Rightarrow z \leq x \wedge y, \\
z \geq x \text{ and } z \geq y \Rightarrow z \geq x \vee y,
\end{align*}
\]

We'll have a lot more to say about lattices in Section 2.

Example 1.1.3 (Boolean algebras). 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 algebra of rank \( n \), denoted here by the symbol \( \mathcal{B}_n \). We may also use \( \mathcal{B}_S \) for the Boolean algebra of subsets of any finite set \( S \); clearly \( \mathcal{B}_S \cong \mathcal{B}_n \). The cardinality of \( S \) is called the rank of \( \mathcal{B}_S \); it is not hard to see that every Boolean algebra is determined up to isomorphism by its rank.
Note that \(2^{[n]}\) is a lattice, with meet and join given by intersection and union respectively.

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 algebra \(2^S\) 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 \(\hat{0}\) and \(\hat{1}\) respectively (or if necessary \(\hat{0}_P\) and \(\hat{1}_P\)).

**Definition 1.1.4.** A poset that has both a \(\hat{0}\) and a \(\hat{1}\) is called **bounded**. An element that covers \(\hat{0}\) is called an **atom**, and an element that is covered by \(\hat{1}\) 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 \(\hat{0}, \hat{1}\) such that \(\hat{0} < x < \hat{1}\) 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

\[
[x, y] := \{z \in P \mid x \leq z \leq y\}.
\]

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 \(x_0 < \cdots < x_n\); equivalently, \(C\) is maximal among all chains with bottom element \(x_0\) and top element \(x_n\). 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 algebra \(B_3\), 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 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. This concept does not typically arise in combinatorics, where one frequently wants to talk about antichains in a bounded posets.

\(^3\)It is very common to drop the braces and commas from subsets of \([n]\), since it is easier and cleaner to write \(\{\emptyset, 3, 123\}\) rather than \(\{\emptyset, \{3\}, \{1, 2, 3\}\}\).
Definition 1.1.7. 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 $\mathcal{L}(P)$ (and sometimes called the Jordan-Hölder set of $P$).

If $P$ is a chain then $\mathcal{L}(P) = \{P\}$, while if $P$ is an antichain then $\mathcal{L}(P) = \Sigma_P$, the set of all permutations (= linear orders) of $P$. In general, the more relations $P$ has, the fewer linear extensions.

Definition 1.1.8. A subposet of $P$ is a subset $Q \subseteq P$ with the same order relation.

Definition 1.1.9. 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 $\langle Q \rangle = \{x \in P \mid x \leq q \text{ for some } q \in Q\}$. Conversely, every order ideal has a unique minimal set of generators, namely its maximal elements (which form an antichain).

Example 1.1.10. Let $\{F_1, \ldots, F_k\}$ be a nonempty family of subsets of $[n]$. The order ideal they generate is

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

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 $\Delta$ contains a simplex, then it contains all sub-simplices. For example, $\Delta$ cannot contain a triangle without also containing its edges and vertices. Simplicial complexes are the fundamental objects of topological combinatorics, and we’ll have much more to say about them in Section 7.

1.2 Operations on posets

Definition 1.2.1. 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 $\mathcal{B}_n$ (via the anti-automorphism $S \mapsto [n] \setminus S$).
- 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 \(B_k \times B_{\ell} \cong B_{k+\ell}\).

Assume that \(P\) has a \(\hat{1}\) and \(Q\) has a \(\hat{0}\). Then the ordinal sum \(P \oplus Q\) is defined by identifying \(\hat{1}_P = \hat{0}_Q\) and setting \(p \leq q\) for all \(p \in P\) and \(q \in Q\).

\[
\text{Figure 1.1: Direct product } \times \text{ and ordinal sum } \oplus.
\]

1.3 Ranked posets

One of the many nice properties of the Boolean algebra \(\mathcal{B}_n\) is that its elements fall into horizontal slices (sorted by their cardinalities). Whenever \(S \prec T\), 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\(^4\) is a little more subtle, but makes perfect sense after a little thought.

**Definition 1.3.1.** Let \(P\) be a poset. For each \(x \in P\), let \(C(x)\) denote the family of chains in \(P\) with top element \(x\). We say that \(P\) is ranked if for every \(x \in P\), all maximal\(^5\) chains in \(C(x)\) have the same length. In this case, we can define a rank function \(r : P \rightarrow \mathbb{N}\) as follows: \(r(x)\) is the length of any (hence every) maximal chain with top element \(x\). Thus \(r(x) = 0\) if and only if \(x\) is a minimal element, and it is easy to show that

\[
x \prec y \implies r(y) = r(x) + 1.
\]

A poset is graded if it is ranked and bounded.

Some notes on these definitions:

1. 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.
2. The literature is not consistent on the usage of the term “ranked”. Sometimes “ranked” is used for the weaker condition that for every pair \(x, y \in P\), all saturated chains from \(x\) to \(y\) have the same length. For example, the poset shown below is ranked in this weaker sense, but is not ranked in the sense of Definition 1.3.1 since \(\{w, x, y\}\) and \(\{z, y\}\) are both maximal elements of \(C_y\).

\(^4\)Which my spell-checker keeps trying to autocorrect to “rancidness”.

\(^5\)In general “maximal” means “maximal under inclusion,” as opposed to “maximum” which means “of the greatest possible size”. In this context, misreading “maximal” as “maximum” would lead to a tautology.
There is no way to equip this poset with a rank function such that both minimal elements have rank 0 and (1.3.1) holds. On the other hand, if $P$ has a unique minimal element, then the strong and weak definitions of rankedness are equivalent.

3. Rankedness in the sense of Definition 1.3.1 is not a self-dual condition; the dual of the poset shown above is ranked. (Weak rankedness, on the other hand, is self-dual.)

4. For any finite poset $P$ (and some infinite ones) one can define a pseudo-rank function $r(x)$ to be the supremum of the lengths of all chains with top element $x$ — but if $P$ is not a ranked poset, then there will be some pair $x, y$ such that $y \succ x$ but $r(y) > r(x) + 1$. For instance, in the bounded poset (known as $N_5$) shown below, $\hat{1} \succ y$ but $r(\hat{1}) = 3$ and $r(y) = 1$.

**Definition 1.3.2.** 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 $k$.

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

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

The expansion of this polynomial is palindromic, because the coefficients are a row of Pascal’s Triangle. That is, $\mathcal{B}_n$ is rank-symmetric. Rank-symmetry also follows from the self-duality of $\mathcal{B}_n$.

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 $F_{P \times Q} = F_P F_Q$.

### 1.4 Lattices

**Definition 1.4.1.** A poset $L$ is a lattice if every pair $x, y \in L$ has a unique meet $x \land y$ and join $x \lor y$. That is,

$$x \land y = \max \{ z \in L \mid z \leq x \text{ and } z \leq y \},$$

$$x \lor y = \min \{ z \in L \mid z \geq x \text{ and } z \geq y \}.$$
Note that, e.g., \( x \land y = x \) 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 \( \land M \) and join \( \lor M \) are well-defined elements of \( L \). In particular, every finite lattice is bounded, with \( \bot = \land L \) and \( \top = \lor L \). For convenience, we set \( \land \emptyset = \top \) and \( \lor \emptyset = \bot \).

**Example 1.4.2 (The partition lattice).** Let \( \Pi_n \) be the poset of all set partitions of \([n]\). E.g., two elements of \( \Pi_5 \) are
\[
\pi = \{\{1,3,4\}, \{2,5\}\} \quad \text{ (abbr.: 134|25)}
\]
\[
\sigma = \{\{1,3\}, \{4\}, \{2,5\}\} \quad \text{ (abbr.: 13|4|25)}
\]
The sets \( \{1,3,4\} \) and \( \{2,5\} \) are called the blocks of \( \pi \). We can impose a partial order on \( \Pi_n \) by putting \( \sigma \leq \pi \) if every block of \( \sigma \) is contained in a block of \( \pi \); for short, \( \sigma \) refines \( \pi \).

![Diagram of \( \Pi_3 \) and \( \Pi_4 \)]

Observe that \( \Pi_n \) is bounded, with \( \emptyset = 1|2|\cdots|n \) and \( \top = 12\cdots n \). For each partition \( \sigma \) in \( \Pi_n \), the partitions that cover \( \sigma \) in \( \Pi_n \) are those obtained from \( \sigma \) by merging two of its blocks into a single block. Therefore, \( \Pi_n \) is ranked (hence graded), with rank function \( r(\pi) = n - |\pi| \). The coefficients of the rank-generating function of \( \Pi_n \) 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, \( \Pi_n \) is a lattice. The meet of two partitions is their coarsest common refinement: \( x, y \) belong to the same block of \( \pi \land \sigma \) if and only if they belong to the same block of \( \pi \) and to the same block of \( \sigma \). The join is the transitive closure of the union of the equivalence relations corresponding to \( \pi \) and \( \sigma \).

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

**Example 1.4.3 (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 \( \Pi_V \) defined by
\[
K(G) = \{ \pi \in \Pi_V \mid G|_X \text{ is connected for every block } X \in \pi \}.
\]
For an example, see Figure 1.2. 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.

**Example 1.4.4 (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, set $\lambda_i = 0$ for all $i > \ell$. 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’ll shortly see another description of the lattice operations.

This is an infinite poset, but the number of partitions at any given rank is finite. So in particular $Y$ is locally finite (if $X$ is any adjective, then “poset $P$ is locally $X$” means “every interval in $P$ is $X$”). 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 $i^{th}$ row contains $\lambda_i$ 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 $\lambda$ contains that of $\mu$. 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 $\tilde{\lambda} \geq \tilde{\mu}$.

**Example 1.4.5 (The subspace lattice).** 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 any old field if you don’t mind infinite posets.)

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, $L_2(2) \cong M_5$.

\[
\begin{array}{c}
\circ \\
M_5
\end{array}
\]

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).

**Example 1.4.6.** Lattices don’t have to be ranked. For example, the poset $N_5$ shown below is a perfectly good lattice.

\[
\begin{array}{c}
\circ \\
M_5
\end{array}
\]
Proposition 1.4.7 (Absorption laws). Let $L$ be a lattice and $x, y \in L$. Then $x \lor (x \land y) = x$ and $x \land (x \lor y) = x$. (Proof left to the reader.)

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

Proposition 1.4.8. Let $P$ be a bounded poset that is a meet-semilattice (i.e., every nonempty $B \subseteq P$ has a well-defined meet $\land B$). Then every finite 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 $B = \{b \in P \mid b \geq a \text{ for all } a \in A\}$. Note that $B \neq \emptyset$ because $\hat{1} \in B$. Then $\land B$ is the unique least upper bound for $A$, for the following reasons. First, $\land B \geq a$ 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 $x \geq \land B$. 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 1.4.9. 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

$$x \land_{L'} y = x \land_L y \text{ and } x \lor_{L'} y = x \lor_L y \text{ for all } x, y \in L'.$$

Equivalently, a sublattice of $L$ is a subset that is closed under meet and join.

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., $L' = \{y \in L \mid x \leq y \leq z\}$) is a sublattice with minimum element $x$ and maximum element $z$. (We might write $0_{L'} = x$ and $1_{L'} = z$.)

Example 1.4.10. Young’s lattice $Y$ is an infinite lattice. Meets of arbitrary sets are well-defined, as are finite joins. There is an $\hat{0}$ element (the empty Ferrers diagram), but no $\hat{1}$. 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 $\hat{0}$ element.

Example 1.4.11. Consider the set $M = \{A \subseteq [4] : A$ has even size$\}$. This is a lattice, but it is not a sublattice of $\mathcal{B}_4$, because for example $12 \land_M 13 = \emptyset$ while $12 \land_{\mathcal{B}_4} 13 = 1$.

Example 1.4.12. [Weak Bruhat order] Let $\mathfrak{S}_n$ be the set of permutations of $[n]$ (i.e., the symmetric group). Write elements $w \in \mathfrak{S}_n$ as strings $w_1w_2\cdots w_n$ of distinct digits, e.g., $47182635 \in \mathfrak{S}_8$. This is

\[\text{That’s a Fraktur S, obtainable in LaTeX as \texttt{\textbackslash mathfrak\{S\}}. The letter S has many other standard uses in combinatorics: Stirling numbers of the first and second kind, Schur symmetric functions, etc. The symmetric group is important enough to merit an ornate symbol!}\]
called one-line notation.) The weak Bruhat order on $S_n$ is defined as follows: $w \preceq v$ if $v$ can be obtained by swapping $w_i$ with $w_{i+1}$, where $w_i < w_{i+1}$. For example,

$$47162835 \preceq 47168235 \quad \text{and} \quad 47162835 \succeq 47162835.$$  

In other words, $s_i < w_{i+1}$ and $v = ws_i$, 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.

The Bruhat order on permutations is a related partial order with more relations (i.e., “stronger”) than the weak order. We first need the notion of inversions: an inversion of $w \in S_n$ is an ordered pair $(i,j)$ such that $i < j$ and $w_i > w_j$. The number of inversions is written $\text{inv}(w)$. The simplest way of describing Bruhat order is as follows: $w \preceq v$ if $\text{inv}(v) = \text{inv}(w) + 1$ and $v = wt$ for some transposition $t$. For example,

$$47162835 \preceq 47182635$$

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, 47162835 is not covered by 47862135 because this transposition increases the inversion number by 5, not by 1.

The Bruhat and weak orders on $S_3$ are shown below. You should be able to see from the picture that Bruhat order is not a lattice.

A Coxeter group is a finite group generated by elements $s_1, \ldots, s_n$, called simple reflections, satisfying $s_i^2 = 1$ and $(s_is_j)^{m_{ij}} = 1$ for all $i \neq j$ and some integers $m_{ij} \geq 23$. For example, setting $m_{ij} = 3$ if $|i - j| = 1$ and $m_{ij} = 2$ if $|i - j| > 1$, we obtain the symmetric group $S_{n+1}$. 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 [Hum09], 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 $S_n$, with the same rank function, namely the number of inversions:

$$r(w) = \left| \{(i,j) \mid i < j \text{ and } w_i > w_j \} \right|.$$  

(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 $S_n$ is a very nice polynomial called the $q$-factorial:

$$F_{S_n}(q) = 1(1 + q)(1 + q + q^2) \cdots (1 + q + \cdots + q^{n-1}) = \prod_{i=1}^{n} \frac{1 - q^i}{1 - q}.$$  

15


1.5 Exercises

Exercise 1.1. (a) Prove that every nonempty interval in a Boolean algebra is itself isomorphic to a Boolean algebra.
(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 \( \Pi_n \) is isomorphic to a partition lattice. (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 \( (x_1, \ldots, x_k) \leq (y_1, \ldots, y_k) \) if \( x_i \leq_{P_i} y_i \) for all \( i \in [k] \).)

Exercise 1.2. A directed acyclic graph (or DAG) is a pair \( \mathbf{G} = (\mathbf{V},\mathbf{E}) \), where \( \mathbf{V} \) is a set of vertices; \( \mathbf{E} \) is a set of edges, each of which is an ordered pair of distinct vertices; and \( \mathbf{E} \) contains no directed cycles, i.e., no subsets of the form \( \{(v_1,v_2), (v_2,v_3), \ldots, (v_{n-1},v_n), (v_n,v_1)\} \) for any \( v_1, \ldots, v_n \in \mathbf{V} \).

(a) Let \( \mathbf{P} \) be a poset with order relation \( < \). Let \( \mathbf{E} = \{(v,w) \mid v, w \in \mathbf{P}, v < w\} \). Prove that the pair \( (\mathbf{P},\mathbf{E}) \) is a DAG.
(b) Let \( \mathbf{G} = (\mathbf{V},\mathbf{E}) \) be a DAG. Define a relation \( < \) on \( \mathbf{V} \) by setting \( v < w \) iff there is some directed path from \( v \) to \( w \) in \( \mathbf{G} \), i.e., iff \( \mathbf{E} \) has a subset of the form \( \{(v = v_1, v_2), (v_2, v_3), \ldots, (v_{n-1}, v_n = w)\} \) with all \( v_i \) distinct. Prove that this relation makes \( \mathbf{V} \) into a poset.

(This problem is purely a technical exercise and is almost tautological, but it does show that posets and DAGs are essentially the same thing.)

Exercise 1.3. Recall from Definition 1.1.7 that \( \mathcal{L}(\mathbf{P}) \) means the set of linear extensions of a poset \( \mathbf{P} \).

(a) Let \( \mathbf{P} \) and \( \mathbf{Q} \) be posets. Describe \( \mathcal{L}(\mathbf{P} + \mathbf{Q}) \) and \( \mathcal{L}(\mathbf{P} \oplus \mathbf{Q}) \) in terms of \( \mathcal{L}(\mathbf{P}) \) and \( \mathcal{L}(\mathbf{Q}) \). (Hint: Start by working out some small examples explicitly. The problem is nontrivial even when \( \mathbf{P} \) and \( \mathbf{Q} \) are both chains of length 1.)
(b) Give a concrete combinatorial description of \( \mathcal{L}(\mathcal{B}_n) \).

Exercise 1.4. Let \( n \) be a positive integer. Let \( \mathbf{D}_n \) be the set of all positive-integer divisors of \( n \) (including \( n \) itself), partially ordered by divisibility.

(a) Prove that \( \mathbf{D}_n \) is a ranked poset, and describe the rank function.
(b) For which values of \( n \) is \( \mathbf{D}_n \) (i) a chain; (ii) a Boolean algebra? For which values of \( n, m \) is it the case that \( \mathbf{D}_n \cong \mathbf{D}_m \)?
(c) Prove that \( \mathbf{D}_n \) is a distributive lattice, i.e., a lattice such that \( x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z) \) for all \( x, y, z \in \mathbf{D}_n \). Describe its meet and join operations and its join-irreducible elements.
(d) Prove that \( \mathbf{D}_n \) is self-dual, i.e., there is a bijection \( f : \mathbf{D}_n \to \mathbf{D}_n \) such that \( f(x) \leq f(y) \) if and only if \( x \geq y \).

Exercise 1.5. Let \( \mathbf{G} \) be a graph on vertex set \( V = [n] \). Recall from Example 1.4.3 that the connectivity lattice of a graph is the subposet \( K(\mathbf{G}) \) of \( \Pi_n \) consisting of set partitions in which every block induces a connected subgraph of \( \mathbf{G} \). Prove that \( K(\mathbf{G}) \) is a lattice. Is it a sublattice of \( \Pi_n \)?
Chapter 2

The Structure of Lattices

2.1 Distributive lattices

Definition 2.1.1. A lattice $L$ is **distributive** if the following two equivalent conditions hold:

$$x \land (y \lor z) = (x \land y) \lor (x \land z) \quad \forall x, y, z \in L,$$

$$x \lor (y \land z) = (x \lor y) \land (x \lor z) \quad \forall x, y, z \in L.$$  

(2.1.1a) \hspace{2cm} (2.1.1b)

Proving that the two conditions (2.1.1a) and (2.1.1b) are equivalent is not too hard, but is not trivial (Exercise 1.4). 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 algebra $2^{[n]}$ 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:

   $\begin{align*}
   (x \lor y) \land z &= \hat{1} \land z = z \\
   (x \land z) \lor (y \land z) &= x \lor \hat{0} = x
   \end{align*}$

4. The partition lattice $\Pi_n$ is not distributive for $n \geq 3$, because $\Pi_3 \cong M_5$, and for $n \geq 4$ every $\Pi_n$ contains a sublattice isomorphic to $\Pi_3$ (see Exercise 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 (proof for homework).

Every poset $P$ gives rise to a distributive lattice in the following way. The set $J(P)$ of order ideals of $P$ (see Definition 1.1.9) is itself a 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 $\mathcal{B}_P$. (See Figure 2.1 for an example.)

![Figure 2.1: A poset $P$ and the corresponding distributive lattice $J(P)$.](image)

For example, if $P$ is an antichain, then every subset is an order ideal, so $J(P) = \mathcal{B}_P$, 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 2.1.2.** There is a natural bijection between $J(P)$ and the set of antichains of $P$, since the maximal elements of any order ideal $A$ 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 2.1.7), 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 2.1.3.** 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 $x = y \lor z$ then either $x = y$ or $x = z$. The subposet (not sublattice!) of $L$ consisting of all join-irreducible elements is denoted $\text{Irr}(L)$. Here is an example.

![image](image)

Equivalently, an element of $L$ is join-irreducible if it covers exactly one other element. (Exercise; not hard.)
Definition 2.1.4. A factorization of $x \in L$ is an equation of the form

$$x = p_1 \lor \cdots \lor p_n$$

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

Provided that $L$ has no infinite descending chains (e.g., $L$ is finite, or is locally finite and has a $\hat{0}$), 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).

For general lattices, irredundant factorizations need not be unique. For example, the $\hat{1}$ 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 2.1.6).

Proposition 2.1.5. Let $L$ be a distributive lattice and let $p \in \text{Irr}(L)$. Suppose that $p \leq q_1 \lor \cdots \lor q_n$. Then $p \leq q_i$ for some $i$.

Proof. By distributivity,

$$p = p \land (q_1 \lor \cdots \lor q_n) = (p \land q_1) \lor \cdots \lor (p \land q_n)$$

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

Proposition 2.1.5 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 2.1.6 (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

$$x = p_1 \lor \cdots \lor p_n = q_1 \lor \cdots \lor q_m$$

with $p_i, q_j \in \text{Irr}(L)$ for all $i, j$. Then $p_i \leq x = q_1 \lor \cdots \lor q_m$, so by Proposition 2.1.5, $p_i \leq q_j$ for some $j$. Again by Proposition 2.1.5, $q_j \leq p_i$ for some $i$. If $i \neq 1$, then $p_1 \leq p_i$, 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 (2.1.2) must be identical. \qed

Theorem 2.1.7 (Birkhoff 1933). Up to isomorphism, the finite distributive lattices are exactly the lattices $J(P)$, where $P$ is a finite poset. Moreover, $L \cong J(\text{Irr}(L))$ for every lattice $L$ and $P \cong \text{Irr}(J(P))$ for every poset $P$.

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

$$\phi(x) = \{p \in \text{Irr}(L) \mid p \leq x\}.$$ 

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 $P \to \text{Irr}(J(P))$ is given by

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

These facts need to be checked; the details are left to the reader (Exercise 2.8).
Corollary 2.1.8. 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 2.1.2, and it is bounded with \( \hat{0} = \emptyset \) and \( \hat{1} = P \). \hfill \Box

Corollary 2.1.9. Let \( L \) be a finite distributive lattice. The following are equivalent:

1. \( L \) is a Boolean algebra.
2. \( \text{Irr}(L) \) 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 \( x \lor \bar{x} = \hat{1} \) and \( x \land \bar{x} = \hat{0} \).
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 \( x \lor u = z \) and \( x \land u = y \).

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

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

\[
(x \lor \bar{x}) \land y = \hat{1} \land y = y \\
(x \lor \bar{x}) \land y = (x \land y) \lor (\bar{x} \land y) = x \lor (\bar{x} \land y)
\]

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

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

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

(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 \( Y \cup U = Z \). \hfill \Box

Join and meet could have been interchanged throughout this section. For example, the dual of Proposition 2.1.6 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.

2.2 Modular lattices

Definition 2.2.1. A lattice \( L \) is modular if every \( x, y, z \in L \) with \( x \leq z \) satisfy the modular equation:

\[
x \lor (y \land z) = (x \lor y) \land z.
\] (2.2.1)

Note that for all lattices, if \( x \leq z \), then \( x \lor (y \land z) \leq (x \lor y) \land z \). Modularity says that, in fact, equality holds.
The term “modularity” arises in algebra: we will shortly prove that $L_n(q)$ is always modular, as well as more generally the poset of modules over any ring, ordered by inclusion (Corollary 2.2.3).

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
   \[ x \lor (y \land z) = (x \lor y) \land (x \lor z) = (x \lor y) \land z. \]
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.

\begin{center}
\begin{tikzpicture}
  \node (x) at (0,0) [circle,fill] {}; \node (y) at (1,0) [circle,fill] {}; \node (z) at (2,0) [circle,fill] {}; \node (w) at (1,1) [circle,fill] {};
  \draw (x) -- (y) -- (z) -- (w) -- (x);
  \node at (0,-1) {$x \lor y$}; \node at (1,-1) {$y \land z$}; \node at (2,-1) {$z$}; \node at (1,1) {$z$};
  \node at (0,-2) {Modular}; \node at (2,-2) {Non-modular};
\end{tikzpicture}
\end{center}

Here $x \leq z$, but

\[ x \lor (y \land z) = x \lor \hat{0} = x, \]
\[ (x \lor y) \land z = \hat{1} \land z = z. \]

In fact, $N_5$ is the unique obstruction to modularity, as we will soon see (Thm. 2.2.5).

5. The nondistributive lattice $M_5 \cong \Pi_3$ is modular. However, $\Pi_4$ is not modular (exercise).

**Theorem 2.2.2.** [Characterizations of modularity] Let $L$ be a lattice. Then the following are equivalent:

(a) $L$ is modular.
(b) For all $x, y, z \in L$, if $x \in [y \land z, z]$, then $x = (x \lor y) \land z$.
(c) For all $x, y, z \in L$, if $x \in [y, y \lor z]$, then $x = (x \land z) \lor y$.
(d) For all $y, z \in L$, the lattices $L' = [y \land z, z]$ and $L'' = [y, y \lor z]$ are isomorphic, via the maps
   \[
   \alpha : L' \to L'' \quad \beta : L'' \to L'
   \]
   \[
   q \mapsto q \lor y, \quad p \mapsto p \land z.
   \]

*Proof.* (a) \implies (b): If $y \land z \leq x \leq z$, then the modular equation $x \lor (y \land z) = (x \lor y) \land z$ reduces to $x = (x \lor y) \land z$. 

21
(b) \implies (a): Suppose that (b) holds. Let \( a, b, c \in L \) with \( a \leq c \). Then
\[
b \land c \leq a \lor (b \land c) \leq c \lor c = c
\]
so applying (b) with \( y = b, z = c, x = a \lor (b \land c) \) gives
\[
a \lor (b \land c) = (a \lor (b \land c)) \lor b \land c = (a \lor b) \land c
\]
which is the modular equation for \( a, b, c \).

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

(b)+(c) \iff (d): The functions \( \alpha \) and \( \beta \) are always order-preserving functions with the stated domains and ranges. Conditions (b) and (c) say respectively that \( \beta \circ \alpha \) and \( \alpha \circ \beta \) are the identities on \( L' \) and \( L'' \); together, these conditions are equivalent to condition (d).

**Corollary 2.2.3.** 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 \( A \lor B = A + B \) and \( A \land B = A \cap B \).

*Proof.* The Second Isomorphism Theorem says that \( B/(A \cap B) \cong (A + B)/A \) for all \( A,B \in L(M) \). Therefore \( L(B/(A \cap B)) \cong L((A + B)/A) \), which says that \( L(M) \) satisfies condition (d) of Theorem 2.2.2. \( \square \)

In particular, the lattices \( L_n(q) \) are modular.

**Example 2.2.4.** For a (finite) group \( G \), let \( L(G) \) denote the lattice of subgroups of \( G \), with operations \( H \land K = H \cap K \) and \( H \lor K = HK \) (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, consider the symmetric group \( \mathfrak{S}_4 \). 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 \( N_5 \):

\[ \begin{array}{c}
\mathfrak{A}_4 \\
\mathfrak{S}_4 \\
\langle (1 2 3) \rangle \\
\{\text{Id}\}
\end{array} \]

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

**Theorem 2.2.5.** 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

\[
(x \lor y) \land z
\]

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

\[
x \land (y \lor z) > (x \land y) \lor (x \land z).
\]

You can then show that

1. this inequality is invariant under permuting \( x, y, z \);
2. \( (x \land (y \lor z)) \lor (y \land z) \) and the two other lattice elements obtained by permuting \( x, y, z \) form a cochain;
3. \( x \lor y = x \lor z = y \lor z \), 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.

### 2.3 Semimodular lattices

Recall that the notation \( x \preceq y \) 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 2.3.1.** A lattice \( L \) is (upper) **semimodular** if for all incomparable \( x, y \in L \),

\[
x \land y \preceq y \quad \Rightarrow \quad x \preceq x \lor y.
\]

Conversely, \( L \) is **lower semimodular** if the converse holds.
Note that $L$ is upper semimodular if and only if its dual $L^*$ is lower semimodular, and that the implication (2.3.1) 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 $[x \land y, x \lor y] \subseteq L$ pictured below, if the southeast relation is a cover, then so is the northwest relation.

![Diagram](image)

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

**Lemma 2.3.2.** If $L$ is modular then it is upper and lower semimodular.

**Proof.** If $x \land y \leq y$, then the sublattice $[x \land y, y]$ has only two elements. If $L$ is modular, then condition (d) of the characterization of modularity (Theorem 2.2.2) implies that $[x \land y, y] \cong [x, x \lor y]$, so $x \preceq x \lor y$. 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

\[
\begin{align*}
r(x \lor y) + r(x \land y) &\leq r(x) + r(y) \quad \text{if } L \text{ is USM}, \\
r(x \lor y) + r(x \land y) &\geq r(x) + r(y) \quad \text{if } L \text{ is LSM}, \\
r(x \lor y) + r(x \land y) &= r(x) + r(y) \quad \text{if } L \text{ is modular}.
\end{align*}
\]

**Lemma 2.3.3.** Suppose $L$ is semimodular and let $q, r, s \in L$. If $q \preceq r$, then either $q \lor s = r \lor s$ or $q \lor s \preceq r \lor s$.

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 $q \lor s$ to $r \lor s$.

**Proof.** Let $p = (q \lor s) \land r$, so that $q \preceq p \preceq r$. Since $q$ is covered by $r$, it follows that either $p = q$ or $p = r$.

- If $p = r$, then $q \lor s \geq r$. So $q \lor s = r \lor (q \lor s) = r \lor s$.
- If $p = q$, then $p = (q \lor s) \land r = q \preceq r$. Applying semimodularity to the diamond figure below, we obtain $(q \lor s) \preceq (q \lor s) \lor r = r \lor s$.

![Diagram](image)
Theorem 2.3.4. Let \( L \) be a lattice. Then \( L \) is semimodular if and only if it is ranked, with rank function \( r \) satisfying the semimodular inequality

\[
    r(x \lor y) + r(x \land y) \leq r(x) + r(y) \quad \forall x, y \in L.
\] (2.3.2)

Proof. ( \( \iff \) ) Suppose that \( L \) is a ranked lattice with rank function \( r \) satisfying (2.3.2). Suppose that \( x, y \) are incomparable and \( x \land y \leq y \). Incomparability implies \( x \lor y > x \). On the other hand, \( r(y) = r(x \land y) + 1 \), so by (2.3.2)

\[
    r(x \lor y) - r(x) \leq r(y) - r(x \land y) = 1
\]

which implies that in fact \( x \lor y > x \).

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

\[
    x \land y \leq x, y \implies x, y \leq x \lor y.
\] (2.3.3)

Denote by \( c(L) \) the maximum length\(^{1} \) 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 \( L = \{0, 1, x_1, \ldots, x_n\} \), where \( n \geq 1 \) and \( 0 \leq x_i < 1 \) for all \( i \). It is easy to see that these lattices are ranked and satisfy (2.3.2). Therefore, suppose \( c(L) = n \geq 3 \). Assume by induction that every semimodular lattice with all chains of length \( < c(L) \) are ranked and satisfy (2.3.2).

First, we show that \( L \) is ranked.

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

Let \( L' = [x_1, 1] \) and \( L'' = [y_1, 1] \). (See Figure 2.2.) 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 \( z_2 = x_1 \lor y_1 \). By (2.3.3), \( z_2 \) covers both \( x_1 \) and \( y_1 \). Let \( z_2, \ldots, 1 \) be a saturated chain in \( L \) (thus, in \( L' \cap L'' \)).

Since \( z \geq x_1 \), the chain \( z_1, \ldots, 1 \) has length \( n - 2 \). So the chain \( y_1, z_1, \ldots, 1 \) has length \( n - 1 \).

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

Second, we show that the rank function \( r \) of \( L \) satisfies (2.3.2). Let \( x, y \in L \) and take a saturated chain

\[
    x \land y = c_0 \leq c_1 \leq \cdots \leq c_{n-1} \leq c_n = x.
\]

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

\[
    y = c_0 \lor y \leq c_1 \lor y \leq \cdots \leq c_n \lor y = x \lor y.
\]

By Lemma 2.3.3, each \( \leq \) in this chain is either an equality or a covering relation. Therefore, the distinct elements \( c_i \lor y \) form a saturated chain from \( y \) to \( x \lor y \), whose length must be \( \leq n \). Hence

\[
    r(x \lor y) - r(y) \leq n = r(x) - r(x \land y)
\]

which implies the semimodular inequality (2.3.2). \( \square \)

\(^{1}\) 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 \( n + 1 \).
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 (2.3.2).

**Theorem 2.3.5.** \( L \) is modular if and only if it is ranked, with rank function \( r \) satisfying the modular equality

\[
  r(x \lor y) + r(x \land y) = r(x) + r(y) \quad \forall x, y \in L.
\]  

**Proof.** If \( L \) is modular, then it is both upper and lower semimodular, so the conclusion follows by Theorem 2.3.4. On the other hand, suppose that \( L \) is a lattice whose rank function \( r \) satisfies (2.3.4). Let \( x \leq z \in L \). We already know that \( x \lor (y \land z) \leq (x \lor y) \land z \), so it suffices to show that these two elements have the same rank. Indeed,

\[
  r(x \lor (y \land z)) = r(x) + r(y \land z) - r(x \land y \land z)
  = r(x) + r(y \land z) - r(x \land y)
  = r(x) + r(y) + r(z) - r(x \lor z) - r(x \land y)
\]

and

\[
  r((x \lor y) \land z) = r(x \lor y) + r(z) - r(x \lor y \lor z)
  = r(x \lor y) + r(z) - r(y \lor z)
  = r(x) + r(y) - r(x \land y) + r(z) - r(y \lor z). \quad \square
\]

### 2.4 Geometric lattices

We begin with a construction that gives the prototype of a geometric lattice. Let \( k \) be a field, let \( V \) be a vector space over \( k \), and let \( E \) be a finite subset of \( V \). Define

\[
  L(E) = \{ W \cap E \mid W \subseteq V \text{ is a vector subspace} \} \tag{2.4.1}
\]
Then $L(E)$ is a subposet of $\mathcal{B}_E$, naturally isomorphic to

$$\{kA \mid A \subseteq E\},$$

the family of vector subspaces of $V$ generated by subsets of $E$. (Of course, 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

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

The elements of $L(E)$ are called flats. For example, $E$ and $\emptyset$ are both flats, because $V \cap E = E$ and $O \cap E = \emptyset$, where $O$ means the zero subspace of $V$. 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 submodular, with rank function $r(A) = \dim kA$. (Exercise: Check that $r$ satisfies the submodular inequality.) It is not in general modular; e.g., see Example 2.4.3 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 $k\langle v_1, \ldots, v_k \rangle = k(v_1 + \cdots + kv_k)$. This motivates the following definition:

**Definition 2.4.1.** 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 $M_3$. (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 construction closely related to $L(E)$ is the lattice

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

(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 $L(E)$, where $E$ 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 2.4.2.** Let $E = \{a, b, c, d\}$, where $a, b, c$ 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).

![Diagram](attachment:affine_representation.png)

**Example 2.4.3.** If $E$ is the point configuration on the left with the only collinear triples $\{a, b, c\}$ and $\{a, d, e\}$, then $L^{\text{aff}}(E)$ is the lattice on the right.

27
This lattice is not modular: consider the two elements $bd$ and $ce$.

Recall that a lattice is relatively complemented if, whenever $y \in [x, z] \subseteq L$, there exists $u \in [x, z]$ such that $y \land u = x$ and $y \lor u = z$.

**Proposition 2.4.4.** Let $L$ be a finite semimodular lattice. Then $L$ is atomic (hence geometric) if and only if it is relatively complemented; that is, whenever $y \in [x, z] \subseteq L$, there exists $u \in [x, z]$ such that $y \land u = x$ and $y \lor u = z$.

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 $U + Y = Z$.

**Proof.** ($\implies$) Suppose that $L$ is atomic. Let $y \in [x, z]$, and choose $u \in [x, z]$ such that $y \land u = x$ (for instance, $u = x$). If $y \lor u = z$ then we are done. Otherwise, choose an atom $a \in L$ such that $a \leq z$ but $a \not\leq y \lor u$. Set $u' = u \lor a$. By semimodularity $u' > u$. Then $u' \lor y > u \lor y$ by Lemma 2.3.3, and $u' \land y = x$ (this takes a little more work; the proof is left as exercise). By repeatedly replacing $u$ with $u'$ if necessary, we eventually obtain a complement for $y$ in $[x, z]$.

($\impliedby$) Suppose that $L$ is relatively complemented and let $x \in L$. We want to write $x$ as the join of atoms. If $x = \hat{0}$ then it is the empty join; otherwise, let $a_1 \leq x$ be an atom and let $x_1$ be a complement for $a_1$ in $[\hat{0}, x]$. Then $x_1 < x$ and $x = a_1 \lor x_1$. Replace $x$ with $x_1$ and repeat, getting

$$x = a_1 \lor x_1 = a_1 \lor (a_2 \lor x_2) = (a_1 \lor a_2) \lor x_2 = \cdots = (a_1 \lor \cdots \lor a_n) \lor x_n = \cdots$$

Then $x > x_1 > x_2 > \cdots$, so eventually $x_n = \hat{0}$, and $x = a_1 \lor \cdots \lor a_n$. \hfill $\square$

What is tricky about the isomorphism in (2.4.1) 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 $\hat{A} := \k A \cap E$ — that is, the set of elements of $E$ in the linear span of $A$.

## 2.5 Exercises

**Exercise 2.1.** 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 $U(\mathcal{A}) = \{\cup \mathcal{A}' \mid \mathcal{A}' \subseteq \mathcal{A}\}$, considered as a poset ordered by inclusion.
Exercise 2.8. Fill in the details in the proof of the FTFDL (Theorem 2.1.7) by showing the following facts.

(a) Prove that $U(A)$ is a lattice. (Hint: Don’t try to specify the meet operation explicitly.)

(b) Construct a set family $A$ such that $U(A)$ is isomorphic to weak Bruhat order on $S_3$ (see Example 2.11).

(c) Construct a set family $A$ such that $U(A)$ is not ranked.

(d) Is every finite lattice of this form?

Exercise 2.2. 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 $\text{inv}(w)$. (For the definition of $\text{inv}(w)$, see Example 1.4.12.)

(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 subword $w = s_{i_{j_1}} \cdots s_{i_{j_l}}$. (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 $\prec$ is precisely Bruhat order on $\mathfrak{S}_n$.

Exercise 2.3. Prove that the rank-generating functions of weak order and Bruhat order on $\mathfrak{S}_n$ are both

$$
\sum_{w \in \mathfrak{S}_n} q^{r(w)} = \prod_{i=1}^{n} \frac{1 - q^i}{1 - q}
$$

where $r(w) = \# \{ \{i, j\} \mid i < j \text{ and } w_i > w_j \}$. (Hint: Induct on $n$, and use one-line notation for permutations, not cycle notation.)

**Distributive lattices**

Exercise 2.4. Prove that the two formulations of distributivity of a lattice $L$ are equivalent, i.e.,

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

Exercise 2.5. In Exercise 1.4 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 $P = \text{Irr}(L)$ is __________.”)

Exercise 2.6. Let $L$ be a lattice and $x \in L$. Prove that $x$ is join-irreducible if it covers exactly one other element.

Exercise 2.7. 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 $\lambda$?

(c) Let $\lambda$ 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 $\lambda$ is a hook shape (i.e., $\lambda = (n+1, 1, 1, \ldots, 1)$, with a total of $m$ copies of 1).

Exercise 2.8. Fill in the details in the proof of the FTFDL (Theorem 2.1.7) by showing the following facts.

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

$$\phi(x) = \{ p \mid p \in \text{Irr}(L), p \leq x \}$$

is indeed a lattice isomorphism.
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).

**Modular lattices**

**Exercise 2.9.** 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 2.2.3).

(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 (2.2.1).)

(b) Calculate the rank-generating function
\[
\sum_{V \in L_n(q)} x^{\dim V} = \sum_{k=0}^n x^k \# \{ V \in L_n(q) : \dim V = k \}.
\]

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 $L_n(q)$.

**Exercise 2.10.** Verify that the lattice $\Pi_4$ is not modular.

**Semimodular and geometric lattices**

**Exercise 2.11.** Prove that the partition lattice $\Pi_n$ is geometric by finding an explicit linear representation $E$. (Hint: What are the atoms of $\Pi_n$? Start by finding vectors corresponding to them. Also, there is a “best possible” construction that works over any field.)

**Exercise 2.12.** The purpose of this exercise is to show that the constructions $L$ and $L^{\text{aff}}$ produce the same class of lattices. Let $k$ be a field and let $E = \{ e_1, \ldots, e_n \} \subseteq k^d$.

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

(b) Let $v$ be a vector in $k^d$ that is not a scalar multiple of any $e_i$, let $H$ be a generic affine hyperplane, let $\tilde{e}_i$ be the projection of $e_i$ onto $H$, and let $\tilde{E} = \{ \tilde{e}_1, \ldots, \tilde{e}_n \}$. Prove that $L(E) = L^{\text{aff}}(\tilde{E})$.

(The first part is figuring out what “generic” means. A generic hyperplane might not exist for all fields, but if $k$ is infinite then almost all hyperplanes are generic.)

**Exercise 2.13.** Fill in the verification that $u' \wedge y = x$ in the first part of the proof of Proposition 2.4.4.
Chapter 3

Poset Algebra

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

3.1 The incidence algebra of a poset

Let \( P \) be a poset and let \( \text{Int}(P) \) denote the set of (nonempty) intervals of \( P \). Recall that an interval is a subset of \( P \) of the form \([x, y] := \{ z \in P \mid x \leq z \leq y \} \); if \( x \nleq y \) then \([x, y] = \emptyset\).

**Definition 3.1.1.** The incidence algebra \( I(P) \) is the set of functions \( \alpha : \text{Int}(P) \to \mathbb{C} \) ("incidence functions"), made into a \( \mathbb{C} \)-vector space with pointwise addition, subtraction and scalar multiplication, and equipped with the convolution product:

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

Here we abbreviate \( \alpha([x, y]) \) by \( \alpha(x, y) \), and it is often convenient to set \( \alpha(x, y) = 0 \) if \( x \nleq y \). Note that the assumption of local finiteness is both necessary and sufficient for convolution to be well-defined.

**Proposition 3.1.2.** Convolution is associative (although it is not in general commutative).
Proof. This is a straight-up calculation:

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

\[
= \sum_{z \in [x, y]} \left( \sum_{w \in [x, z]} \alpha(x, w)\beta(w, z) \right) \gamma(z, y)
\]

\[
= \sum_{w, z: x \leq w \leq z \leq y} \alpha(x, w)\beta(w, z)\gamma(z, y)
\]

\[
= \sum_{w \in [x, y]} \alpha(x, w) \left( \sum_{z \in [w, y]} \beta(w, z)\gamma(z, y) \right)
\]

\[
= \sum_{w \in [x, y]} \alpha(x, w) \cdot (\beta * \gamma)(w, y)
\]

\[
= [\alpha * (\beta * \gamma)](x, y).
\]

The multiplicative identity of \( I(P) \) is 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 \( \delta \).

Proposition 3.1.3. 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

\[
\alpha^{-1}(x, y) = \begin{cases} 
\alpha(x, x)^{-1} & \text{if } x = y, \\
-\alpha(y, y)^{-1} \sum_{z: x \leq z < y} \alpha^{-1}(x, z)\alpha(z, y) & \text{if } x < y.
\end{cases}
\] (3.1.1)

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 \( \beta \) be a left convolution inverse of \( \alpha \). In particular, \( \alpha(x, x) = \beta(x, x)^{-1} \) for all \( x \), 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 \( \beta(x, y) \) gives the formula (3.1.1) (pull the term \( \beta(x, y)\alpha(y, y) \) out of the sum), which is well-defined provided that \( f(y, y) \neq 0 \). So the nonzero condition is also sufficient. A similar argument shows that the nonzero condition is necessary and sufficient for \( \alpha \) 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.

The zeta function and eta function of \( P \) are defined as

\[
\zeta(x, y) = \begin{cases} 
1 & \text{if } x \leq y, \\
0 & \text{if } x \not\leq y,
\end{cases} \quad \eta(x, y) = \begin{cases} 
1 & \text{if } x < y, \\
0 & \text{if } x \not< y,
\end{cases}
\]

i.e., \( \eta = \zeta - 1 = \zeta - \delta \). Note that \( \zeta \) is invertible and \( \eta \) 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,

\[
\zeta^2(x, y) = \sum_{z \in [x, y]} \zeta(x, z)\zeta(z, y) = \sum_{z \in [x, y]} 1
\]

\[
= \# \{ z : x \leq z \leq y \},
\]

\[
\zeta^3(x, y) = \sum_{z \in [x, y]} \sum_{w \in [z, y]} \zeta(x, z)\zeta(z, w)\zeta(w, y) = \sum_{x \leq z \leq w \leq y} 1
\]

\[
= \# \{ (z, w) : x \leq z \leq w \leq y \},
\]

\[
\zeta^k(x, y) = \# \{ (x_1, \ldots, x_{k-1}) : x_1 \leq x_2 \leq \cdots \leq x_{k-1} \leq y \}.
\]

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 \( \zeta \) with \( \eta \), then the calculations all work the same way, except that all the \( \leq \)'s are replaced with \( < \)'s, so we get

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

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, 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 \( F : I(P) \otimes I(Q) \to I(P \times Q) \). 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 \( \text{Int}(P) \times \text{Int}(Q) \to \text{Int}(P \times Q) \) given by \((I, J) \to I \times J\), and \( F(\chi_I \otimes \chi_J) = \chi_{I \times J}\) (where \( \chi_I \) 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 3.1.4.** The map \( F \) just defined is a ring isomorphism. That is, for all \( \alpha, \beta \in I(P) \) and \( \phi, \psi \in I(Q) \),

\[
\alpha \phi \ast \beta \psi = (\alpha \ast \beta)(\phi \ast \psi).
\]

Furthermore, the incidence functions \( \delta \) and \( \zeta \) are multiplicative on direct products, i.e.,

\[
\delta_{P \times Q} = \delta_P \delta_Q \quad \text{and} \quad \zeta_{P \times Q} = \zeta_P \zeta_Q.
\]

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

\[
(\alpha \phi \ast \beta \psi)[(x, x'), (y, y')] = \sum_{(z, z') \in [(x, x'), (y, y')]} \alpha \phi[(x, x'), (z, z')] \cdot \beta \psi[(z, z'), (y, y')]
\]

\[
= \sum_{z \in [x, y]} \sum_{z' \in [x', y']} \alpha(x, z) \phi(x', z') \beta(z, y) \psi(z', y')
\]

\[
= \left[ \sum_{z \in [x, y]} \alpha(x, z) \beta(z, y) \right] \left[ \sum_{z' \in [x', y']} \phi(x', z') \psi(z', y') \right]
\]

\[
= (\alpha \ast \beta)(x, y) \cdot (\phi \ast \psi)(x', y').
\]

Multiplicativity of \( \delta \) and \( \zeta \) is immediate from their definitions. \(\qed\)

---

1 See §9.5 for an extremely brief introduction to the tensor product operation \( \otimes \).
3.2 The Möbius function

The Möbius function $\mu_P$ 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. Proposition 3.1.3 provides a recursive formula for $\mu$:

$$\mu(x, y) = \begin{cases} 0 & \text{if } y \not\geq x \text{ (i.e., if } [x, y] = \emptyset) \text{,} \\ 1 & \text{if } y = x \text{,} \\ -\sum_{x \leq z < y} \mu(x, z) & \text{if } x < y. \end{cases} \quad (3.2.1)$$

**Example 3.2.1.** 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 3.2.2.** Here are the Möbius functions $\mu_P(x) = \mu_P(\hat{0}, x)$ for the lattices $N_5$ and $M_5$:

And here are the Boolean lattice $B_3$ and the divisor lattice $D_{24}$:

**Example 3.2.3** (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 (3.2.1). The value of $\mu_{\Pi_4}(\hat{0}, \pi)$ depends only on the block sizes of $\pi$, in fact, $[\hat{0}, \pi] \cong \Pi_{x_1} \times \cdots \times \Pi_{x_k}$

<table>
<thead>
<tr>
<th>Block sizes</th>
<th>Number of $\pi$’s</th>
<th>Isomorphism type of $[\hat{0}, \pi]$</th>
<th>$\mu(\hat{0}, \pi)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,1,1,1</td>
<td>1</td>
<td>$\Pi_1$</td>
<td>1</td>
</tr>
<tr>
<td>2,1,1</td>
<td>6</td>
<td>$\Pi_2$</td>
<td>-1</td>
</tr>
<tr>
<td>2,2</td>
<td>3</td>
<td>$\Pi_2 \times \Pi_2$</td>
<td>1</td>
</tr>
<tr>
<td>3,1</td>
<td>4</td>
<td>$\Pi_3$</td>
<td>2</td>
</tr>
</tbody>
</table>
Therefore, $\mu(\Pi_4) = -(1 \cdot 1 - 1 \cdot 6 + 1 \cdot 3 + 2 \cdot 4) = -6$. Let’s try $n = 5$:

<table>
<thead>
<tr>
<th>Block sizes</th>
<th>Number of $\pi$’s</th>
<th>$\mu(0, \pi)$</th>
<th>Contribution to $-\mu(\Pi_5)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,1,1,1</td>
<td>1</td>
<td>$\mu(\Pi_1) = 1$</td>
<td>1</td>
</tr>
<tr>
<td>2,1,1,1</td>
<td>10</td>
<td>$\mu(\Pi_2) = -1$</td>
<td>-10</td>
</tr>
<tr>
<td>2,2,1</td>
<td>15</td>
<td>$\mu(\Pi_2 \times \Pi_2) = 1$</td>
<td>15</td>
</tr>
<tr>
<td>3,1,1</td>
<td>10</td>
<td>$\mu(\Pi_3) = 2$</td>
<td>20</td>
</tr>
<tr>
<td>3,2</td>
<td>10</td>
<td>$\mu(\Pi_3 \times \Pi_2) = -2$</td>
<td>-20</td>
</tr>
<tr>
<td>4,1</td>
<td>5</td>
<td>$\mu(\Pi_4) = -6$</td>
<td>-30</td>
</tr>
</tbody>
</table>

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 $\mu(P) = \mu(P(\hat{0}, \hat{1}))$ tells you a lot about a bounded poset $P$. Confusingly, this number itself is sometimes called the “Möbius function” of $P$ (a better term would be “Möbius number”). Here is the reason.

**Definition 3.2.4.** A family $\mathcal{F}$ of posets is **hereditary** if, for each $P \in \mathcal{F}$, every interval in $P$ is isomorphic to some [other] poset in $\mathcal{F}$. It is **semi-hereditary** if every interval in a member of $\mathcal{F}$ is isomorphic to a product of members of $\mathcal{F}$.

For example, the families of Boolean lattices, divisor lattices, and subspace lattices are all hereditary, and the family of partition lattices is semi-hereditary. 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 3.2.5.** The Möbius function is multiplicative on direct products, i.e., $\mu_P \times Q = \mu_P \mu_Q$ (in the notation of Proposition 3.1.4).

**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 $\mu_P \mu_Q = \zeta_{P \times Q}^{-1} = \mu_{P \times Q}$. (It is also possible to prove that $\mu_P \mu_Q = \mu_{P \times Q}$ directly from the definition.)

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 $\mu(\Pi_9) = \mu(\Pi_3)\mu(\Pi_2)^2$.

Since $\mu(\mathcal{B}_1) = -1$ and $\mathcal{B}_n$ is a product of $n$ copies of $\mathcal{B}_1$, an immediate consequence of Proposition 3.2.5 is the formula

$$\mu(\mathcal{B}_n) = (-1)^n.$$  

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

$$\mu(\mathcal{B}_n) = - \sum_{A \subseteq [n]} (-1)^{|A|} = - \sum_{k=0}^{n-1} (-1)^k \binom{n}{k} \quad \text{(by induction)}$$

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

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

**Example 3.2.6.** Let \( P \) be a product of \( k \) chains of lengths \( a_1, \ldots, a_n \). Equivalently,
\[
P = \{(x_1, \ldots, x_k) \mid 0 \leq x_i \leq a_i \text{ for all } i \in [k]\},
\]
ordered by \( x \preceq y \) iff \( x_i \leq y_i \) for all \( i \). Then Prop. 3.2.5 together with the formula for the Möbius function of a chain (above) gives
\[
\mu(\hat{0}, x) = \begin{cases} 0 & \text{if } x_i \geq 2 \text{ for at least one } i; \\
(-1)^s & \text{if } x \text{ consists of } s \, 1's \text{ and } n-s \, 0's. 
\end{cases}
\]
(The Boolean algebra is the special case that \( a_i = 2 \) for every \( i \).) This conforms to the definition of Möbius function that you may have seen in Math 724 or a class on 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.

**Theorem 3.2.7** (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
\[
c_k = c_k(P) = \left| \{(x_0, \ldots, x_k) : \hat{0} = x_0 < x_1 < \cdots < x_k = \hat{1}\} \right|.
\]
the number of chains of length \( k \) between \( \hat{0} \) and \( \hat{1} \). Then
\[
\mu_P(\hat{0}, \hat{1}) = \sum_k (-1)^k c_k.
\]

**Proof.** Recall that \( c_k = \eta^k(\hat{0}, \hat{1}) = (\zeta - \delta)^k(\hat{0}, \hat{1}) \). 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 \( \eta \) and \( 1 \) with \( \delta \), we get
\[
(\delta + \eta)^{-1} = \left( \sum_{k=0}^{\infty} (-1)^k \eta^k \right).
\]
Despite looking like an infinite power series, this is actually a valid polynomial equation in \( I(P) \), because \( \eta^k = 0 \) for \( k \) sufficiently large. Evaluating both sides on \( [\hat{0}, \hat{1}] \) gives
\[
\sum_{k=0}^{\infty} (-1)^k c_k = \sum_{k=0}^{\infty} (-1)^k \eta^k(\hat{0}, \hat{1}) = (\delta + \eta)^{-1}(\hat{0}, \hat{1}) = \zeta^{-1}(\hat{0}, \hat{1}) = \mu(\hat{0}, \hat{1}). \quad \square
\]

This alternating sum looks like an Euler characteristic. In fact it is.

**Corollary 3.2.8.** 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 1.1.10) whose vertices are the elements of \( P \setminus \{\hat{0}, \hat{1}\} \) and whose simplices are chains. Each chain \( x_0 = \hat{0} \prec x_1 \prec \cdots \prec x_k = \hat{1} \) 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
\[
\tilde{\chi}(\Delta(P)) \overset{\text{def}}{=} \sum_{k \geq -1} (-1)^k f_k(\Delta(P)) = \sum_{k \geq 1} (-1)^{k-2} c_k(P) \mu_P(\hat{0}, \hat{1}). \quad \square
\]

**Example 3.2.9.** Let \( P \) be the poset of Example 3.2.2. Then \( c_0 = 0, c_1 = 1, c_2 = 6, c_3 = 6, \) and \( c_k = 0 \) for \( k > 3 \). So \( c_0 - c_1 + c_2 - c_3 = -1 = \mu_P(\hat{0}, \hat{1}) \).

36
3.3 Möbius inversion

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

**Theorem 3.3.1 (Möbius inversion formula).** Let $P$ be a poset in which every principal order ideal is finite, and let $f, g : P \to \mathbb{C}$. Then

$$g(x) = \sum_{y : y \leq x} f(y) \quad \forall x \in P \iff f(x) = \sum_{y : y \leq x} \mu(y, x)g(y) \quad \forall x \in P, \quad (3.3.1a)$$

$$g(x) = \sum_{y : y \geq x} f(y) \quad \forall x \in P \iff f(x) = \sum_{y : y \geq x} \mu(x, y)g(y) \quad \forall x \in P. \quad (3.3.1b)$$

*Proof.* Stanley calls the proof “A trivial observation in linear algebra”. We regard the incidence algebra as acting $\mathbb{C}$-linearly on the vector space $V$ of functions $f : P \to \mathbb{Z}$ by$^2$

$$(f \cdot \alpha)(x) = \sum_{y : y \leq x} \alpha(y, x)f(y),$$

$$(\alpha \cdot f)(x) = \sum_{y : y \geq x} \alpha(x, y)f(y).$$

for $\alpha \in I(P)$. In terms of these actions, formulas (3.3.1a) and (3.3.1b) are respectively just the “trivial” observations

$$g = f \cdot \zeta \iff f = g \cdot \mu, \quad (3.3.2a)$$

$$g = \zeta \cdot f \iff f = \mu \cdot g. \quad (3.3.2b)$$

We just have to prove that these actions are indeed actions, i.e.,

$$(\alpha \ast \beta) \cdot f = \alpha \cdot (\beta \cdot f) \quad \text{and} \quad f \cdot (\alpha \ast \beta) = (f \cdot \alpha) \cdot \beta.$$

We prove the second identity:

$$(f \cdot (\alpha \ast \beta))(y) = \sum_{x : x \leq y} (\alpha \ast \beta)(x, y)f(x)$$

$$= \sum_{x : x \leq y} \left( \sum_{z : z \in [x, y]} \alpha(x, z)\beta(z, y) \right)f(x)$$

$$= \sum_{z : z \leq y} \left( \sum_{x : x \leq z} \alpha(x, z)f(x) \right)\beta(z, y)$$

$$= \sum_{z : z \leq y} (f \cdot \alpha)(z)\beta(z, y) = ((f \cdot \alpha) \cdot \beta)(y).$$

and the other verification is a mirror image of this one. \hfill \Box

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

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

$^2$The action is denoted by $\cdot$ to avoid confusion with other multiplications, such as scalar multiplication and convolution.
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 algebra 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.

**Example 3.3.2.** Here’s an oldie-but-goodie. A derangement is a permutation \( \sigma \in S_n \) with no fixed points. If \( D_n \) is the set of derangements in \( S_n \), then

\[
\begin{align*}
|D_1| &= 0, \\
|D_2| &= 1 = |\{21\}|, \\
|D_3| &= 2 = |\{231, 312\}|, \\
|D_4| &= 9 = |\{2341, 2314, 2413, 3142, 3412, 3421, 4123, 4312, 4321\}|, \\
&\ldots
\end{align*}
\]

The problem is to determine \( |D_n| \) in general. For \( S \subseteq [n] \), let

\[
\begin{align*}
f(S) &= \{\sigma \in S_n \mid \sigma(i) = i \text{ iff } i \in S\}, \\
g(S) &= \{\sigma \in S_n \mid \sigma(i) = i \text{ if } i \in S\}.
\end{align*}
\]

Thus \( D_n = f(\emptyset) \).

It’s easy to count \( g(S) \) 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’s hard to count \( f(S) \) 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 (3.3.1b),

\[
f(S) = \sum_{R \supseteq S} \mu(S, R) g(R) = \sum_{R \supseteq S} (-1)^{|R| - |S|} (n - |R|)! = \sum_{r=0}^{n} \binom{n-s}{r-s} (-1)^{r-s} (n-r)!.
\]

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 3.3.3.** 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 \( P = \{p_1, \ldots, p_s\} \). We work in the lattice \( D_n \cong C_{a_1} \times \cdots \times C_{a_s} \), and 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

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

38
Applying formulation (3.3.1b) of Möbius inversion gives

\[ f(x) = \sum_{y \leq x} \mu(x, y)g(y). \]

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

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

Therefore,

\[ \phi(n) = f(1) = \sum_{y \in D_n} \mu(1, y)(n/y) \]

\[ = n \sum_{Q \subseteq P} \left( -1 \right)^{|Q|} \prod_{p_i \in Q} p_i \]

\[ = \frac{n}{p_1 \cdots p_r} \sum_{Q \subseteq P} (-1)^{|Q|} p_1 \cdots p_r \prod_{p_i \in Q} p_i \]

\[ = \frac{n}{p_1 \cdots p_r} \sum_{R = P \setminus Q \subseteq P} (-1)^{|s|} \prod_{p_i \in R} p_i \]

\[ = \frac{n}{p_1 \cdots p_r} (-1)^s \prod_{i=1}^s (1 - p_i) \]

\[ = p_1^{\alpha_1-1} \cdots p_r^{\alpha_r-1}(p_1 - 1) \cdots (p_r - 1) \]

as is well known.

**Example 3.3.4.** 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 \( \kappa \) 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 1.4.3 and Exercise 1.5). For each \( t \)-coloring \( \kappa \), let \( G_\kappa \) 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 \( G_\kappa \); then \( P(\kappa) \) is an element of \( K(G) \). The coloring \( \kappa \) is proper if and only if \( P(\kappa) = \emptyset_{K(G)} \), the partition of \( V(G) \) into singleton blocks. Accordingly, if we define \( f : K(G) \to \mathbb{N} \) by

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

then the number of proper \( t \)-colorings is \( f(\emptyset) \). 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 \( \pi \) 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 (3.3.1b)) says that

\[ p_G(t) = f(\emptyset) = \sum_{\pi \in K(G)} \mu(\emptyset, \pi)g(\pi) = \sum_{\pi \in K(G)} \mu(\emptyset, \pi)t^{||\pi||}. \]
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**.

If $G = K_n$ is the complete graph, then the connectivity lattice $K(K_n)$ is just the full partition lattice $\Pi_n$. On the other hand, we can calculate the chromatic polynomial of $K_n$ directly: it is $p_G(t) = t(t-1)(t-2) \cdots (t-n+1)$ (since a proper coloring must assign different colors to all vertices). Equating our two expressions for $p_G(t)$ gives

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

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

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

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

**Example 3.3.5.** Here is another way to use Möbius inversion to compute the Möbius function itself. In this example, we will do this for the lattice $L_n(q)$. The partition lattice can be handled the same way (Exercise 3.3).

For small $n$, it is possible to work out the Möbius function of $L_n(q)$ by hand. For instance, $\mu(L_1(q)) = \mu(\mathcal{P}_1) = -1$, 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 $\mu(L_4(q)) = q^n$. 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

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

since $\tilde{\phi}$ 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 $0 = \hat{0}_L$, we obtain

$$
f(\hat{0}) = \sum_{U \in L} \mu_L(\hat{0}, U) t^{n-\dim U} \\
= \# \{ 1\text{-1 linear maps } V \to X \} \\
= (t-1)(t-q)(t-q^2) \cdots (t-q^{n-1}). \tag{3.3.3}
$$
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 (3.3.3) 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 become

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

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

### 3.4 The characteristic polynomial

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

\[
\chi(P; t) = \sum_{x \in P} \mu(\hat{0}, x) t^{n-r(x)}
\]

where \( r \) is the rank function. This is called the characteristic polynomial of \( P \). Thus we have shown that

\[
\chi(\Pi_n; t) = t(t-1)(t-n+1), \\
\chi(K(G); t) = p_G(t), \\
\chi(L_n(q); t) = (t-1)(t-q)(t-q^2) \cdots (t-q^{n-1}).
\]

Moreover, the characteristic polynomial of the Boolean lattice \( B_n \) is

\[
\chi(B_n; t) = \sum_{j=0}^{n} (-1)^j \binom{n}{j} t^{n-j} = (t-1)^n.
\]

In fact, since the Möbius function is multiplicative on direct products of posets (Proposition 3.1.4), so is the characteristic polynomial.

Note that \( \chi(P,0) = \mu(P) \), so the characteristic polynomial generalizes the Möbius number of a poset. It contains additional information about \( P \) 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 Section 6, is that \( |\chi_P(-1)| \) and \( |\chi_P(1)| \) count the number of regions and bounded regions of \( X \), respectively.

### 3.5 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). The main algebraic object is the following ring.

**Definition 3.5.1.** Let \( L \) be a lattice. The Möbius algebra \( A(L) \) 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, \( \mathbf{1} \) is the multiplicative unit of \( A(L) \).)
The elements of \( L \) form a vector space basis of \( A(L) \) 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 \( A(L) \cong \mathbb{C}[x_1, \ldots, x_n]/(x_1^2 - x_1, \ldots, x_n^2 - x_n) \), with a natural vector space basis given by square-free monomials.

It seems as though \( A(L) \) could have a complicated ring structure, but actually it is quite simple.

**Proposition 3.5.2.** Let \( L \) be a finite lattice. For \( x \in L \), define

\[
\varepsilon_x = \sum_{y \leq x} \mu(y, x) y \in A(L).
\]

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

**Proof.** By Möbius inversion,

\[
x = \sum_{y \leq x} \varepsilon_y
\]

so \( B \) is a vector space basis for \( A(L) \) 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 \( \phi : A(L) \to \mathbb{C}^{|L|} \) given by

\[
\phi(\varepsilon_x) = 1_x
\]

is a ring isomorphism. It is certainly a vector space isomorphism, and (3.5.1) implies that

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

This proof almost looks like an illusion, since we never calculated \( \varepsilon_x \varepsilon_y \) explicitly.

**Remark 3.5.3.** Darij Grinberg observes that Prop. 3.5.2 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 \( A(L) \) that it must have a unit, but the existence of a unit is implied by the isomorphism \( A(L) \cong \mathbb{C}^{|L|} \) (in fact the unit is \( \sum_{x \in L} \varepsilon_x \)).

What is the Möbius algebra good for? Well, we will establish identities in \( A(L) \) that permit computation \( \mu(x, y) \) by summing over a cleverly chosen subset of \( [x, y] \), rather than the entire interval. Of course we know that \( \mu(P) = -\sum_{x \neq 1} \mu(0, x) \) for any poset \( P \), but this leads to a recursive computation that can be quite inefficient. The special structure of a lattice \( L \) leads to much more streamlined expressions for \( \mu(L) \).

**Proposition 3.5.4** (Weisner’s theorem). Let \( L \) be a finite lattice with at least two elements. Then for every \( a \in L \setminus \{ 1 \} \) we have the equation

\[
\sum_{x \in L : x \wedge a = 0} \mu(x, 1) = 0.
\]

In particular, pulling off the \( x = 0 \) summand gives

\[
\mu(L) = \mu_L(0, 1) = -\sum_{x \in L \setminus \{ 0 \} : x \wedge a = 0} \mu(x, 1).
\]
Proof. We work in \( A(L) \) and calculate \( a \varepsilon_1 \) in two ways. On the one hand

\[
a \varepsilon_1 = \left( \sum_{b \leq a} \varepsilon_b \right) \varepsilon_1 = 0.
\]

On the other hand

\[
a \varepsilon_1 = a \sum_{x \in L} \mu(x, \hat{1}) x = \sum_{x \in L} \mu(x, \hat{1}) x \wedge a.
\]

Now taking the coefficient of \( \hat{0} \) on both sides gives the desired equation. \( \square \)

Example 3.5.5 (The Möbius function of the partition lattice \( \Pi_n \)). Let \( a = 1|2\cdot3\cdot\ldots\cdot n \in \Pi_n \). Then the partitions \( x \) that show up in the sum of (3.5.2) are just the atoms whose non-singleton block is \( \{1, i\} \) for some \( i > 1 \). For each such \( x \), the interval \([x, \hat{1}] \subseteq \Pi_n \) is isomorphic to \( \Pi_{n-1} \), so (3.5.2) 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 3.5.6 (The Möbius function of the subspace lattice \( L_n(q) \)). Let \( L = L_n(q) \), and let \( A = \{(v_1, \ldots, v_n) \in \mathbb{F}_q^n \mid v_n = 0\} \). This is a codimension-1 subspace in \( \mathbb{F}_q^n \), hence 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 \((x_1, \ldots, x_n)\) with \( x_n \neq 0 \). We may as well assume \( x_n = 1 \) and choose \( x_1, \ldots, x_{n-1} \) arbitrarily, so there are \( q^{n-1} \) such lines. Moreover, the interval \([X, \hat{1}] \subseteq L \) is isomorphic to \( L_{n-1}(q) \). Therefore

\[
\mu(L_n(q)) = -q^{n-1} \mu(L_{n-1}(q))
\]

and by induction

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

Here is an important consequence of Weisner’s theorem.

Theorem 3.5.7. The Möbius function of any semimodular lattice \( L \) weakly alternates in sign. That is, \((-1)^{r(x)} \mu(\hat{0}, x) \geq 0\) for all \( x \in L \).

Proof. First, note that \( \mu(P) = \mu(P^*) \) for any poset \( P \) (the proof is an exercise). Therefore, the dual form of Weisner’s theorem says that for any \( a \in L \setminus \{ \hat{0} \} \) we have the equation

\[
\sum_{x \in L : x \wedge a = \hat{1}} \mu(\hat{0}, x) = 0. \quad (3.5.3)
\]

Now, suppose \( L \) is semimodular of rank \( n \). The theorem is certainly true if \( n \leq 1 \), so we proceed by induction. Suppose we take \( a \) to be an atom. If \( x \vee a = \hat{1} \), then

\[
r(x) \geq r(x \vee a) + r(x \wedge a) - r(a) \geq n - 1,
\]

so \( x \) is either \( \hat{1} \) or a coatom whose meet with \( a \) is \( \hat{0} \). Therefore, (3.5.3) can be rewritten as

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

But each interval \([\hat{0}, x] \) is itself a submodular lattice of rank \( n - 1 \), so by induction each summand has sign \((-1)^{n-1} \), which completes the proof. \( \square \)
A drawback of Weisner’s theorem is that it is still recursive; the right-hand side of (3.5.2) 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 \( [0, x] \cong L_k \) (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 3.5.8.** Let \( L \) be a lattice. An upper crosscut of \( L \) is a set \( X \subseteq L \setminus \{ \hat{1} \} \) such that if \( y \in L \setminus X \setminus \{ \hat{1} \} \), then \( y < x \) for some \( x \in X \). A lower crosscut of \( L \) is a set \( X \subseteq L \setminus \{ \hat{0} \} \) such that if \( y \in L \setminus X \setminus \{ \hat{0} \} \), 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 3.5.9** (Rota’s crosscut theorem). Let \( L \) be a finite lattice and let \( X \) be an upper crosscut. Then

\[
\mu(L) = \sum_{Y \subseteq X: \bigwedge Y = \hat{0}} (-1)^{|Y|}.
\]

(3.5.4a)

Dually, if \( X \) is a lower crosscut, then

\[
\mu(L) = \sum_{Y \subseteq X: \bigvee Y = \hat{1}} (-1)^{|Y|}.
\]

(3.5.4b)

**Proof.** We prove only (3.5.4a); the proof of (3.5.4b) is dual. For any \( x \in L \), we have the following simple equation in the Möbius algebra of \( L \):

\[
\hat{1} - x = \sum_{y \in L} \varepsilon_y - \sum_{y \leq x} \varepsilon_y = \sum_{y \nleq x} \varepsilon_y.
\]

Therefore, for any \( X \subseteq L \),

\[
\prod_{x \in X} (\hat{1} - x) = \prod_{x \in X} \sum_{y \nleq x} \varepsilon_y = \sum_{y \in Y} \varepsilon_y
\]

where \( Y = \{ y \in L \mid y \nleq x \text{ for all } x \in X \} \). (Expand the sum and recall that \( \varepsilon_y \varepsilon_y' = \delta_{yy'} \varepsilon_y \)) But if \( X \) is an upper crosscut, then \( Y = \{ \hat{1} \} \), and this last equation becomes

\[
\prod_{x \in X} (\hat{1} - x) = \varepsilon_1 = \sum_{y \in L} \mu(y, \hat{1}) y.
\]

(3.5.5)

On the other hand, a direct binomial expansion gives

\[
\prod_{x \in X} (\hat{1} - x) = \sum_{A \subseteq X} (-1)^{|A|} \bigwedge A.
\]

(3.5.6)

Now equating the coefficients of \( \hat{0} \) on the right-hand sides of (3.5.5) and (3.5.6) yields (3.5.4a).

**Corollary 3.5.10.** Let \( L \) be a lattice in which \( \hat{1} \) is not a join of atoms (for example, a distributive lattice that is not Boolean). 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 3.5.11. Let $L$ be a geometric lattice of rank $r$, and let $P = L \setminus \{0,1\}$. 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 $r - 2$.

3.6 Exercises

Exercise 3.1. Let $P$ be a locally finite poset. Consider the incidence function $\kappa \in I(P)$ defined by

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

(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 $\kappa$ and its convolution powers whether $P$ is ranked?
(c) Give combinatorial interpretations of $\kappa \ast \zeta(x, y)$ and $\zeta \ast \kappa(x, y)$.

Exercise 3.2. 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 3.1.4) directly from the definition of $\mu$.

Exercise 3.3. Let $\Pi_n$ be the lattice of set partitions of $[n]$. Recall that the order relation on $\Pi_n$ is given as follows: if $\pi, \sigma \in \Pi_n$, then $\pi \leq \sigma$ if every block of $\pi$ is contained in some block of $\sigma$ (for short, “$\pi$ refines $\sigma$”). In this problem, you will calculate the number $\mu_n := \mu_{\Pi_n}(0, 1)$. You may want to first review Example 3.3.5.

(a) Calculate $\mu_n$ by brute force for $n = 1, 2, 3, 4$. Make a conjecture about the value of $\mu_n$ in general.
(b) Let $T$ be a finite set of cardinality $t$. Define functions $f : \Pi_n \to \mathbb{Q}[x]$ by

$$
f(\pi) = \# \{ h : [n] \to T \mid h(s) = h(s') \text{ iff } s \sim_\pi s' \},
$$

$$
g(\pi) = \# \{ h : [n] \to T \mid h(s) = h(s') \text{ if } s \sim_\pi s' \},
$$

where $\sim_\pi$ denotes the equivalence relation induced by $\pi$. Find explicit formulas for $f(\pi)$ and $g(\pi)$.
(c) Using Möbius inversion, find the characteristic polynomial $\chi(\Pi_n, t)$ and the Möbius function $\mu(\Pi_n)$.

Exercise 3.4. 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 $F(\alpha \ast \beta) = F(\beta) \ast F(\alpha)$.
(b) Show that $F(\delta_P) = \delta_{P^*}$ and $F(\zeta_P) = \zeta_{P^*}$. Conclude that $\mu(P) = \mu(P^*)$.

Exercise 3.5. A set partition in $\Pi_n$ 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 $\text{NC}_n$. Ordering by reverse refinement makes $\text{NC}_n$ into a subposet of the partition lattice $\Pi_n$. Note that $\text{NC}_n = \Pi_n$ 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 $\xi = 1|2 5|3|4|6 8 12|7|9|10 11 \in \text{NC}_{12}$ is shown in Figure 3.1(a).

(a) Prove that $\text{NC}_n$ is a ranked lattice. Is it a sublattice of $\Pi_n$?
(b) Prove that the numbers $nc_n = |NC_n|$ satisfy the Catalan recurrence is the $n^{th}$ Catalan number $C_n = \frac{1}{n+1} \binom{2n}{n}$. (Find a bijection to something, or possibly establish the recurrence.)

(c) Prove that the operation of Kremeras complementation is an anti-automorphism of $\Pi_n$. To define the Kremeras complement $K(\pi)$ of $\pi \in NC_n$, start with the chord diagram of $\pi \in NC_n$ 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 $\pi$. For instance, the Kremeras complement of the noncrossing partition $\xi \in NC_{12}$ shown above is $K(\xi) = 1 5 12 | 2 3 4 | 6 7 | 8 9 11 | 10$ (see Figure 3.1(b)).

(d) Use Weisner’s theorem to prove that $\mu(\Pi_n) = (-1)^{n-1}C_{n-1}$ 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).

**Exercise 3.6.** This problem is about how far Proposition 3.5.2 can be extended. Suppose that $R$ is a commutative $C$-algebra of finite dimension $n$ as a $C$-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 $R \cong C^n$ as rings.

**Exercise 3.7.** The $q$-binomial coefficient is the rational function

$$\left[\begin{array}{c} n \\ k \end{array}\right]_q = \frac{(q^n - 1)(q^n - q) \cdots (q^n - q^{k-1})}{(q^k - 1)(q^k - q) \cdots (q^k - q^{k-1})}.$$  

(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:

$$\left[\begin{array}{c} n \\ k \end{array}\right]_q = q^k \left[\begin{array}{c} n - 1 \\ k \end{array}\right]_q + \left[\begin{array}{c} n - 1 \\ k - 1 \end{array}\right]_q$$  

and

$$\left[\begin{array}{c} n \\ k \end{array}\right]_q = \left[\begin{array}{c} n - 1 \\ k \end{array}\right]_q + q^{n-k} \left[\begin{array}{c} n - 1 \\ k - 1 \end{array}\right]_q.$$  

Deduce that $\left[\begin{array}{c} n \\ k \end{array}\right]_q$ is actually a polynomial in $q$ (not merely a rational function).

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

$$\prod_{k=0}^{n-1} (x - q^k) = \sum_{k=0}^{n} \left[\begin{array}{c} n \\ k \end{array}\right]_q (-1)^k q^{\binom{k}{2}} x^{n-k}.$$  

(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.)
Exercise 3.8. (Stanley, EC1, 3.129) Here is a cute application of combinatorics to elementary number theory. Let $P$ be a finite poset, and let $\hat{P} = P \cup \{\hat{0}, \hat{1}\}$. 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 $\mu_{\hat{P}}(\hat{0}, \hat{1}) \equiv -1 \pmod{p}$. What does this say in the case that $\hat{P} = \Pi_p$?
Chapter 4

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].

4.1 Closure operators

In what follows, there will be a lot of singleton sets. For brevity, we frequently abbreviate $\{e\}$ by $e$ in notation such as $A \cup e$, $A \setminus e$, and $A \vee e$. Would it be better to abbreviate $A + e = A \cup \{e\}$ and $A - e = A \setminus \{e\}$ to avoid overloading the symbols $\cup$ and $\setminus$?

Definition 4.1.1. 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 $\bar{A} = A$. A matroid closure operator is a closure operator that satisfies in addition

$$\text{if } e \not\in \bar{A} \text{ but } e \in \bar{A} \cup e', \text{ then } e' \in \bar{A} \cup e.$$  \hspace{1cm} (4.1.1)

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.

For any closure operator (not necessarily matroidal), any two subsets $A, B \subseteq E$ satisfy $\bar{A \cap B} \subseteq \bar{A}$ and

48
\( A \cap B \subseteq B \) (by (iii)), hence \( A \cap B \subseteq A \cap B \). In particular, if \( F \) and \( G \) are flats, then
\[
F \cap G \subseteq F \cap G = F \cap G \quad (4.1.2)
\]
so equality holds. That is, the intersection of flats is a flat.

**Example 4.1.2. Vector matroids.** Let \( V \) be a vector space over a field \( k \), and let \( E \subseteq V \) be a finite set. Then
\[
A \mapsto \bar{A} := kA \cap E
\]
is a matroid closure operator on \( E \). It is easy to check the conditions for a closure operator. To check condition (4.1.1), if \( e \in A \cup e' \), 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 \( k \). 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 \( f' \) as a linear combination of the vectors in \( A \cup e \), obtaining (4.1.1). 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).

### 4.2 Matroids and geometric lattices

The following theorem says that simple matroids and geometric lattices are essentially the same things.

**Theorem 4.2.1.** 1. Let \( M \) be a simple matroid with finite ground set \( E \). Let \( L(M) \) be the poset of flats of \( M \), ordered by inclusion. Then \( L(M) \) is a geometric lattice, under the operations \( F \wedge G = F \cap G, F \vee G = F \cup G \).

2. Let \( L \) be a geometric lattice and let \( E \) be its set of atoms. Then the function \( A \mapsto \bar{A} = \{ e \in E \mid e \leq \bigvee A \} \) is a matroid closure operator for a simple matroid on \( E \).

3. These constructions are mutual inverses.

**Proof.** (1) Let \( M \) be a simple matroid on \( E \).

First, we show that \( L(M) \) is an atomic lattice. The intersection of flats is a flat by (4.1.2), so the operation \( F \wedge G = F \cap G \) makes \( L(M) \) into a meet-semilattice. It’s bounded (with \( \emptyset = \emptyset \) and \( \mathbf{1} = E \)), so it’s a lattice by Proposition 1.4.8. By the way, the join operation is \( F \vee G = F \cup G \), which is by definition the smallest flat containing \( F \cup G \), so it is the meet of all flats containing both \( F \) and \( G \). (Note that this argument shows that any closure operator, not necessarily matroidal, gives rise to a lattice.)

By definition of a simple matroid, the singleton subsets of \( E \) are atoms in \( L(M) \). Every flat is the join of the atoms corresponding to its elements, so \( L(M) \) is atomic.

At this point we prove a useful lemma about covering relations in \( L(M) \)

¹ Usually one of the first two terms is used for a matroid defined by a set of vectors; “representable” suggests that the matroid *could* be represented in that way.
Lemma 4.2.2. If \( F \in L(M) \) and \( e \in E \setminus F \) (so that \( F < F \lor e \)), then in fact \( F \preceq F \lor e \).

**Proof.** Suppose that there is a flat \( G \) such that
\[
F \subseteq G \subseteq F \lor e = \overline{F \cup e}. \tag{4.2.1}
\]
Let \( e' \in G \setminus F \). Then \( e' \in \overline{F \cup e} \), so the exchange axiom (4.1.1) implies \( e \in \overline{F \cup e'} \), which in turn implies that \( F \lor e \subseteq F \lor e' \subseteq G \). Hence the \( \subseteq \) in (4.2.1) is actually an equality. We have shown that there are no flats strictly between \( F \) and \( F \lor e \), proving the claim. 

Of course, if \( F \preceq G \) then \( G = F \lor e \) for any \( e \in G \setminus F \). So we have essentially characterized all the covering relations in \( L(M) \), which is very useful.

Suppose now that \( F \) and \( G \) are incomparable and that \( G \succeq F \land G \). Then \( G \) is of the form \((F \land G) \lor e\), and we can take \( e \) to be any element of \( G \setminus F \). In particular \( F < F \lor e \), so by Lemma 4.2.2, \( F \preceq F \lor e \). Moreover,
\[
F \lor G = F \lor ((F \land G) \lor e) = (F \lor e) \lor (F \land G) = F \lor e.
\]

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

\[
\begin{array}{ccc}
F & \lor & G = G \lor e \\
& \lor & \\
F \land G & \lor & G = (F \land G) \lor e
\end{array}
\]

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 \( \bar{A} = \{ e \in E \mid e \leq \bigvee A \} \) 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 (4.1.1).

Recall that if \( L \) is semimodular and \( x, e \in L \) with \( e \) an atom and \( x \npreceq e \) (so that \( x < x \lor e \)), then in fact \( x \preceq x \lor e \), because
\[
r(x \lor e) - r(x) \leq r(e) - r(x \land e) = 1 - 0 = 1.
\]
Accordingly, let \( A \subseteq E \) and let \( e, f \in E \setminus A \). Suppose that \( e \in \overline{A \cup f} \); we must show that \( f \in \overline{A \cup e} \). Let \( x = \bigvee A \in L \). Then
\[
x \preceq x \lor f \quad \text{(by the previous remark)} \quad \text{and} \quad x \preceq x \lor e \preceq x \lor f,
\]
which together imply that \( x \lor f = x \lor e \). In particular \( f \preceq x \lor e \), i.e., \( f \in \overline{A \cup e} \), proving that we have a matroid closure operator.

Part (3) is left as an exercise. \( \square \)
In view of Theorem 4.2.1, 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 4.2.3.** A **matroid rank function** on $E$ is a function $r : 2^E \rightarrow \mathbb{N}$ satisfying the following conditions for all $A, B \subseteq E$:

1. **(R1)** $r(A) \leq |A|$.
2. **(R2)** If $A \subseteq B$ then $r(A) \leq r(B)$.
3. **(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

$$\bar{A} = \{ e \in E : r(A \cup e) = r(A) \}.$$  

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 4.2.4.** 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 4.2.3. The corresponding matroid is called the **uniform matroid** $U_k(n)$. Its closure operator is

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

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 algebra $2^{[n]}$ that has been truncated at the $k^{th}$ rank: that is, all elements of rank $\geq k$ have been deleted and replaced with a single $1$. For $n = 3$ and $k = 2$, this lattice is $M_5$. For $n = 4$ and $k = 3$, the Hasse diagram is as shown below.

![Hasse diagram](image)

If $E$ is a set of $n$ vectors in general position in $k^k$, 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 $k$ 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, $k$ must be sufficiently large (in terms of $n$) in order for $k^k$ to have $n$ points in general position: for instance, $U_2(4)$ cannot be represented as a matroid over $\mathbb{F}_2$ simply because $\mathbb{F}_2^3$ contains only three nonzero vectors.
Definition 4.2.5. 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 \( f : E \to E' \) meeting any (hence all) of the following:

1. There is a lattice isomorphism \( L(M) \cong L(M') \);
2. \( r(A) = r(f(A)) \) for all \( A \subseteq E \). (Here \( f(A) = \{ f(a) \mid a \in A \} \).)
3. \( \overline{f(A)} = f(\overline{A}) \) for all \( A \subseteq E \).

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

### 4.3 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 4.3.1.** 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

\[
\overline{A} = \{ e = xy \in E \mid x, y \text{ belong to the same component of } G|_A \}. \tag{4.3.1}
\]

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

![Graph Example](image)

**Proposition 4.3.2.** The operator \( A \mapsto \overline{A} \) defined by (4.3.1) is a matroid closure operator.

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

Finally, suppose \( e = xy \not\in A \) but \( e \in \overline{A} \). Let \( P \) be an \((A \cup f)\)-path from \( x \) to \( y \). Then \( f \in P \) (because there is no \( A \)-path from \( x \) to \( y \)) and \( P \cup e \) is a cycle. Deleting \( f \) produces an \((A \cup e)\)-path between the endpoints of \( f \). \( \square \)
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 of \( A \) (or of \( G|_A \)). They are the bases of the graphic matroid \( M(G) \).

Theorem 4.3.3. 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 1.4.3.

Example 4.3.4. 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 4.3.5. 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 4.3.6. 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 \( \Pi_n \).

### 4.4 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.

Definition 4.4.1. A (matroid) independence system \( \mathcal{I} \) is a family of subsets of \( E \) such that

(I1) \( \emptyset \in \mathcal{I} \);

(I2) if \( I \in \mathcal{I} \) and \( I' \subseteq I \), then \( I' \in \mathcal{I} \);

(I3) ("Donation") if \( I, J \in \mathcal{I} \) and \( |I| < |J| \), then there is some \( x \in J \setminus I \) such that \( I \cup x \in \mathcal{I} \).

Note that conditions (I1) and (I2) say that \( \mathcal{I} \) is an abstract simplicial complex on \( E \) (see Example 1.1.10).
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 4.4.2.** Let \( \mathcal{I} \) be a matroid independence system on \( E \). Suppose that \( I \in \mathcal{I} \) 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. \( \square \)

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 \( \mathcal{I} \) is pure — an induced subcomplex is something of the form \( \mathcal{I}_X = \{ I \in \mathcal{I} \mid I \subseteq X \} \), for \( X \subseteq E \), and “pure” means that all maximal faces have the same cardinality.) This actually characterizes independence complexes among all simplicial complexes; we will state a stronger result (Theorem 4.4.4) soon.

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

**Proposition 4.4.3.** Let \( E \) be a finite set.

1. If \( r \) is a matroid rank function on \( E \), then
\[
\mathcal{I} = \{ A \subseteq E : r(A) = |A| \}
\]
   is an independence system.

2. If \( \mathcal{I} \) is an independence system on \( E \), then
\[
r(A) = \max\{|I| : I \subseteq A, I \in \mathcal{I}\}
\]
   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 \( \mathcal{I} \) as in (4.4.1a). First, \( r(I) \leq |I| \) for all \( I \subseteq E \), so (I1) follows immediately. Second, suppose \( I \in \mathcal{I} \) 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)
\[
\emptyset \subseteq \{x_1\} \subseteq \{x_1, x_2\} \subseteq \cdots \subseteq I' \subseteq \cdots \subseteq I.
\]
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 \( I' \in \mathcal{I} \), establishing (I2).

To show (I3), let \( I, J \in \mathcal{I} \) with \( |I| < |J| \) and let \( J \setminus I = \{x_1, \ldots, x_n\} \). If \( n = 1 \) then \( J = I \cup \{x_1\} \) and there is nothing to show. Now suppose that \( n \geq 1 \) and \( r(I \cup \{x_1\}) = r(I) \) for every \( k \in [n] \). By submodularity,
\[
\begin{align*}
r(I \cup \{x_1, x_2\}) &\leq r(I \cup x_1) + r(I \cup x_2) - r(I) = r(I), \\
r(I \cup \{x_1, x_2, x_3\}) &\leq r(I \cup \{x_1, x_2\}) + r(I \cup x_3) - r(I) = r(I), \\
&\vdots \\
r(I \cup \{x_1, x_2, \ldots, x_n\}) &\leq r(I \cup \{x_1, \ldots, x_{n-1}\}) + r(I \cup x_n) - r(I) = r(I),
\end{align*}
\]
54
and equality must hold throughout. But then \( r(I \cup J) = r(I) < r(J) \), which is a contradiction.

**Part 2:** Now suppose that \( \mathcal{I} \) is an independence system on \( E \), and define a function \( r : 2^E \to \mathbb{Z} \) as in (4.4.1b). It is immediate from the definition that \( r(A) \leq |A| \) and that \( A \subseteq B \) implies \( r(A) \leq r(B) \) for all \( A,B \in \mathcal{I} \).

To prove submodularity, let \( A,B \subseteq E \) and let \( I \) be a basis of \( A \cap B \). By Lemma 4.4.2, 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 \( A \cap B \). Therefore, if we let \( J_A = J \cap A \) and \( J_B = J \cap B \), then

\[
J_A \cap J_B = I \quad \text{and} \quad J_A \cup J_B = J.
\]

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

\[
r(A \cup B) + r(A \cap B) = |I| + |J| = |J \cap A| + |J \cap B| \leq r(A) + r(B). \tag*{\square}
\]

If \( M = M(G) \) is a graphic matroid, the associated independence system \( \mathcal{I} \) 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 \( r(A \setminus e) < r(A) \) if and only if deleting \( e \) breaks a component of \( G|A \) into two smaller components (so that in fact \( r(A \setminus e) = r(A) - 1 \)). 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, \( I \cup e \) is acyclic.

We return to the question of which abstract simplicial complexes are matroid independence complexes. Here is the promised generalization of Lemma 4.4.2.

**Theorem 4.4.4.** Let \( \Delta \) be an abstract simplicial complex on \( E \). The following are equivalent:

1. \( \Delta \) is a matroid independence complex.
2. For every \( F \subseteq E \), the induced subcomplex \( \Delta|_F = \{ \sigma \in \Delta \mid \sigma \subseteq F \} \) is shellable.
3. For every \( F \subseteq E \), the induced subcomplex \( \Delta|_F \) is Cohen-Macaulay.
4. For every \( F \subseteq E \), the induced subcomplex \( \Delta|_F \) is pure.

Conditions (2) and (3) make sense only if you know what “shellable” and “Cohen-Macaulay” mean; we will get to that in Section 7.

**Proof.** The implications (2) \( \implies \) (3) \( \implies \) (4) are consequences of the material in Section 7 (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 \( I \cup \{x\} \) is a face of \( \Delta \), hence of \( \Delta|_F \).

(4) \( \implies \) (2): More interesting; left as an exercise. \( \square \)
The bases of $M$ (the maximal independent sets) provide another way of defining a matroid. Notational remark: Here and subsequently, notation such as $B \setminus e \cup e'$ should be read as $(B \setminus e) \cup e'$, not as $B \setminus \{e, e'\}$.

**Definition 4.4.5.** A (matroid) basis system on $E$ is a family $\mathcal{B} \subseteq 2^E$ such that for all $B, B' \in \mathcal{B}$,

1. $(B_1)$ $|B| = |B'|$;
2. $(B_2)$ For all $e \in B \setminus B'$, there exists $e' \in B' \setminus B$ such that $(B \setminus e) \cup e' \in \mathcal{B}$;
3. $(B_2')$ For all $e \in B \setminus B'$, there exists $e' \in B' \setminus B$ such that $(B' \cup e) \setminus e' \in \mathcal{B}$.

In fact, given $(B_1)$, the conditions $(B_2)$ and $(B_2')$ are equivalent, although this require some proof (Exercise 4.2).

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 $(B_2)$.

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 $(B_2)$. Let $G$ be a connected graph, let $B, B'$ be spanning trees, and let $e \in B \setminus B'$. Then $B \setminus e$ 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 $B \setminus e \cup e'$ is a spanning tree.

As for $(B_2')$, if $e \in B \setminus B'$, then $B' \cup e$ must contain a unique cycle $C$ (formed by $e$ together with the unique path in $B'$ between the endpoints of $e$). Deleting any edge $e' \in C \setminus e$ will produce a spanning tree.
If $G$ is a graph with edge set $E$ and $M = M(G)$ is its graphic matroid, then

\[ \mathcal{I} = \{ A \subseteq E \mid A \text{ is acyclic} \}, \]

\[ \mathcal{B} = \{ A \subseteq E \mid A \text{ is a spanning forest of } G \}. \]

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

\[ \mathcal{I} = \{ A \subseteq S \mid A \text{ is linearly independent} \}, \]

\[ \mathcal{B} = \{ A \subseteq S \mid A \text{ is a basis for } \text{span}(S) \}. \]

**Proposition 4.4.6.** Let $E$ be a finite set.

1. If $\mathcal{I}$ is an independence system on $E$, then the family of maximal elements of $\mathcal{I}$ is a basis system.
2. If $\mathcal{B}$ is a basis system, then $\mathcal{I} = \bigcup_{B \in \mathcal{B}} 2^B$ 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. So Proposition 4.4.6 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 4.4.7.** A (matroid) circuit system on $E$ is a family $\mathcal{C} \subseteq 2^E$ such that, for all $C,C' \in \mathcal{C}$,

1. $\emptyset \notin \mathcal{C}$;
2. $C \not\subseteq C'$;
3. For all $e \in C \cap C'$, the set $C \cup C' \setminus e$ contains an element of $\mathcal{C}$.

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 $C \cup C' \setminus e$ is dependent, hence contains a circuit.

In a graph, if two cycles $C,C'$ meet in a (non-loop) edge $e = xy$, then $C \setminus e$ and $C' \setminus e$ 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 4.4.8.** Let $E$ be a finite set.
1. If $\mathcal{I}$ is an independence system on $E$, then $\{C \notin \mathcal{I} \mid C' \in \mathcal{I} \forall C' \subseteq C\}$ is a circuit system.
2. If $\mathcal{E}$ is a circuit system, then $\{I \subseteq E \mid C \subseteq I \forall C \in \mathcal{E}\}$ 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. 7.3.1). 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 $\Delta$ be an abstract simplicial complex on $E$ (see Definition 4.4.1). 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 $\Delta$ and $w$, is the following “greedy” algorithm (known as Kruskal’s algorithm):

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

**Proposition 4.4.9.** $\Delta$ 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 10.17.2 below.

<table>
<thead>
<tr>
<th>Summary of Matroid Axiomatizations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Geometric lattice:</strong> lattice that is atomic and semimodular. Corresponds to a simple matroid.</td>
</tr>
<tr>
<td><strong>Rank function:</strong> function $r : 2^E \to \mathbb{N}$ such that $r(A) \leq</td>
</tr>
<tr>
<td><strong>Closure operator:</strong> function $2^E \to 2^E$, $A \mapsto \bar{A}$ such that $A \subseteq \bar{A} = \overline{\overline{A}}$; $A \subseteq B \implies \bar{A} \subseteq B$; and $x \notin A$, $x \in \bar{A} \cup y \implies y \in \overline{A \cup x}$. Simple if $\bar{A} = A$ whenever $</td>
</tr>
<tr>
<td><strong>Independence system:</strong> set family $\mathcal{I} \subseteq 2^E$ such that $\emptyset \in \mathcal{I}$; $I \in \mathcal{I}$, $I' \subseteq I \implies I' \in \mathcal{I}$; and $I, J \in \mathcal{I}$, $</td>
</tr>
<tr>
<td><strong>Basis system:</strong> set family $\mathcal{B} \subseteq 2^E$ such that $\emptyset \in \mathcal{B}$; $I \in \mathcal{B}$, $I' \subseteq I$ \implies $I' \in \mathcal{B}$; and $I, J \in \mathcal{B}$, $</td>
</tr>
<tr>
<td><strong>Circuit system:</strong> set family $\mathcal{C} \subseteq 2^E$ such that no element contains any other, and $C, C' \in \mathcal{C}$, $e \in C \cap C'$ \implies $\exists C'' \in \mathcal{C}$ : $C'' \subseteq C \cup C' \setminus e$. Simple if all elements have size at least 3.</td>
</tr>
<tr>
<td><strong>Greedy algorithm:</strong> simplicial complex $\Delta$ on $E$ such that the greedy algorithm successfully constructs a maximum-weight facet for every weight function $w : E \to \mathbb{R}_{\geq 0}$.</td>
</tr>
</tbody>
</table>

### 4.5 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 4.5.1.** Let $M$ be a matroid and $V$ a vector space over a field $k$. A set of vectors $S \subseteq V$ represents or realizes $M$ over $k$ if the linear matroid $M(S)$ associated with $S$ is isomorphic to $M$.

For example:

- The matroid $U_2(3)$ is representable over $F_2$ (in fact, over any field): we can take $S = \{(1, 0), (0, 1), (1, 1)\}$, and any two of these vectors form a basis of $F_2^2$.
- If $k$ has at least three elements, then $U_2(4)$ is representable, by, e.g., $S = \{(1, 0), (0, 1), (1, 1), (1, a)\}$ where $a \in k \setminus \{0, 1\}$. Again, any two of these vectors form a basis of $k^2$.
- On the other hand, $U_2(4)$ is not representable over $F_2$, because $F_2^2$ 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 $F_q$ of order $q$. Then each element of $E$ must be represented by some nonzero vector in $F_q^n$, and no two vectors can be scalar multiples of each other. Therefore,

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

**Example 4.5.2. The Fano plane.** Consider the affine point configuration with 7 points and 7 lines (one of which looks like a circle), as shown:

![Diagram of the Fano plane]

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 $F_2$ by the columns of the matrix

$$
\begin{bmatrix}
1 & 1 & 0 & 0 & 0 & 1 & 1 \\
0 & 1 & 1 & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 1 & 1 & 1
\end{bmatrix}
\in (F_2)^{3 \times 7}
$$

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 $F_2^3$.

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\(^2\): 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 4.5.3** (Finite projective planes). Let \(q \geq 2\) be a positive integer. A **projective plane of order\(^2\) \(q\)** 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 \(\mathbb{P}^2_q\) whose points and lines are the 1- and 2-dimensional subspaces \(\mathbb{F}^3_q\), 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 \(2\) (mod \(4\)), 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 \(\mathbb{P}^2_q\). 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 \(U_k(n)\) can be represented over \(\mathbb{F}_q\) only if \(k \leq q^n - 1\) (so that there are enough nonzero vectors in \(\mathbb{F}^n_q\)). For example, \(U_2(4)\) is not representable over \(\mathbb{F}_2\). However, this inequality

\(^2\text{But not self-dual as a matroid in the sense to be defined in §4.7.}\)
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 \(-1\).

**Theorem 4.5.4.** 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 4.5.5.** 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 \( e_i - e_j \) for \( ij \in E \). (Or \( e_j - e_i \); it doesn’t matter, since scaling a vector does not change the matroid.) Here \( \{e_1, \ldots, e_n\} \) 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

\[
(e_{v_1} - e_{v_2}) + (e_{v_2} - e_{v_3}) + \cdots + (e_{v_k} - e_{v_1}) = 0,
\]

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 \( \pm 1 \). 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 \( \pm 1 \) and the rest zero. Expanding on that row gives \( \det X_{WB} = \pm \det X_{W\setminus w, B \setminus b} \), 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. \( \square \)

**Example 4.5.6.** 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 4.5.7.** 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

\[
x = \overline{AB} \cap \overline{Ac}, \quad y = \overline{Ac} \cap \overline{Bc}, \quad z = \overline{Bc} \cap \overline{Ac}.
\]
Accordingly, there is a rank-3 simple matroid on ground set $E = \{a, b, c, A, B, C, x, y, z\}$ whose flats are

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

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 $\pm 1$ coefficients), it follows that $NP$ is not representable over any field.

## 4.6 Direct sum

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

**Definition 4.6.1.** Let $M_1, M_2$ be matroids on disjoint sets $E_1, E_2$, with basis systems $\mathcal{B}_1, \mathcal{B}_2$. The **direct sum** $M_1 \oplus M_2$ is the matroid on $E_1 \cup E_2$ with basis system

$$\mathcal{B} = \{B_1 \cup B_2 \mid B_1 \in \mathcal{B}_1, \ B_2 \in \mathcal{B}_2\}.$$ 

I will omit the routine proof that $\mathcal{B}$ 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).
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 4.6.2.** A matroid that cannot be written nontrivially as a direct sum of two smaller matroids is called **connected** or **indecomposable.**

**Proposition 4.6.3.** 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’ll prove the other direction later (maybe).

**Remark 4.6.4.** 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 algebra $2^{[n]}$ (and, for that matter, so is the lattice of flats).

In light of Proposition 4.6.3, 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 2-connected graphs below are not isomorphic, but have isomorphic graphic matroids.

---

3The first term is more common among matroid theorists, but I prefer “indecomposable” to avoid potential confusion with the graph-theoretic meaning of “connected”.
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
\[
r_{M_1 \oplus M_2}(A_1 \cup A_2) = r_{M_1}(A_1) + r_{M_2}(A_2).
\]
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 \( i \).

### 4.7 Duality

**Definition 4.7.1.** Let \( M \) be a matroid on ground set \(|E|\) with basis system \( \mathcal{B} \). The dual matroid \( M^* \) (also known as the orthogonal matroid to \( M \) and denoted \( M^\perp \)) has basis system
\[
\mathcal{B}^* = \{E \setminus B \mid B \in \mathcal{B}\}.
\]
We often write \( e^* \) for elements of the ground set when talking about their behavior in the dual matroid.

Clearly the elements of \( \mathcal{B}^* \) 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 (Exercise 4.2) then you also believe that \( \mathcal{B}^* \) 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 \( B \in \mathcal{B} \) if and only if \( E \setminus A \supseteq E \setminus B \)), and vice versa. The rank function \( r^* \) of the dual is given by
\[
r^*(A) = r(E \setminus A) + |A| - r(E) \quad \forall A \subseteq E.
\]
(4.7.1)
The proof is left as Exercise 4.7.

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

Let \( Y \) be any \((n - r) \times n\) matrix with rowspan(\(Y\)) = nullspace(\(X\)). That is, the rows of \( Y \) span the orthogonal complement of rowspan(\(X\)), according to the standard inner product.

**Theorem 4.7.2.** With this setup, the columns of \( Y \) are a representation for \( M^* \).

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

**Example 4.7.3.** 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 for $Y$ 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).

**Proof of Theorem 4.7.2.** First, note that invertible row operations on a matrix $X \in k^{r \times n}$ (i.e., multiplication on the left by an element of $GL_r(F)$) do not change the matroid represented by its columns; they simply change the basis of $k^r$.

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

**Definition 4.7.4.** Let $M$ be a matroid on $E$. A **loop** is an element of $E$ that does not belong 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 4.4.7). Set-theoretically, a cocircuit is a minimal set not contained in any basis of $M^*$, so it is a minimal set that meets every basis of $M$. 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.

### 4.8 Deletion and contraction

**Definition 4.8.1.** Let $M$ be a matroid on $E$ with basis system $\mathcal{B}$, and let $e \in E$.

1. If $e$ is not a coloop, then the **deletion** of $e$ is the matroid $M\backslash e$ (or $M - e$) on $E \setminus \{e\}$ with bases
   \[ \{B \mid B \in \mathcal{B}, e \notin B\}. \]

2. If $e$ is not a loop, then the **contraction** of $e$ is the matroid $M/e$ (or $M : e$) on $E \setminus \{e\}$ with bases
   \[ \{B \setminus e \mid B \in \mathcal{B}, e \in B\}. \]

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.,

\[ M(G\backslash e) = M(G)\backslash e, \quad M(G/e) = M(G)/e. \]  

\[(4.8.1)\]
Notice that contracting can cause some edges to become parallel, and can cause other edges (namely, those parallel to the edge you’re contracting) 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.

Deletion and contraction are interchanged by duality:

\[(M\setminus e)^* \cong M^*/e^* \quad \text{and} \quad (M/e)^* \cong M^*/e^*\]  \hspace{1cm} (4.8.2)

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 \(k\), let \(E \subseteq V\) be a set of vectors spanning \(V\), let \(M = M(E)\), and let \(e \in E\). Then:

1. \(M \setminus e = M(E \setminus \{e\})\). (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 \(E \setminus \{e\}\) in the quotient space \(V/ke\). (For this quotient to be nontrivial, \(e\) must not be a loop.)

Thus both operations preserve representability over \(k\). For instance, to find an explicit representation of \(M/e\), apply a change of basis to \(V\) so that \(e\) is the \(i\)th standard basis vector, then simply erase the \(i\)th coordinate of every vector in \(E \setminus \{e\}\).

A matroid \(M'\) obtained from \(M\) by some sequence of deletions and contractions is called a **minor** of \(M\).

**Proposition 4.8.2.** Every minor of a graphic (resp., linear, uniform) matroid is graphic (resp., linear, uniform).

**Proof.** The graphic case follows from (4.8.1), and the linear case from the previous discussion. For uniform matroids, the definitions imply that

\[U_k(n)\setminus e \cong U_k(n-1) \quad \text{and} \quad U_k(n)/e \cong U_{k-1}(n-1)\]

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 4.8.1.

**Proposition 4.8.3.** 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

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

**Example 4.8.4.** If \(M \cong U_k(n)\), then \(b(M) = \binom{n}{k}\), and the recurrence of Proposition 4.8.3 is just the Pascal relation \(\binom{n}{k} = \binom{n-1}{k} + \binom{n}{k-1}\).

Deletion and contraction can be described nicely in terms of the independence complex. If \(\Delta = \mathcal{I}(M)\) is the independence complex of \(M - e\) and \(M/e\), then

\[\mathcal{I}(M\setminus e) = \text{del}_\Delta(e) = \{\sigma \in \Delta \mid e \notin \sigma\},\]
\[\mathcal{I}(M/e) = \text{lk}_\Delta(e) = \{\sigma \in \Delta \mid e \notin \sigma \text{ and } \sigma \cup e \in \Delta\}.\]
These subcomplexes of $\Delta$ are known respectively as the deletion and link of $e$. Informally speaking, the deletion is the part of the complex that is far away from $e$, while the link is what you see if you stand at $e$ and look around you.

Recall that the reduced Euler characteristic of $\Delta$ is

$$\tilde{\chi}(\Delta) = \sum_{\sigma \in \Delta} (-1)^{\dim \sigma} = \sum_{\sigma \in \Delta} (-1)^{|\sigma|-1}. \quad (4.8.3)$$

This important topological invariant also satisfies a deletion-contraction recurrence. For any $e \in V$,

$$\tilde{\chi}(\Delta) = \sum_{\sigma \in \Delta : e \notin \sigma} (-1)^{\dim \sigma} + \sum_{\sigma \in \Delta : e \in \sigma} (-1)^{\dim \sigma}$$

$$= \sum_{\sigma \in \text{del}_\Delta(e)} (-1)^{\dim \sigma} + \sum_{\tau \in \text{lk}_\Delta(e)} (1+\dim \tau)$$

$$= \tilde{\chi}(\text{del}_\Delta(e)) - \tilde{\chi}(\text{lk}_\Delta(e)). \quad (4.8.4)$$

These observations are the tip of an iceberg that we will explore in Section 5.

### 4.9 Exercises

**Exercise 4.1.** Determine, with proof, all pairs of integers $k \leq n$ such that there exists a graph $G$ with $\mathcal{M}(G) \cong U_k(n)$. (Here $U_k(n)$ denotes the uniform matroid of rank $k$ on $n$ elements; see Example 4.2.4.). Hint: Use Proposition 4.8.2.

**Exercise 4.2.** Prove the equivalence of the two forms of the basis exchange condition (B2) and (B2'). (Hint: Examine $|B \setminus B'|$.)

**Exercise 4.3.** (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 $B \setminus e \cup \phi(e)$ is a basis of $M$ for every $e \in B \setminus B'$.

**Exercise 4.4.** Prove Proposition 4.4.8, which describes the equivalence between matroid independence systems and matroid circuit systems.

**Exercise 4.5.** Complete the proof of Theorem 4.4.4 by showing that if $\Delta$ is a simplicial complex in ground set $V$ in which every induced subcomplex is pure, then $\Delta$ is shellable. To do this, pick a vertex $v$. Use induction to show that the two complexes

$$\Delta_1 = \langle \text{facets } F \in \Delta \mid v \notin F \rangle,$$

$$\Delta_2 = \langle \text{facets } F \in \Delta \mid v \in F \rangle$$

are both shellable, then concatenate the shelling orders to produce a shelling order on $\Delta$. Derive a relationship among the $h$-polynomials of $\Delta$, $\Delta_1$, and $\Delta_2$.

**Exercise 4.6.** 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 $r(E) = r(E_1) + \cdots + r(E_n)$. 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.)

**Exercise 4.7.** 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 $A \subseteq E$. 

68
Exercise 4.8. Let $E$ be a finite set and let $\Delta$ 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 $\Delta$, stop.

Otherwise, find $e \in E \setminus A$ of maximal weight such that $A \cup \{e\} \in \Delta$ (if there are several such $e$, pick one at random), and replace $A$ with $A \cup \{e\}$.

**Step 3:** Repeat Step 2 until $A$ is a facet of $\Delta$.

This algorithm may or may not work for a given simplicial complex $\Delta$ and weight function $w$. 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) $\Delta$ is a matroid independence system.

Note: This result does follow from Theorem 4.4.4. However, that is a substantial result, so don’t use it unless you first do Exercise 4.5. It is possible to do this exercise by working directly with the definition of a matroid independence system.

Exercise 4.9. Prove that duality interchanges deletion and contraction, as in equation (4.8.2).

Exercise 4.10. 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 $\mathcal{I}$ be the family of all subsets of $X$ that have a transversal; in particular $\mathcal{I}$ is a simplicial complex.

Prove that $\mathcal{I}$ 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.

Exercise 4.11. (Requires a bit of abstract algebra.) Let $n$ be a positive integer, and let $\zeta$ be a primitive $n^{th}$ 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 $\phi$ function. Prove the following:

(a) if $n$ is prime, then $Y_n \cong U_{n-1}(n)$.

(b) if $m$ is the square-free part of $n$ (i.e., the product of all the distinct primes dividing $n$ — e.g., the square-free 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 $K_{p,q}$.
5.1 The two definitions of the Tutte polynomial

For $A \subseteq E$, we define

$$\text{corank } A = r(E) - r(A),$$

$$\text{nullity } A = |A| - r(A).$$

Corank and nullity measure how far $A$ is from being spanning and independent, respectively. 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 5.1.1.** The Tutte polynomial of $M$ is

$$T_M = T_M(x, y) := \sum_{A \subseteq E} (x - 1)^{r(E) - r(A)}(y - 1)^{|A| - r(A)}. \quad (5.1.1)$$

**Example 5.1.2.** 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 5.1.3.** For uniform matroids, corank and nullity depend only on cardinality, making the Tutte polynomials family 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 5.1.4.** Let \( G \) be the graph below (known as the “diamond”):

![Diagram](image)

The formula (5.1.1) gives

| \( A \)         | \(|A|\) | \( r(A)\) | \(|A| - r(A)\) | Contribution to (5.1.1) |
|----------------|--------|------------|----------------|--------------------------|
| 1 empty set    | 0      | 0          | 3              | 0                        |
| 5 singletons   | 1      | 1          | 2              | 0                        |
| 10 doubletons  | 2      | 2          | 1              | 0                        |
| 2 triangles    | 3      | 2          | 1              | 1                        |
| 8 spanning trees | 3  | 3          | 0              | 0                        |
| 5 quadruplets  | 4      | 3          | 0              | 1                        |
| 1 whole set    | 5      | 3          | 0              | 2                        |
| 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. \( T_M(2, 2) = \sum_{A \subseteq E} 1 = 2^{|E|} \).
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 \( 0^0 = 1 \). So \( T_M(1, 1) = b(M) \), the number of bases. We previously observed that this quantity satisfies a deletion/contraction recurrence; 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 of the independence complex of \( M \):
   \[
   T_M(0, 1) = \sum_{A \subseteq E} (-1)^{r(E) - r(A)}q^{|A| - r(A)} = \sum_{A \subseteq E \text{ independent}} (-1)^{r(E) - r(A)}
   \]
   \[
   = (-1)^{r(E)} \sum_{A \subseteq \mathcal{I}(M)} (-1)^{|A|}
   \]
   \[
   = (-1)^{r(E)-1} \tilde{\chi}(\mathcal{I}(M)).
   \]
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 5.1.5.** 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 $T_M = y T_{M \setminus e}$.
- **(T2b)** If $e \in E$ is a coloop, then $T_M = x T_{M/e}$.
- **(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 5.1.6.** 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 5.1.7.** Let $M \cong U_1(2)$ (the graphic matroid of the “digon”, two vertices joined by two parallel edges). Let $e \in E$; then

$$T_M = T_{M \setminus e} + T_{M/e} = T(U_1(1)) + T(U_0(1)) = x + y.$$  

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 (5.1.3).

The Tutte recurrence says we can represent a calculation of $T_M$ by a binary tree in which moving down corresponds to deleting or contracting:

```
          M
         /   \   /
M \setminus e  M \setminus f  M \setminus e \setminus f  M \setminus e / f  M \setminus f / e  M \setminus f / e / f
            :            :            :            :
```

**Example 5.1.8.** Consider the diamond of Example 5.1.4. 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

$$T(G \setminus a) = x(x^2 + x + y)$$
by Example 5.1.7. Next, apply the Tutte recurrence to the edge $b$ in $G/a$. The graph $G/a\setminus b$ has a coloop $c$, contracting which produces a digon. Meanwhile, $M(G/a\setminus b) \cong U_1(3)$. Therefore

$$T(G/a\setminus b) = x(x+y) \quad \text{and} \quad T(G/a\setminus b) = x + y + y^2.$$ 

Putting it all together, we get

$$T(G) = T(G\setminus a) + T(G/a)$$
$$= T(G\setminus a) + T(G/a\setminus b) + T(G/a/b)$$
$$= x(x^2 + x + y) + x(x + y) + (x + y + y^2)$$
$$= x^3 + 2x^2 + 2xy + x + y + y^2.$$

On the other hand, we could have recursed first on $e$, getting

$$T(G) = T(G\setminus e) + T(G/e)$$
$$= T(G\setminus e\setminus c) + T(G\setminus e/c) + T(G/e\setminus c) + T(G/e/c)$$
$$= x^3 + (x^2 + x + y) + x(x + y) + y(x + y)$$
$$= x^3 + 2x^2 + 2xy + x + y + y^2.$$

Proof of Theorem 5.1.5. Let $M$ be a matroid on ground set $E$. The definitions of rank function, deletion, and contraction imply that for any $e \in E$ and $A \subseteq E \setminus \{e\}$:

$$73$$
1. If \( e \) is not a coloop, then \( r'(A) = r_{M\setminus\{e\}}(A) = r_M(A) \).
2. If \( e \) is not a loop, then \( r''(A) = r_{M/e}(A) = r_M(A \cup e) - 1 \).

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

\[
T_M = \sum_{A \subseteq E} X^{r(E) - r(A)} Y^{|A| - r(A)} \\
= \sum_{A \subseteq E; e \notin A} X^{r(E) - r(A)} Y^{|A| - r(A)} + \sum_{B \subseteq E; e \in B} X^{r(E) - r(B)} Y^{|B| - r(B)} \\
= \sum_{A \subseteq E \setminus \{e\}} X^{r'(E \setminus \{e\}) - r'(A)} Y^{|A| - r'(A)} + \sum_{A \subseteq E \setminus \{e\}} X^{r'(E \setminus \{e\}) - r'(A)} Y^{|A| - r'(A) + 1 - r'(A)} \\
= (1 + Y) \sum_{A \subseteq E \setminus \{e\}} X^{r'(E \setminus \{e\}) - r'(A)} Y^{|A| - r'(A)} = yT_{M \setminus \{e\}}.
\]

For (T2b), let \( e \) be a coloop. Then

\[
T_M = \sum_{A \subseteq E} X^{r(E) - r(A)} Y^{|A| - r(A)} \\
= \sum_{e \notin A \subseteq E} X^{r(E) - r(A)} Y^{|A| - r(A)} + \sum_{e \in B \subseteq E} X^{r(E) - r(B)} Y^{|B| - r(B)} \\
= \sum_{A \subseteq E \setminus \{e\}} X^{r''(E \setminus \{e\}) - r''(A)} Y^{|A| - r''(A)} + \sum_{A \subseteq E \setminus \{e\}} X^{r''(E \setminus \{e\}) - r''(A)} Y^{|A| - r''(A) + 1 - r''(A)} \\
= (X + 1) \sum_{A \subseteq E \setminus \{e\}} X^{r''(E \setminus \{e\}) - r''(A)} Y^{|A| - r''(A)} = xT_{M/e}.
\]

For (T3), suppose that \( e \) is ordinary. Then

\[
T_M = \sum_{A \subseteq E} X^{r(E) - r(A)} Y^{|A| - r(A)} \\
= \sum_{A \subseteq E \setminus \{e\}} \left[ X^{r(E) - r(A)} Y^{|A| - r(A)} \right] + \left[ X^{r(E) - r(A \cup \{e\})} Y^{|A \cup \{e\}| - r(A \cup \{e\})} \right] \\
= \sum_{A \subseteq E \setminus \{e\}} \left[ X^{r'(E \setminus \{e\}) - r'(A)} Y^{|A| - r'(A)} \right] + \left[ X^{r''(E + 1) - r''(A) + 1} Y^{|A| + 1 - r''(A) - 1} \right] \\
= \sum_{A \subseteq E \setminus \{e\}} X^{r'(E \setminus \{e\}) - r'(A)} Y^{|A| - r'(A)} + \sum_{A \subseteq E \setminus \{e\}} X^{r''(E \setminus \{e\}) - r''(A)} Y^{|A| - r''(A)} \\
= T_{M \setminus \{e\}} + T_{M/e}.
\]

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.

74
2. Duality interchanges $x$ and $y$, i.e.,

$$T_M(x, y) = T_M^\ast(y, x). \tag{5.1.2}$$

This can be deduced either from the Tutte recurrence (since duality interchanges deletion and contraction; see (4.8.2)) or from the corank-nullity generating function, by expressing $r^\ast$ in terms of $r$ (see Exercise 4.7).

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 (5.1.1).

### 5.2 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 5.2.1** (Tutte Recipe Theorem for Matroids). Let $u(M)$ be a matroid isomorphism invariant that satisfies a recurrence of the form

$$u(M) = \begin{cases} 
1 & \text{if } E = \emptyset, \\
Xu(M/e) & \text{if } e \in E \text{ is a coloop}, \\
Yu(M\setminus e) & \text{if } e \in E \text{ is a loop}, \\
anu(M/e) + bun(M\setminus e) & \text{if } e \in E \text{ is ordinary}.
\end{cases}$$

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

$$r(M) = r(M\setminus e) = r(M/e) + 1 \quad \text{and} \quad n(M) = n(M\setminus e) + 1 = n(M/e)$$

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

$$a^r b^n \tilde{u}(M^E) = \begin{cases} 
1 & \text{if } E = \emptyset, \\
Xa^{-1}b^n\tilde{u}(M/e) & \text{if } e \in E \text{ is a coloop}, \\
Ya^{-1}b^n\tilde{u}(M\setminus e) & \text{if } e \in E \text{ is a loop}, \\
a^r b^n u(M/e) + a^r b^n \tilde{u}(M\setminus e) & \text{if } e \in E \text{ is ordinary}.
\end{cases}$$

Setting $X = xa$ and $Y = yb$, we see that $\tilde{u}(M) = T_M(x, y) = T_M(X/a, Y/b)$ by Theorem 5.1.5, and rewriting in terms of $u(M)$ gives the desired formula. \(\square\)

Bollobás [Bol98, p.340] gives an analogous result for graphs:
Theorem 5.2.2 (Tutte Recipe Theorem for Graphs). Let \( u(G) \) be a graph isomorphism invariant that satisfies a recurrence of the form

\[
    u(G) = \begin{cases} 
        a |V| & \text{if } E = \emptyset, \\
        Xu(G\setminus e) & \text{if } e \in E \text{ is a coloop}, \\
        Y u(G\setminus e) & \text{if } e \in E \text{ is a loop}, \\
        bu(G\setminus e) + cu(G/e) & \text{if } e \in E \text{ is ordinary},
    \end{cases}
\]

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, and that deletion of a cut-edge makes sense in the context of graphs (no, that's not a misprint in the coloop case of Theorem 5.2.2!). The invariant \( U \) is described by Bollobás as "the universal form of the Tutte polynomial."

5.3 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 \( i, e : \mathcal{B}(M) \to \mathbb{N} \) such that

\[
    T_M(x,y) = \sum_{B \in \mathcal{B}(M)} x^{i(B)} y^{e(B)}.
\]

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 5.3.1. Let \( M \) be a matroid on \( E \) with basis system \( \mathcal{B} \) and let \( B \in \mathcal{B} \). For \( e \in B \), the fundamental cocircuit of \( e \) with respect to \( B \), denoted \( C^*(e,B) \), is the unique cocircuit in \( (E \setminus B) \cup e \). That is,

\[
    C^*(e,B) = \{ e' | B \setminus e \cup e' \in \mathcal{B} \}.
\]

Dually, for \( e \notin B \), then the fundamental circuit of \( e \) with respect to \( B \), denoted \( C(e,B) \), is the unique circuit in \( B \cup e \). That is,

\[
    C(e,B) = \{ e' | B \setminus e' \cup e \in \mathcal{B} \}.
\]

In other words, the fundamental cocircuit is the set of all elements outside \( B \) that could replace \( e \) in a basis exchange, while the fundamental circuit is the set of all elements outside \( B \) that could be replaced by \( e \).

Suppose that \( M = M(G) \), where \( G \) is a connected graph, and \( B \) is a spanning tree. For all \( e \in B \), the graph \( B \setminus e \) has two components \( X, Y \), 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 \( B \cup e \) 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 \( B \setminus e \), and \( C(e,B) \) is the unique linearly dependent subset of \( B \cup e \).
**Definition 5.3.2.** 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 $i(B) = \# \{e \in B \mid e \text{ is internally active with respect to } B\}$\hspace{1cm}e(B) = \# \{e \in E \setminus B \mid e \text{ is externally active with respect to } B\}$

Note that these numbers depend on the choice of ordering of $E$.

**Example 5.3.3.** 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:

- $C(e_1, B) = \{e_1, e_4, e_5\}$ so $e_1$ is externally active;
- $C(e_3, B) = \{e_2, e_3, e_5\}$ so $e_3$ is not externally active;
- $C^*(e_2, B) = \{e_2, e_3\}$ so $e_2$ is internally active;
- $C^*(e_4, B) = \{e_1, e_4\}$ so $e_4$ is not internally active;
- $C^*(e_5, B) = \{e_1, e_3, e_5\}$ so $e_5$ is not internally active.

**Theorem 5.3.4.** 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

$$T_M(x, y) = \sum_{B \in \mathcal{B}(M)} x^{i(B)} y^{e(B)}.$$ \hspace{1cm}(5.3.1)

For instance, in Example 5.3.3, the spanning tree $B$ contributes the monomial $xy = x^1y^1$ to $T(G; x, y)$.

The proof of Theorem 5.3.4 is omitted. Like the proof of Theorem 5.1.5, it requires careful bookkeeping but is not really hard—it boils down to showing that the generating function on the right-hand side of (5.3.1) satisfies the Tutte recurrence. 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 (5.3.1) is divisible by $y$ (resp. $x$).
5.4 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 5.4.1.** Let $L$ be a geometric lattice with atoms $E$. Let $M$ be the corresponding matroid on $E$, and let $r$ be its rank function. 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

$$(-1)^{r(M)} T_M(1 - k, 0) = (-1)^{r(A)} \sum_{A \subseteq E} (-k)^{r(A) - r(K)} (-1)^{|A| - r(A)}$$

$$= \sum_{A \subseteq E} (-1)^{|A|} k^{r(M) - r(A)}$$

$$= \sum_{K \in L} \left( \sum_{A \subseteq E : \bar{A} = K} (-1)^{|A|} \right) k^{r(M) - r(K)}.$$

We now claim that $f(K) = \mu_L(\emptyset, K)$, which we will prove by Möbius inversion on $L$. For a flat $K \in L$, let

$$g(K) = \sum_{A \subseteq E : A \subseteq K} (-1)^{|A|} = \sum_{J \leq K} f(J)$$

so by (3.3.1a)

$$f(K) = \sum_{J \leq K} \mu(J, K) g(J).$$

But $g(J)$ is the binomial expansion of $(1 - 1)^{|J|}$, which is zero unless $J = \emptyset$. So the previous equation reduces to $f(K) = \mu(\emptyset, K)$ as desired. \hfill \Box$

Theorem 5.4.1 gives another way to prove that the Möbius function of a semimodular lattice weakly alternates in sign (Theorem 3.5.7). First, we can reduce from the semimodular to the geometric case by Corollary 3.5.10 (if $L$ is a lattice in which $1$ is not a join of atoms, then $\mu(L) = 0$). Recall that $\mu(L) = \chi(L; 0)$, which by Theorem 5.4.1 is $(-1)^{r(M)} T_M(1, 0)$. But $T_M(1, 0) \geq 0$ for every matroid $M$, because $T_M \in \mathbb{N}[x, y]$.

For a graphic matroid, the characteristic polynomial has a classical combinatorial interpretations in terms of colorings, as we now explain.

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 3.3.4 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 $p_G(k) = k \chi_K(G)(k)$. 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 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 5.4.2.** 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
\[
\begin{align*}
p_G(k) &= k^n & \text{if } E = \emptyset; & (5.4.1) \\
p_G(k) &= 0 & \text{if } G \text{ has a loop;} & (5.4.2) \\
p_G(k) &= (k-1)p_{G/e}(k) & \text{if } e \text{ is a coloop;} & (5.4.3) \\
p_G(k) &= p_{G\setminus e}(k) - p_{G/e}(k) & \text{otherwise.} & (5.4.4)
\end{align*}
\]

We already know (5.4.1) and (5.4.2). 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 (5.4.4) 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 (5.4.3).

The graph \( G \setminus e \) has \( n \) vertices and either \( e+1 \) or \( c \) components, according as \( e \) is or is not a coloop. Meanwhile, \( G/e \) has \( n-1 \) vertices and \( c \) components. By induction,
\[
(-1)^{n-c}k^c T_G(1-k,0) = \begin{cases} 
  k^n & \text{if } E = \emptyset, \\
  0 & \text{if } e \text{ is a loop,} \\
  (1-k)(-1)^{n+1-c} k^c T_{G/e}(1-k,0) & \text{if } e \text{ is a coloop,} \\
  (-1)^{n-c}k^c (T_{G\setminus e}(1-k,0) + T_{G/e}(1-k,0)) & \text{otherwise}
\end{cases}
\]
\[
= \begin{cases} 
  k^n & \text{if } E = \emptyset, \\
  0 & \text{if } e \text{ is a loop,} \\
  (k-1)p_{G/e}(k) & \text{if } e \text{ is a coloop,} \\
  p_{G\setminus e}(k) - p_{G/e}(k) & \text{otherwise}
\end{cases}
\]
which is exactly the recurrence satisfied by the chromatic polynomial. \( \square \)

**Remark 5.4.3.** It is also possible to prove Theorem 5.4.2 by invoking the Tutte Recipe Theorem for Graphs (Theorem 5.2.2). To do so, one would need to replace case (5.4.3) of the chromatic recurrence with the statement \( p_G(k) = \frac{k-1}{k} p_{G-e}(k) \).

79
Recall that 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 5.4.1 and 5.4.2 gives

$$p_G(k) = k^c \chi(K(G); k).$$

## 5.5 Acyclic orientations

An orientation $O$ of a graph $G = (V,E)$ is an assignment of a direction to each edge $xy \in E$ (either $\overrightarrow{xy}$ or $\overrightarrow{yx}$). A directed cycle is a sequence $(x_0, x_1, \ldots, x_{n-1})$ of vertices such that $\overrightarrow{x_i x_{i+1}}$ 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 then $a(G) = 1$. 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 $(O,f)$, where $O$ 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 $\text{Com}(G,k)$ be the set of compatible $k$-pairs of $G$.

**Theorem 5.5.1** (Stanley’s Acyclic Orientation Theorem). *For every graph $G$ and positive integer $k$,*

$$|\text{Com}(G,k)| = (-1)^n p_G(-k) = k^c T_G(1 + k, 0).$$

(5.5.1)

**Proof.** The second equality follows from Theorem 5.4.2, so we prove the first one. Let $n = |G|$.

If $G$ has no edges then $|\text{Com}(G,k)| = k^n = (-1)^n(-k)^n = (-1)^n p_G(-k)$, confirming (5.5.1).

If $G$ has a loop then it has no acyclic orientations, hence no $k$-pairs for any $k$, so both sides of (5.5.1) are zero.
Let $e = xy$ be an edge of $G$ that is not a loop. Denote the left-hand side of (5.5.1) by $\bar{p}_G(k)$. Then
\[
\bar{p}_G(k) = (-1)^n p_G(-k) = (-1)^n (p_G, e(-k) - p_G/e(-k))
\]
\[
= (-1)^n ((-1)^n \bar{p}_G, e(k)) - (-1)^{n-1} \bar{p}_G/e(k))
\]
\[
= \bar{p}_G, e(k) + \bar{p}_G/e(k)
\]
so all we have to show is that $|\text{Com}(G, k)|$ satisfies the same recurrence. Write
\[
\text{Com}(G, k) = \left\{ ((O, f) : f(x) \neq f(y)) \cup \left\{ (O, f) : f(x) = f(y) \right\} \right\}.
\]

Observe that it always possible to extend a $k$-pair of $G \setminus e$ to $G$ by choosing an appropriate orientation for $e$. (If neither choice worked, then the orientation on $G \setminus e$ must contain a directed path from each of $x, y$ to the other, and then the orientation wouldn’t be acyclic.) Let $C_{\text{rev}}$ denote the set of pairs of $G \setminus e$ that are reversible (either choice of orientation for $e$ works), and let $C_{\text{irr}}$ denote those that are irreversible. So $\text{Com}(G \setminus e, k) = C_{\text{rev}} \cup C_{\text{irr}}$.

For $(O, f) \in C'$, we can simply delete $e$ to obtain a $k$-pair of $G \setminus e$. The $k$-pairs we obtain are precisely the irreversible ones.

On the other hand, for $(O, f) \in C''$, deleting $e$ produces a reversible $k$-pair of $G \setminus e$, and contracting $e$ produces a compatible $k$-pair of $G/e$ (whose coloring is well-defined because $f(x) = f(y)$). Since $e$ could, in fact, be oriented in two ways, we obtain every reversible element of $\text{Com}(G \setminus e, k)$, and every element of $\text{Com}(G/e, k)$, twice. That is,
\[
|C''| = 2|C_{\text{rev}}| = 2|\text{Com}(G/e, k)|
\]
but then
\[
|C''| = |C_{\text{rev}}| + |\text{Com}(G/e, k)|
\]
and so
\[
|\text{Com}(G, k)| = |C'| + |C''| = |C_{\text{irr}}| + |C_{\text{rev}}| + |\text{Com}(G/e, k)|
\]
\[
= |\text{Com}(G \setminus e, k)| + |\text{Com}(G/e, k)|
\]
as desired. \[\square\]

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:

**Theorem 5.5.2.** 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 6.5 and/or see a source such as Beck and Robins [BR07], but here is a brief taste.

Let $G$ be a simple graph on $n$ vertices. The **graphic arrangement** $A_G$ 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 $\mathbb{R}^n \setminus A_G$ 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 $A_G$. 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).
5.6 The Tutte polynomial and linear codes

**Definition 5.6.1.** A linear code \( C \) 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 \( C \). The elements \( c = (c_1, \ldots, c_n) \in C \) are called codewords. The support of a codeword is \( \text{supp}(c) = \{ i \in [n] \mid c_i \neq 0 \} \), and its weight is \( \text{wt}(c) = |\text{supp}(c)| \).

The weight enumerator of \( C \) is the polynomial

\[
W_C(t) = \sum_{c \in C} t^{\text{wt}(c)}.
\]

For example, let \( C \) 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 \( C = \{000, 101, 011, 110\} \), and \( W_C(t) = 1 + 3t^2 \).

The dual code \( C^\perp \) is the orthogonal complement under the standard inner product. This inner product is nondegenerate, i.e., \( \dim C^\perp = n - \dim C \). (Note, though, that a subspace and its orthogonal complement can intersect nontrivially. A space can even be its own orthogonal complement, such as \( \mathbb{F}_2^3 \). 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, \( C^\perp = \{000, 111\} \) and \( W_{C^\perp}(t) = 1 + 3t^2 \).

**Theorem 5.6.2** (Curtis Greene, 1976). Let \( C \) 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 \( C \). Then

\[
W_C(t) = t^{n-r}(1-t)^r T_M \left( \frac{1+(q-1)t}{1-t}, \frac{1}{t} \right)
\]

The proof is again a deletion-contraction argument and is omitted. As an example, if \( C = \{000, 101, 011, 110\} \subseteq \mathbb{F}_2^3 \) as above, then the matroid \( M \) is \( U_2(3) \). Its Tutte polynomial is \( x^2 + x + y \), and Greene’s theorem gives

\[
W_C(t) = t(1-t)^2 T_M \left( \frac{1+t}{1-t}, \frac{1}{t} \right)
= t(1-t)^2 \left( \frac{1+t}{1-t} \right)^2 + \left( \frac{1+t}{1-t} \right) + \frac{1}{t}
= t(1+t)^2 + t(1+t)(1-t) + (1-t)^2
= (t + 2t^2 + t^3) + (t - t^3) + (1 - 2t + t^2)
= 1 + 3t^2.
\]

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 (5.1.2), so setting \( s = (1-t)/(1+(q-1)t) \) (so \( t = (1-s)/(1+(q-1)s) \); isn’t that convenient?) gives

\[
W_{C^\perp}(t) = t^r(1-t)^{n-r} T_M \left( \frac{1+(q-1)s}{1-s}, \frac{1}{s} \right)
= t^r(1-t)^{n-r}s^{r-n}(1-s)^{-r}W_C(s),
\]

or rewriting in terms of \( t \),

\[
W_{C^\perp}(t) = \frac{1+(q-1)t^n}{q^r}W_C \left( \frac{1-t}{1+(q-1)t} \right)
\]

which is known as the MacWilliams identity and is important in coding theory.
5.7 Exercises

Exercise 5.1. 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 $T_G(0,2)$.

Exercise 5.2. Let $G = (V,E)$ be a connected graph and $A$ be an abelian group, written additively. An $A$-flow on $G$ is a function $f : E \to A$ such that for every vertex $v$,
$$
\sum_{e \in E(v)} f(e) = 0
$$
where $E(v)$ means the set of edges incident to $v$. Prove that the number of everywhere nonzero $A$-flows is $(-1)^{|E|-|V|-1}T_G(0,1-|A|)$.

Exercise 5.3. 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.

Exercise 5.4. Let $M$ be a matroid on $E$ and let $0 \leq p \leq 1$. 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 $\{X_e \mid e \in E\}$, each of which is 1 with probability $p$ and 0 with probability $1-p$. Let $A = \{e \in E \mid X_e = 1\}$. Then $R_M(p)$ is the probability that $r(A) = r(E)$.) Express $R_M(p)$ as an evaluation of the Tutte polynomial, using the closed-form formula.

Exercise 5.5. 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 Exercise 4.5.)

Exercise 5.6. Merino’s theorem on critical configurations of the chip-firing game.

Exercise 5.7. Prove Theorem 5.3.4.

Exercise 5.8. Prove Theorem 5.6.2.

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 6

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].

6.1 Basic definitions

**Definition 6.1.1.** Let $k$ be a field, typically either $\mathbb{R}$ or $\mathbb{C}$, and let $n \geq 1$. A **linear hyperplane** in $k^n$ is a vector subspace of codimension 1. An **affine hyperplane** is a translate of a linear hyperplane. A **hyperplane arrangement** $\mathcal{A} \subseteq k^n$ is a finite set of (distinct) hyperplanes $H_1, \ldots, H_k \subseteq k^n$. The number $n$ is called the **dimension** of $\mathcal{A}$, and the space $k^n$ is its **ambient space**.

**Definition 6.1.2.** The **intersection poset** $L(\mathcal{A})$ of an arrangement $\mathcal{A}$ is the poset of all intersections of subsets of $\mathcal{A}$, ordered by reverse inclusion. The **characteristic polynomial** of $\mathcal{A}$ is

$$
\chi_{\mathcal{A}}(k) = \sum_{x \in L(\mathcal{A})} \mu(\hat{0}, x) k^{\dim x}. \tag{6.1.1}
$$

This is essentially the same as the characteristic polynomial of $L(\mathcal{A})$, up to a correction factor that we will explain soon.

We will frequently be interested in the complement $X_\mathcal{A} := k^n \setminus \mathcal{A}$. (Here we introduce a convenient abuse of notation: identifying the family of hyperplanes $\mathcal{A}$ with the point set $\bigcup_{H \in \mathcal{A}} H$.)

**Example 6.1.3.** Two line arrangements in $\mathbb{R}^2$ are shown below. The arrangement $\mathcal{A}_1$ is linear; it consists of the lines $x = 0$, $y = 0$, and $x = y$. The arrangement $\mathcal{A}_2$ is affine; it 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.
Their intersection posets of the 2-dimensional arrangements $A_1$ and $A_2$ of Example 6.1.3 are

and their characteristic polynomials are $x^2 - 3x + 2$ and $x^2 - 4x + 5$ respectively.

Every hyperplane is the zero set of some linear form, so $H_1 \cup \cdots \cup H_s$ is the zero set of the product of those $s$ linear forms. We can specify an arrangement concisely by that product, called the **defining polynomial** of $A$ (which is well-defined up to a scalar multiple). For example, the defining polynomials of $A_1$ and $A_2$ are $xy(x - y)$ and $xy(y - 1)(y + 1)$ respectively.

**Example 6.1.4.** The **Boolean arrangement** $B_n$ (or **coordinate arrangement**) consists of the $n$ coordinate hyperplanes in $n$-space. Its defining polynomial is $x_1 x_2 \ldots x_n$, and its intersection poset is the Boolean algebra $B_n$. More generally, any arrangement whose intersection poset is $B_n$ might be referred to as a Boolean arrangement.

**Example 6.1.5.** The **braid arrangement** $Br_n$ consists of the $\binom{n}{2}$ hyperplanes $x_i = x_j$ in $n$-space. Its defining polynomial is $\prod_{1 \leq i < j \leq n}(x_i - x_j)$, the **Vandermonde determinant**. Its intersection poset is naturally identified with the partition lattice $\Pi_n$. 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 in this way. For instance, the intersection poset of $Br_3$ is as follows:

Note that the poset $\Pi_n = L(\text{Br}_n)$ has characteristic polynomial $k^3 - 3k^2 + 2k$, but the arrangement $\text{Br}_3$ has characteristic polynomial $k^2 - 3k + 2$.

**Example 6.1.6.** If $G = (V, E)$ is a simple graph on vertex set $V = [n]$, then the corresponding **graphic arrangement** $A_G$ is the subarrangement of $\text{Br}_n$ consisting of those hyperplanes $x_i = x_j$ for which $ij \in E$. 

85
Thus $B_{rn}$ itself is the graphic arrangement of the complete graph $K_n$. Moreover, the intersection poset of $A_G$ is precisely the connectivity lattice $K(G)$ defined in Example 1.4.3.

Here are some pictures. Note that every hyperplane in $B_{rn}$ contains the line $x_1 = x_2 = \cdots = x_n$, so projecting $\mathbb{R}^4$ along that line allows us to picture $B_{r4}$ as an arrangement $ess(B_{r4})$ in $\mathbb{R}^3$. (The symbol “ess” means essentialization, to be explained soon).

\[ \begin{array}{c}
\mathcal{B}_3 \\
B_{r3} \\
\text{ess}(B_{r4})
\end{array} \]

The second two figures were produced using the computer algebra system Sage [S+14].

The poset $L(A)$ is the fundamental combinatorial invariant of $A$. Some easy observations:

1. If $T : \mathbb{R}^n \to \mathbb{R}^n$ is an invertible linear transformation, then $L(T(A)) \cong L(A)$, where $T(A) = \{ T(H) \mid H \in A \}$. 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 $L(A)$ is a meet-semilattice, with meet given by

\[
\left( \bigcap_{H \in B} H \right) \land \left( \bigcap_{H \in C} H \right) = \bigcap_{H \in B \cap C} H
\]

for all $B, C \subseteq A$. Its $\emptyset$ element is the empty intersection, which by convention is $k^n$.

3. $L(A)$ is ranked, with rank function $r(X) = n - \dim X$. To see this, observe that each covering relation $X \prec Y$ 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. $L(A)$ has a $1$ element if and only if the center $\cap A := \bigcap_{H \in A} H$ is nonempty. Such an arrangement is called central. In this case $L(A)$ is a lattice (and is referred to as the intersection lattice of $A$). Since translation does not affect whether an arrangement is central (or indeed any of its combinatorial structure), we will typically assume that $\cap A$ contains the zero vector, which is to say that every hyperplane in $A$ is a linear hyperplane in $k^n$. (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 $L(\mathbb{k}^n)^*$ (the lattice of all subspaces of $\mathbb{k}^n$ ordered by reverse inclusion). The associated matroid $M(\mathcal{A}) = M(L(\mathcal{A}))$ is represented over $\mathbb{k}$ by any family of vectors \( \{n_H \mid H \in \mathcal{A}\} \) where $n_H$ is normal to $H$. (Note that any normals will do, since the matroid is unchanged by scaling the $n_H$ 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 $B_r$ were a “genuine” 4-dimensional arrangement then we would not be able to represent it in $\mathbb{R}^3$. The reason we could do so is that the hyperplanes of $B_r$ intersected in a positive-dimensional space that we could effectively “squash down to nothing,” reducing the ambient dimension without changing the intersection poset. This observation motivates the following definition.

**Definition 6.1.7.** Let $\mathcal{A} \subseteq \mathbb{k}^n$ be an arrangement and let

\[
N(\mathcal{A}) = \mathbb{k}\langle n_H \mid H \in \mathcal{A} \rangle.
\]

The **essentialization** of $\mathcal{A}$ is the arrangement

\[
\text{ess}(\mathcal{A}) = \{ H \cap N(\mathcal{A}) \mid H \in \mathcal{A} \} \subseteq N(\mathcal{A}).
\]

If $N(\mathcal{A}) = \mathbb{k}^n$ then we say that $\mathcal{A}$ is **essential**. So $\text{ess}(\mathcal{A})$ is always essential, and $L(\text{ess}(\mathcal{A})) \cong L(\mathcal{A})$ as posets. The **rank** of $\mathcal{A}$ is the dimension of its essentialization.

We can now explain the correction factor in the characteristic polynomial: it is

\[
\chi_{\mathcal{A}}(k) = k^{\dim N(\mathcal{A})} \chi_{L(\mathcal{A})}(k) = k^{\dim \mathcal{A} - \text{rank} \mathcal{A}} \chi_{L(\mathcal{A})}(k).
\]

Thus the two polynomials coincide precisely for essential arrangements.

For example, rank $B_r = \dim \text{ess}(B_r) = n - 1$, and 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 $\text{ess}(\mathcal{A}) = \{ H/V \mid H \in \mathcal{A} \} \subseteq \mathbb{k}^n/V$. Thus $\mathcal{A}$ is essential if and only if $\cap \mathcal{A} = 0$. Moreover, if $\mathcal{A}$ is linear then $\text{rank}(\mathcal{A})$ is the rank of its intersection lattice — so rank is a combinatorial invariant, unlike dimension.

There are two natural operations that go back and forth between central and non-central arrangements, called **projectivization** and **coning**.

**Definition 6.1.8.** Let $\mathbb{k}$ be a field and $n \geq 1$. The set of lines through the origin in $\mathbb{k}^n$ is called **$n$-dimensional projective space** over $\mathbb{k}$ and denoted by $\mathbb{P}^{n-1}\mathbb{k}$.

If $\mathbb{k} = \mathbb{R}$, 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 $\mathbb{k}^n \setminus \{0\}$, and $\mathbb{P}^{n-1}$ is the set of equivalence classes.

Linear hyperplanes in $\mathbb{k}^n$ correspond to affine hyperplanes in $\mathbb{P}^{n-1}\mathbb{k}$. Thus, given a central arrangement $\mathcal{A} \subseteq \mathbb{k}^n$, we can construct its **projectivization** $\text{proj}(\mathcal{A}) \subseteq \mathbb{P}^{n-1}\mathbb{k}$.

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 $H \in \mathcal{A}$. Regard $H_0 \cap S$ as the equator and project the northern hemisphere into your piece of paper.
Of course, a diagram of \( \text{proj}(\mathcal{A}) \) only shows the “upper half” of \( \mathcal{A} \). We can recover \( \mathcal{A} \) from \( \text{proj}(\mathcal{A}) \) by “reflecting the interior of the disc to the exterior” (Stanley); e.g., for the Boolean arrangement \( \mathcal{A} = \mathcal{B}_3 \), the picture is as shown below. In general, \( r(\text{proj}(\mathcal{A})) = \frac{1}{2}r(\mathcal{A}) \).

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 6.1.9.** Let \( \mathcal{A} \subseteq \mathbb{k}^n \) be a hyperplane arrangement, not necessarily central. The **cone** \( c\mathcal{A} \) is the central arrangement in \( \mathbb{k}^{n+1} \) defined as follows:

- **Geometrically:** Make a copy of \( \mathcal{A} \) in \( \mathbb{k}^{n+1} \), 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 \( \mathbb{k}^{n+1} \)). Then, toss in one more hyperplane containing \( p \) and in general position with respect to every \( H' \).
- **Algebraically:** For \( H = \{ x \mid L(x) = a_i \} \subseteq \mathcal{A} \) (with \( L \) a homogeneous linear form on \( \mathbb{k}^n \) and \( a_i \in \mathbb{k} \)), construct a hyperplane \( H' = \{ (x_1, \ldots, x_n, y) \mid L(x) = a_i y \} \subseteq \mathbb{k}^{n+1} \) in \( c\mathcal{A} \). Then, toss in the hyperplane \( y = 0 \).

For example, if \( \mathcal{A} \) consists of the points \( x = 0, x = -5 \) and \( x = 3 \) 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 6.1.10. $\chi_{c\mathcal{A}}(k) = (k - 1)\chi_{\mathcal{A}}(k)$.

6.2 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 $\text{ess}(\mathcal{A})$ is bounded. (Note that if $\mathcal{A}$ is not essential then every region is unbounded.) Let

$$
\begin{align*}
\text{r}(\mathcal{A}) &= \text{number of regions of } \mathcal{A}, \\
\text{b}(\mathcal{A}) &= \text{number of relatively bounded regions of } \mathcal{A}.
\end{align*}
$$

Example 6.2.1. For the arrangements $\mathcal{A}_1$ and $\mathcal{A}_2$ shown in Example 6.1.3,

$$
\begin{align*}
\text{r}(\mathcal{A}_1) &= 6, \\
\text{b}(\mathcal{A}_1) &= 0, \\
\text{r}(\mathcal{A}_2) &= 10, \\
\text{b}(\mathcal{A}_2) &= 2.
\end{align*}
$$

Example 6.2.2. The Boolean arrangement $\mathcal{B}_n$ 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, $\chi_{\mathcal{B}_n}(k) = (k - 1)^n$. The complement $\mathbb{R}^n \setminus \mathcal{B}_n$ is $\{(x_1, \ldots, x_n) \mid x_i \neq 0 \text{ for all } i\}$, and the connected components are the open orthants, specified by the signs of the $n$ coordinates. Therefore, $\text{r}(\mathcal{B}_n) = 2^n$ and $\text{b}(\mathcal{B}_n) = 0$.

Example 6.2.3. 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. Each bounded region contains exactly one point where two lines of $\mathcal{A}$ meet, and each unbounded face has
four sides. Therefore

\[ v = r, \quad (6.2.1a) \]
\[ f = 1 + \left( \frac{m}{2} \right) = \frac{m^2 - m + 2}{2}, \quad (6.2.1b) \]
\[ 4(f - 1) = 2e - (r - b). \quad (6.2.1c) \]

Moreover, the number \( r - b \) of unbounded regions 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 (6.2.1c) and (6.2.1b) we obtain

\[ e = m + 2(f - 1) = m^2. \quad (6.2.1d) \]

Euler’s formula for planar graphs says that \( v - e + f = 2 \). Substituting in (6.2.1a), (6.2.1b) and (6.2.1d) and solving for \( r \) gives

\[ r = \frac{m^2 + m + 2}{2} \]

and therefore

\[ b = r - 2m = \frac{m^2 - 3m + 2}{2} = \left( m - 1 \right) \]

**Example 6.2.4.** The braid arrangement \( B_n \) consists of the \( \binom{n}{2} \) hyperplanes \( H_{ij} = \{ x \mid x_i = x_j \} \) in \( \mathbb{R}^n \). The complement \( \mathbb{R}^n \setminus B_n \) 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:

\[
\begin{align*}
  y = x \\
  y < x < z & \quad x < y < z \\
  y < z < z & \quad x < z < y \\
  z = y & \quad z < x < y \\
  z = x & \quad z < y < x \\
\end{align*}
\]

Therefore, \( r(B_n) = n! \). (Stanley: “Rarely is it so easy to compute the number of regions!”) Furthermore,

\[ \chi_{B_n}(k) = k(k - 1)(k - 2) \cdots (k - n + 1). \]

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 6.2.5.** Let \( G = (V, E) \) be a simple graph with \( V = [n] \). The graphic arrangement \( A_G \subseteq \mathbb{R}^n \) is the subarrangement of \( B_n \) containing the hyperplanes \( H_{ij} \). Thus \( L(A_G) \) is a geometric lattice, and the corresponding matroid is just the graphic matroid of \( G \), since that matroid is represented by the vectors \( \{ e_i - e_j \mid ij \in E \} \), which are normal to the hyperplanes \( H_{ij} \). In particular, the characteristic polynomial of
$L(\mathcal{A}_G)$ is precisely the chromatic polynomial of $G$ (see Section 5.4). We will see another explanation for this fact later; see Example 6.4.4.

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

$$r(\mathcal{A}_G) = \text{number of acyclic orientations of } G = |\chi_{L(\mathcal{A}_G)}(-1)|.$$

### 6.3 Zaslavsky’s theorem

Example 6.2.5 motivates the main result of this section, Theorem 6.3.6, which was historically the first major theorem about hyperplane arrangements, due to Zaslavsky [Zas75]. Let $\mathcal{A}$ be a real hyperplane arrangement, and let $\chi_{\mathcal{A}}$ be the characteristic polynomial of its intersection poset. Zaslavsky’s theorem(s) say(s) that the numbers of regions and relatively bounded regions are given by

$$r(\mathcal{A}) = (-1)^{\dim \mathcal{A}} \chi_{\mathcal{A}}(-1),$$
$$b(\mathcal{A}) = (-1)^{\text{rank } \mathcal{A}} \chi_{\mathcal{A}}(1).$$

The proof combines geometry and combinatorics.

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

$$\mathcal{A}_x = \{H \in \mathcal{A} \mid H \supseteq x\},$$
$$\mathcal{A}^x = \{H \cap x \mid H \in \mathcal{A} \setminus \mathcal{A}_x\}. \quad (6.3.1)$$

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,

$$L(\mathcal{A}_x) \cong \{y \in L(\mathcal{A}) \mid y \leq x\}, \quad L(\mathcal{A}^x) \cong \{y \in L(\mathcal{A}) \mid y \geq x\}.$$

**Example 6.3.1.** 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 posets (in this case, lattices) $L(\mathcal{A})$, $L(\mathcal{A}_p)$, and $L(\mathcal{A}^H)$ are shown below.
Henceforth, let $M(A)$ be the matroid represented by the normal vectors $\{n_H : H \in A\}$. Fix a hyperplane $H \in A$ and let $A' = A \setminus H$ and $A'' = A^H$.

**Proposition 6.3.2.** 1. $r(A) = r(A') + r(A'')$.

2. $b(A) = \begin{cases} 0 & \text{if rank } A = \text{rank } A' + 1 \quad (i.e., n_H \text{ is a coloop in } M(A)), \\ b(A') + b(A'') & \text{if rank } A = \text{rank } A' \quad (i.e., it isn't). \end{cases}$

**Proof.** (1) Consider what happens when we add $H$ to $A'$ to obtain $A$. Some regions of $A'$ 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 $A'$ (in the figure above, $S = 2$ and $U = 4$). The unsplit regions each contribute 1 to $r(A)$. The split regions each contribute 2 to $r(A)$, but they also correspond bijectively to the regions of $A''$. (See, e.g., Example 6.3.1.) So

$$r(A') = S + U, \quad r(A) = 2S + U, \quad r(A'') = S$$

and so $r(A) = r(A') + r(A'')$, proving the first assertion of Proposition 6.3.2. By the way, if (and only if) $H$ is a coloop then it borders every region of $A$, so $r(A) = 2r(A')$ in this case.

(2) Now we count bounded regions. If rank $A = \text{rank } A' + 1$, then $N(A') \subseteq \mathbb{R}^n$, i.e., $A'$ is not essential. In that case, every region of $A'$ must contain a line (or possibly a bigger space) orthogonal to $N(A')$, which gets squashed down to a point upon essentialization. Therefore, every region of $A$ contains a ray, and $b(A) = 0$. This takes care of the second case.

Otherwise, the bounded regions of $A$ come in a few different flavors.

- If $W$ is a bounded region not bordered by $H$, then it is also a bounded region of $A'$.
- If $X, Y$ are two bounded regions separated by part of $H$, then they merge into one bounded region $XY$ in $A'$. Also, the part of $H$ that separates them (namely $s = \partial X \cap H = \partial Y \cap H$) is itself a bounded region of $A''$.  

If $Z$ is a bounded region of $A$ whose neighbor across $H$ is an unbounded region $U$, then the merged region $ZU$ is unbounded in $A'$, but $\partial Z \cap H$ is nevertheless bounded in $A''$.

In short, here are the possibilities:

<table>
<thead>
<tr>
<th>Description</th>
<th>Contributions to...</th>
</tr>
</thead>
<tbody>
<tr>
<td>(W) bounded regions that don’t touch $H$</td>
<td>$b(A)$ $b(A')$ $b(A'')$</td>
</tr>
<tr>
<td>(X, Y) pairs of bounded regions separated by $H$</td>
<td>1 1 1</td>
</tr>
<tr>
<td>(Z) bounded, neighbor across $H$ is unbounded</td>
<td>1 0 1</td>
</tr>
</tbody>
</table>

In all cases the contribution to $b(A)$ equals the sum of those to $b(A')$ and $b(A'')$, establishing the second desired recurrence.

Proposition 6.3.2 looks a lot like a Tutte polynomial deletion/contraction recurrence. This suggests that we should be able to interpret $r(A)$ and $b(A)$ in terms of the characteristic polynomial $\chi_A$. The first step is to find a more convenient form for the characteristic polynomial.

**Proposition 6.3.3** (Whitney’s formula). *For any real hyperplane arrangement $A$,*

$$\chi_A(k) = \sum_{central \ B \subseteq A} (-1)^{|B|} k^{dim A} A - rank B.$$  

*Proof.* Consider the interval $[\hat{0}, x]$. The atoms in this interval are the hyperplanes of $A$ containing $x$, and they form a lower crosscut of $[\hat{0}, x]$. Using the formula for $\mu(\hat{0}, x)$ given by the second form of Thm. 3.5.9 gives

$$\chi_A(k) = \sum_{x \in L(A)} \mu(\hat{0}, x) k^{dim x}$$

$$= \sum_{x \in L(A)} \sum_{B \subseteq A: x = \cap B} (-1)^{|B|} k^{dim x}$$ (note that $1_{[\hat{0}, x]} = x$)

$$= \sum_{B \subseteq A: \cap B \neq 0} (-1)^{|B|} k^{dim (\cap B)}$$

$$= \sum_{central \ B \subseteq A} (-1)^{|B|} k^{dim A} A - rank B$$

as desired. Note that the empty subarrangement is considered central for the purpose of this formula, corresponding to the summand $x = \hat{0}$ and giving rise to the leading term $k^{dim A}$ of $\chi_A(k)$. 

93
Whitney’s formula leads to the following deletion/restriction recurrence for the characteristic polynomial.

**Proposition 6.3.4.** $\chi_A(k) = \chi_{A'}(k) - \chi_{A''}(k)$.

**Sketch of proof.** Splitting up Whitney’s formula gives

$$
\chi_A(k) = \sum_{\text{central } B \subseteq A: H \notin B} (-1)^{|B|} k^{n - \text{rank } B} + \sum_{\text{central } B \subseteq A: H \in B} (-1)^{|B|} k^{n - \text{rank } B}.
$$

Then $\Sigma_1 = \chi_{A'}(k)$ (it is just Whitney’s formula for $A'$), so it remains to show that $\Sigma_2 = -\chi_{A''}(k)$. This is a little trickier, because different hyperplanes in $A$ can have the same intersection with $H$, which means that subarrangements of $A$ can give rise to the same subarrangement of $A''$. Specifically, let $A'' = \{K_1, \ldots, K_s\}$, and split $A'$ into blocks $A_i = \{J \in A : J \cap H = K_i\}$. Each arrangement $B$ arising as a summand of $\Sigma_2$ gives rise to a central subarrangement of $A''$, namely

$$
\pi(B) = \{J \cap H : J \in B\},
$$

that depends only on which $A_i$’s it meets. That is, for each central $C \subseteq A''$, the set of summands $B$ of $\Sigma_2$ such that $\pi(B) = C$ is

$$
\pi^{-1}(C) = \{B \mid \pi(B) = C\} = \left\{B = H \cup B', \text{ where } B' = \bigcup_{i: K_i \in C} B_i \mid \emptyset \subseteq B_i \subseteq A_i\right\}.
$$

Also, since $H \in B$, we have $\cap B = \cap \pi(B)$ and therefore

$$
n - \text{rank } B = \dim(\cap B) = \dim(\cap \pi(B)) = \dim H - \text{rank } B'.
$$

(6.3.2)

Now we break up the sum $\Sigma_2$ into subsums depending on $\pi(B)$:

$$
\Sigma_2 = \sum_{\text{central } C \subseteq A''} \sum_{B \in \pi^{-1}(C)} (-1)^{|B|} k^{n - \text{rank } B}
$$

$$
= \sum_{B''} \left( \sum_{B \in \Psi(C)} (-1)^{|B|} \right) k^{\dim H - \text{rank } B''}
$$

$$
= \sum_{B''} \left( \sum_{B = H \cup B' \in \Psi(C)} -(-1)^{|B'|} \right) k^{\dim H - \text{rank } B''}
$$

$$
= -\sum_{B''} \left( \prod_{i} \sum_{K_i \in B' \neq B_i \subseteq A_i} (-1)^{|B_i|} \right) k^{\dim H - \text{rank } B''}
$$

(to see this, expand the product and observe that equals the inner sum in the previous line). But $\sum_{\emptyset \neq B_i \subseteq A_i} (-1)^{|B_i|} = -1$, because it is the binomial expansion of $(1 - 1)^{|C|} = 0$, with one +1 term (namely $B_i = \emptyset$) removed. Therefore, the whole thing boils down to

$$
-\sum_{B''} (-1)^{|B''|} k^{\dim H - \text{rank } B''}
$$

which is just Whitney’s formula for $-\chi_{A''}(k)$. □
Remark 6.3.5. This recurrence is strongly reminiscent of the chromatic recurrence (5.4.4). Indeed, if \( A = A_G \) is a graphic arrangement in \( \mathbb{R}^n \), \( e \) is an edge of \( G \), and \( H_e \) is the corresponding hyperplane in \( A_G \), then it is clear that \( A_G \setminus e = A_G \setminus \{H_e\} \). 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 \( A_G / e \) can be identified with \( (A_G)^{H_e} \) (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.

**Theorem 6.3.6** (Zaslavsky’s Theorem). Let \( A \) be a real hyperplane arrangement, and let \( \chi_A \) be the characteristic polynomial of its intersection poset. Then

\[
\begin{align*}
    r(A) &= (-1)^{\dim A} \chi_A(-1) \quad \text{and} \\
    b(A) &= (-1)^{\rank A} \chi_A(1).
\end{align*}
\]

**Proof.** We have already done the hard work, and just need to put all the pieces together.

If \( |A| = 1 \), then \( L(A) \) is the lattice with two elements, namely \( \mathbb{R}^n \) and a single hyperplane \( H \), and its characteristic polynomial is \( k^n - k^{n-1} \). Thus \( \tilde{r}(A) = (-1)^n((-1)^n - (-1)^{n-1}) = 2 \) and \( \tilde{b}(A) = -(1 - 1) = 0 \), which match \( r(A) \) and \( b(A) \).

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. 6.3.2). First,

\[
\begin{align*}
    \tilde{r}(A) &= (-1)^{\dim A} \chi_A(-1) \\
    &= (-1)^{\dim A} (\chi_A(-1) - \chi_{A'}(-1)) \quad \text{(by Prop. 6.3.4)} \\
    &= (-1)^{\dim A'} \chi_A'(-1) + (-1)^{\dim A''} \chi_A''(-1) \\
    &= \tilde{r}(A') + \tilde{r}(A'').
\end{align*}
\]

As for \( \tilde{b} \), if \( \rank A = \rank A' + 1 \), then in fact \( A' \) and \( A'' \) have the same essentialization, hence the same rank, and their characteristic polynomials only differ by a factor of \( k \). The deletion/restriction recurrence (Prop. 6.3.4) therefore implies \( \tilde{b}(A) = 0 \).

On the other hand, if \( \rank A = \rank A' \), then \( \rank A'' = \rank A - 1 \) and a calculation similar to that for \( \tilde{r} \) (replacing dimension with rank) shows that \( \tilde{b}(A) = \tilde{b}(A') + \tilde{b}(A'') \).

**Corollary 6.3.7.** Let \( A \subseteq \mathbb{R}^n \) be a central hyperplane arrangement and let \( M = M(A) \) be the matroid represented by normals. Then

\[
\begin{align*}
    r(A) &= T_M(2, 0), \\
    b(A) &= T_M(0, 0) = 0.
\end{align*}
\]

**Proof.** Combine Zaslavsky’s theorem with the formula \( \chi_A(k) = (-1)^n T_M(1 - k, 0) \).

**Remark 6.3.8.** The formula for \( r(A) \) could be obtained from the Tutte Recipe Theorem (Thm. 5.2.1). But this would not work for \( b(A) \), which is not an invariant of \( M(A) \).

**Example 6.3.9.** Let \( s \geq n \), and let \( A \) be an arrangement of \( s \) linear hyperplanes in general position in \( \mathbb{R}^n \). “General position” means precisely that the corresponding matroid \( M \) is \( U_n(s) \), whose rank function

95
$r : 2^s \to \mathbb{N}$ is given by $r(A) = \min(n, |A|)$. Therefore,

$$r(A) = T_M(2, 0) = \sum_{A \subseteq [s]} (2 - 1)^{n - r(A)}(0 - 1)^{|A| - r(A)}$$

$$= \sum_{A \subseteq [s]} (-1)^{|A| - r(A)}$$

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

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

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

$$= 2 \left( \binom{s}{n-1} + \binom{s}{n-3} + \binom{s}{n-5} + \cdots \right).$$

For instance, if $n = 3$ then

$$r(A) = 2 \binom{s}{2} + 2 \binom{s}{0} = s^2 - s + 2.$$

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 $r(A)$ and $b(A)$ for that arrangement is left as an exercise.

### 6.4 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 6.4.1.** Let $\mathbb{F}_q$ be the finite field of order $q$, and let $A \subseteq \mathbb{F}_q^n$ be a hyperplane arrangement. Then

$$|\mathbb{F}_q^n \setminus A| = \chi_A(q).$$

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.** By inclusion-exclusion,

$$|\mathbb{F}_q^n \setminus A| = \sum_{B \subseteq A} (-1)^{|B|} |\bigcap B|.$$

If $B$ is not central, then by definition $|\bigcap B| = 0$. Otherwise, $|\bigcap B| = q^{n - \text{rank} B}$. So the sum becomes

$$|\mathbb{F}_q^n \setminus A| = \sum_{\text{central } B \subseteq A} (-1)^{|B|} q^{n - \text{rank} B}$$

which is just Whitney’s formula for $\chi_A(q)$ (Prop. 6.3.3).
This fact has a much more general application, which was systematically mined by Athanasiadis, e.g., [Ath96].

**Definition 6.4.2.** Let $A \subseteq \mathbb{R}^n$ be an integral hyperplane arrangement (i.e., whose hyperplanes are defined by equations with integer coefficients). For a prime $p$, let $A_p = A \otimes \mathbb{F}_p$ be the arrangement in $\mathbb{F}_p^n$ defined by regarding the equations in $A$ as lying over $\mathbb{F}_p$. We say that $A$ reduces correctly modulo $p$ if $L(A_p) \cong L(A)$.

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, to choose $p$ larger than the absolute value of any minor of $M$, so that a 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:

**Corollary 6.4.3.** Let $A \subseteq \mathbb{R}^n$ be an integral hyperplane arrangement. Then $\chi_A(p)$ is the polynomial that counts points in the complement of $A_p$, for large enough primes $p$.

**Example 6.4.4.** Let $G = ([n], E)$ be a simple graph and let $A_G$ be the corresponding graphic arrangement in $\mathbb{R}^n$. Note that $A_G$ reduces correctly over every finite field $\mathbb{F}_q$ (because graphic matroids are regular). A point $(x_1, \ldots, x_n) \in \mathbb{F}_q^n$ 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 $\mathbb{F}_q^n \setminus A_G$. The number of such colorings is $p_G(k)$ (the chromatic polynomial of $G$ evaluated at $q$). On the other hand, by Theorem 6.4.1, it is also the characteristic polynomial $\chi_{A_G}(q)$. Since $p_G(k) = \chi_{A_G}(q)$ for infinitely many $q$ (namely, all integer prime powers), the polynomials must be equal.

**Example 6.4.5.** The Shi arrangement is the arrangement of $n(n-1)$ hyperplanes in $\mathbb{R}^n$ defined by

$$\text{Shi}_n = \{x_i = x_j, \ x_i = x_j + 1 \ | \ 1 \leq i < j \leq n\}.$$

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 $\text{ess}(\text{Shi}_2)$ consists of two points on a line, while $\text{ess}(\text{Shi}_3)$ is shown below.

**Proposition 6.4.6.** The characteristic polynomial of the Shi arrangement is

$$\chi_{\text{Shi}_n}(q) = q(q - n)^{n-1}.$$
In particular, the numbers of regions and bounded regions are respectively

\[ r(\text{Shi}_n) = |\chi(-1)| = (n+1)^{n-1}, \quad b(\text{Shi}_n) = |\chi(1)| = (n-1)^{n-1}. \]

The number \((n+1)^{n-1}\) should look familiar — by Cayley’s formula, it is the number of spanning trees of the complete graph \(K_{n+1}\).

**Proof.** It suffices to count the points in \(\mathbb{F}_q^n \setminus \text{Shi}_n\) for a large enough prime \(q\). Let \(x = (x_1, \ldots, x_n) \in \mathbb{F}_q^n \setminus \text{Shi}_n\). 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 \(k\) on the \(x_i\) bead. For example, let \(n = 6\) and \(q = 11\). Then the necklace for \(x = (2, 5, 6, 10, 3, 7)\) is as follows:

![Necklace Diagram](image)

The requirement that \(x\) 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 \(\mid\) for each bead without a red number, we get

\[ 15 \mid 236 \mid 4 \mid \]

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 \(1 \in B_1\).

Note that each block of \(\Pi(x)\) corresponds to a contiguous set of values among the coordinates of \(x\). 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 has its red numbers in strictly increasing order counterclockwise.

Thus we have bijections

\[ \mathbb{F}_q^n \setminus \text{Shi}_n \leftrightarrow \{\text{necklaces}\} \leftrightarrow \{\text{OWPs of } [n] \text{ with } q-n \text{ blocks}\}. \]

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 \(f : [2, n] \to [q-n]\), where \(f(i)\) is the index of the block containing \(i\) (note that \(f(1)\) must equal 1), and there are \((q-n)^{n-1}\) such things. Since there are \(q\) choices for the bead containing the red 1, we obtain

\[ |\mathbb{F}_q^n \setminus \text{Shi}_n| = q(q-n)^{n-1} = \chi_{\text{Shi}_n}(q). \]
6.5 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 $\mathcal{A}_G$. (This is another reason for the correction factor $k^{\dim \mathcal{A} - \rank \mathcal{A}}$ in (6.1.2).) For some graphs, the chromatic polynomial factors into linear terms. 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.

Example 6.5.1. 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 $r(F) = \min(|F|, 3)$. Since the Möbius function of an element of $L$ depends only on its rank, it is easy to check that

$$\chi_L(k) = k^3 - 4k^2 + 6k - 3 = (k-1)(k^2 - 3k + k).$$

Multiplying by $k^{\dim \mathcal{A}_L - \rank \mathcal{A}_L} = k^{4-3}$ gives the characteristic polynomial of $\mathcal{A}_L$, which is the chromatic polynomial of $C_4$:

$$\chi_{C_4}(k) = k(k-1)(k^2 - 3k + k).$$

For which graphs does the chromatic polynomial factor into linear terms? More generally, for which arrangements $\mathcal{A}$ does the characteristic polynomial $\chi_\mathcal{A}(k)$ factor? A useful sufficient condition is that the intersection poset be a supersolvable lattice.

Let $L$ be a lattice. Recall from (2.2.1) that $L$ is modular if it is ranked, and its rank function $r$ satisfies

$$r(x) + r(y) = r(x \lor y) + r(x \land y)$$

for every $x, y \in L$. This is not how we first defined modular lattices, but we proved that it is an equivalent condition (Theorem 2.3.5).

Definition 6.5.2. An element $x \in L$ is a modular element if $r(x) + r(y) = r(x \lor y) + r(x \land y)$ holds for every $y \in L$. Thus $L$ is modular if and only if every element of $L$ is modular.

- The elements $\hat{0}$ and $\hat{1}$ 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 $y = x \lor y$ and $x = x \land y$, while if $y \nleq x$ then $y \land x = \hat{0}$ and $y \lor x = \hat{1}$.
- The coatoms of a geometric lattice need not be modular. For example, let $L = \Pi_n$, and recall that $\Pi_n$ has rank function $r(\pi) = n - |\pi|$. Let $x = 12|34, y = 13|24 \in \Pi_4$. Then $r(x) = r(y) = 2$, but $r(x \lor y) = r(1) = 3$ and $r(x \land y) = r(\hat{0}) = 0$. So $x$ is not a modular element.

Proposition 6.5.3. The modular elements of $\Pi_n$ 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

$$X = \{C \in \sigma \mid C \cap B \neq \emptyset\}, \quad Y = \{C \in \sigma \mid C \cap B = \emptyset\}.$$
Then
\[ \pi \land \sigma = \left\{ C \cap B \mid C \in X \right\} \cup \left\{ \{i\} \mid i \not\in B \right\}, \quad \pi \lor \sigma = \left\{ \bigcup_{C \in X} C \right\} \cup Y \]

so
\[
|\pi \land \sigma| + |\pi \lor \sigma| = (|X| + n - |B|) + (1 + |Y|)
= (n - |B| + 1) + (|X| + |Y|) = |\pi| + |\sigma|, \]
proving that \( \pi \) is a modular element.

For the converse, suppose \( B, C \) are nonsingleton blocks of \( \pi \), with \( i, j \in B \) and \( k, \ell \in C \). Let \( \sigma \) be the partition with exactly two nonsingleton blocks \( \{i, k\}, \{j, \ell\} \). Then \( r(\sigma) = 2 \) and \( r(\pi \land \sigma) = r(\emptyset) = 0 \), but
\[
r(\pi \lor \sigma) = r(\pi) + 1 < r(\pi) + r(\sigma) - r(\pi \land \sigma)
\]
so \( \pi \) is not a modular element.

The usefulness of a modular element is that if one exists, we can factor the characteristic polynomial of \( L \).

**Theorem 6.5.4.** Let \( L \) be a geometric lattice of rank \( n \), and let \( z \in L \) be a modular element. Then
\[
\chi_L(k) = \chi_{\{0,z\}}(k) \sum_{y : y \land z = 0} \mu_L(\emptyset, y)k^{n-r(y)-r(\emptyset)}. \tag{6.5.1}
\]

Here is a sketch of the proof; for the full details, see [Sta07, pp. 440–441]. We work in the dual Möbius algebra \( A^\ast(L) = A(L^\ast) \); that is, the vector space of \( \mathbb{C} \)-linear combinations of elements of \( L \), with multiplication given by join (rather than meet as in §3.5). Thus the “algebraic” basis of \( A^\ast(L) \) is
\[
\{ \sigma_y \equiv \sum_{x : x \geq y} \mu(y, x) x \mid y \in L \}.
\]

First, show by direct calculation that
\[
\sigma_\emptyset = \sum_{x \in L} \mu(x)x = \left( \sum_{v : v \leq z} \mu(v)v \right) \left( \sum_{y : y \lor z = 0} \mu(y)y \right) \tag{6.5.2}
\]
for any \( z \in L \). Second, for \( z, y, v \in L \) such that \( z \) is modular, \( v \leq z \), and \( y \land z = 0 \), one shows first that \( z \land (v \lor y) = v \) (by rank considerations) and then that \( \text{rank}(v \lor y) = \text{rank}(v) + \text{rank}(y) \). Third, make the substitutions \( v \mapsto k^{\text{rank} z - \text{rank} v} \) and \( y \mapsto k^{\text{rank} y - \text{rank} z} \) in the two sums on the RHS of (6.5.2). Since \( v y = v \lor y \), the last observation implies that substituting \( x \mapsto k^{n - \text{rank} z} \) on the LHS preserves the product, and the equation becomes (6.5.1).

In particular, every atom \( a \) is modular, so
\[
\chi_L(k) = (k - 1) \sum_{x \in L} \mu_L(\emptyset, x)k^{r(L) - 1 - r(x)}.
\]

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 \( \chi_L(1) = \sum_{x \in L} \mu_L(\emptyset, x) = 0 \). Also, the sum in the expression is not the characteristic polynomial of a lattice.

On the other hand, if the modular element \( z \) is a *coatom*, then Theorem 6.5.4 is much more useful, since it lets us peel off a linear factor:
Corollary 6.5.5. Let $L$ be a geometric lattice, and let $z \in L$ be a coatom that is a modular element. Then

\[ \chi_L(k) = (k - e)\chi_{\hat{0},z}(k), \]

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

\[ \hat{0} = x_0 \ll x_1 \ll \cdots \ll x_{n-1} \ll x_n = \hat{1}. \]

In this case, we can apply Corollary 6.5.5 successively with $z = x_{n-1}$, $z = x_{n-2}$, \ldots, $z = x_1$ to split the characteristic polynomial completely into linear factors:

\[ \chi_L(k) = (k - e_{n-1})\chi_{\hat{0},x_{n-1}}(k) = (k - e_{n-1})(k - e_{n-2})\chi_{\hat{0},x_{n-2}}(k) = \cdots = (k - e_{n-1})(k - e_{n-2}) \cdots (k - e_0), \]

where

\[ e_i = \# \{ \text{atoms } a \text{ of } [\hat{0},x_{i+1}] \mid a \not\leq x_i \} = \# \{ a \in A \mid a \leq x_{i+1}, a \not\leq x_i \}. \]

Definition 6.5.6. A geometric lattice $L$ is supersolvable if it has a modular saturated chain, that is, a saturated chain $\hat{0} = x_0 \ll x_1 \ll \cdots \ll x_n = \hat{1}$ 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 6.5.7. Every modular lattice is supersolvable, because every saturated chain is modular. In particular, the characteristic polynomial of every modular lattice splits into linear factors.

Example 6.5.8. The partition lattice $\Pi_n$ (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 6.5.3. There are $n-1$ atoms $a \not\leq z$, namely the partitions whose nonsingleton 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 6.5.9. 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 6.5.10. 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

\[ C_i := \{ v_j \mid j \leq i, v_i v_j \in E \} \]

forms a clique in $G$. 

101
Such an ordering is called a **perfect elimination ordering**. The proof of Theorem 6.5.10 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 $\geq 4$, 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 6.5.10, 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

$$\chi(G; k) = \prod_{i=1}^{n} (k - |C_i|).$$

### 6.6 Arrangements over $\mathbb{C}$

What if $A \subseteq \mathbb{C}^n$ is a **complex** hyperplane arrangement? Since the hyperplanes of $A$ have codimension 2 as real vector subspaces, the complement $X = \mathbb{C}^n \setminus A$ 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 6.6.1** (Brieskorn [Bri73]). The homology groups $H_i(X, \mathbb{Z})$ are free abelian, and the Poincaré polynomial of $X$ is the characteristic polynomial backwards:

$$\sum_{i=0}^{n} \text{rank}_\mathbb{Z} H_i(X, \mathbb{Z})q^i = (-q)^n \chi_{L(A)}(-1/q).$$

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 $L(A)$, thereby proving that the cohomology is a combinatorial invariant of $A$. (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!

### 6.7 Exercises

**Exercise 6.1.** Let $m > n$, and let $A$ be the arrangement of $m$ affine hyperplanes in general position in $\mathbb{R}^n$ (so this is not the same thing as Example 6.3.9, where the hyperplanes are linear). Calculate $\chi_A(k)$, $r(A)$, and $b(A)$.

**Exercise 6.2.** (Stanley, HA, 2.5) Let $G$ be a graph on $n$ vertices, let $A_G$ be its graphic arrangement in $\mathbb{R}^n$, and let $B_G = \mathcal{B}_n \cup A_G$. (That is, $B$ 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 $\chi_{B_G}(q)$ in terms of $\chi_{A_G}(q)$.

**Exercise 6.3.** (Stanley, EC2, 3.115(d)) The **type B braid arrangement** $O_n \subseteq \mathbb{R}^n$ is the arrangement with two hyperplanes $x_i = x_j$, $x_i = -x_j$ for every pair $i, j$ with $1 \leq i < j \leq n$ (so $n(n-1)$ hyperplanes in all). Calculate the characteristic polynomial and the number of regions of $O_n$. 

102
Exercise 6.4. Recall that each permutation \( w = (w_1, \ldots, w_n) \in S_n \) 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 \( C^w \). For any set \( W \subseteq S_n \), consider the closed fan \( F(W) = \bigcup_{w \in W} C_w = \{(x_1, \ldots, x_n) \in \mathbb{R}^n | x_{w_1} \leq \cdots \leq x_{w_n} \text{ for some } w \in W\} \).

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 \( x \prec y \).)

Exercise 6.5. 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 \( i \)-th 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.

Exercise 6.6. Prove Theorem 6.5.10.
Chapter 7

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).

7.1 Basic definitions and terminology

Definition 7.1.1. Let $V$ be a finite set of vertices. An (abstract) simplicial complex $\Delta$ 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, $\Delta$ is an order ideal in the Boolean algebra $2^V$. The elements of $\Delta$ are called its faces or simplices. A face that is maximal with respect to inclusion is called a facet.

The dimension of a face $\sigma$ 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 $\Delta$ is $\dim \Delta = \max \{ \dim \sigma \mid \sigma \in \Delta \}$. A complex is pure if all its facets have the same dimension.

A fairly common convention is to set $\dim \Delta = d - 1$ so that $d$ is the maximum number of vertices in a face. (It is a constant source of irritation that the dimension of a simplex is one less than its cardinality.)

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 $\{v\} \not\in \Delta$, then $v \not\in \sigma$ 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 $e_1, \ldots, e_n \in \mathbb{R}^n$, and a simplex $\sigma = \{v_1, \ldots, v_r\}$ is then the convex hull of the
corresponding points:

$$|\sigma| = \text{conv}\{e_{v_1}, \ldots, e_{v_r}\} = \{c_1 e_{v_1} + \cdots + c_r e_r \mid 0 \leq c_i \leq 1 \, (\forall i), \ c_1 + \cdots + c_r = 1\}.$$ 

For example, faces of sizes 1, 2, and 3 correspond respectively to vertices, line segments, and triangles. Taking \(\{e_i\}\) to be the standard basis of \(\mathbb{R}^n\) gives the standard geometric realization \(|\Delta|\) of \(\Delta\):

$$|\Delta| = \bigcup_{\sigma \in \Delta} \text{conv}\{e_i : i \in \sigma\}.$$ 

It is usually possible to realize \(\Delta\) 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 \(\Delta\) and \(|\Delta|\); if we want to be specific we will use terminology like “geometric realization of \(\Delta\)” or “face poset of \(\Delta\).” A triangulation of a topological space \(X\) is a simplicial complex whose geometric realization is homeomorphic to \(X\).

Here are geometric realizations of the simplicial complexes \(\Delta_1 = (124, 23, 24, 34)\) and \(\Delta_2 = (12, 14, 23, 24, 34)\).

The filled-in triangle indicates that 124 is a face of \(\Delta_1\), but not of \(\Delta_2\). Note that \(\Delta_2\) is the subcomplex of \(\Delta_1\) consisting of all faces of dimensions \(\leq 1\) — that is, it is the 1-skeleton of \(\Delta_1\).

**Definition 7.1.2.** Let \(\Delta\) be a simplicial complex of dimension \(d-1\). The \(f\)-vector of \(\Delta\) is \((f_{d-1}, f_0, f_1, \ldots, f_{d-1})\), where \(f_i = f_i(\Delta)\) is the number of faces of dimension \(i\). The term \(f_{d-1}\) is often omitted, because \(f_{d-1} = 1\) unless \(\Delta\) is the void complex. The \(f\)-polynomial is the generating function for the nonnegative \(f\)-numbers (essentially the rank-generating function of \(\Delta\) 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 \(\Delta_1, \Delta_2\) are the simplicial complexes pictured above, then

$$f(\Delta_1) = (4, 5, 1) \quad \text{and} \quad f(\Delta_2) = (4, 5).$$

**Example 7.1.3.** 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.

What if \(P\) itself is the set of faces of a simplicial complex? In this case \(\Delta(P)\) is the barycentric subdivision. Combinatorially, the vertices of \(\text{Sd}(\Delta)\) correspond to the faces of \(\Delta\); a collection of vertices of \(\text{Sd}(\Delta)\) forms a face if the corresponding faces of \(\Delta\) are a chain in its face poset. Topologically, \(\text{Sd}(\Delta)\) can be constructed by drawing a vertex in the middle of each face of \(\Delta\) and connecting them — this is best illustrated by a picture.
Each vertex (black, red, blue) of Sd(Δ) 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.

### 7.2 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 $\mathbb{R}$, 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 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 $1 \leq v_0 < \cdots < v_k \leq n$, then

$$\partial_k[\sigma] = \sum_{i=0}^{k} (-1)^i [v_0, \ldots, \hat{v_i}, \ldots, v_k] \in C_{k-1}(\Delta)$$

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 $\partial_k$ by $\partial$, 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
\begin{bmatrix}
12 & 13 & 14 & 23 & 24
\end{bmatrix}
\partial_2

C_1 = \mathbb{R}^5
\begin{bmatrix}
12 & 13 & 14 & 23 & 24
\end{bmatrix}
\partial_1

C_0 = \mathbb{R}^4
\begin{bmatrix}
1 & 2 & 3 & 4
\end{bmatrix}
\partial_0

C_{-1} = \mathbb{R}
\begin{bmatrix}
1 & 2 & 3 & 4
\end{bmatrix}
```

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:

$$\partial^2 = 0.$$
An equivalent condition is that $\ker \partial_k \supseteq \text{im} \partial_{k+1}$ for all $k$. In particular, we can define the reduced simplicial homology groups\(^2\)

$$\tilde{H}_k(\Delta) = \tilde{H}_k(\Delta) = \ker \partial_k / \text{im} \partial_{k+1}.$$  

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 of $\Delta$ and traditionally notated $\beta_k$. They can be calculated using the rank-nullity formula: in general

$$\beta_k(\Delta) = \dim \tilde{H}_k(\Delta) = \dim \ker \partial_k - \dim \text{im} \partial_{k+1} = f_k - \text{rank} \partial_k - \text{rank} \partial_{k+1}.$$  

In the example above, this formula gives

$$\tilde{\beta}_0(\Delta) = 4 - 1 - 3 = 0, \quad \tilde{\beta}_1(\Delta) = 5 - 3 - 1 = 1, \quad \tilde{\beta}_2(\Delta) = 1 - 1 - 0 = 0$$

(note that $\partial_3$ 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 elaborate to give here,\(^4\) but provides a crucial tool for studying simplicial complexes: we can now ask how the topology of $\Delta$ 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 $\Delta$ is acyclic over every ring. If $\Delta \cong S^d$ (i.e., $|\Delta|$ is a $d$-dimensional sphere), then

$$\tilde{H}_k(\Delta) \cong \begin{cases} \mathbb{R} & \text{if } k = d, \\ 0 & \text{if } k < d. \end{cases} \tag{7.2.1}$$  

### 7.3 Stanley-Reisner theory

There is an extremely important connection, the **Stanley-Reisner correspondence**, between simplicial complexes and commutative algebra. Let $R = k[x_1, \ldots, x_n]$ be the ring of polynomials in $n$ variables over your favorite field $k$. Define the **support** of a monomial $\mu \in R$ as

$$\text{supp} \mu = \{i \mid x_i \text{ divides } \mu\}.$$

**Definition 7.3.1.** Let $\Delta$ be a simplicial complex on vertex set $[n]$. Its **Stanley-Reisner ideal** in $R$ is

$$I_\Delta = \langle x_{i_1} \cdots x_{i_r} \mid \{i_1, \ldots, i_r\} \not\in \Delta \rangle.$$  

The **Stanley-Reisner ring** or **face ring** is $k[\Delta] := R/I_\Delta$.

---

\(^2\)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.

\(^4\)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. See [Hat02, §2.1] for the full story. Or take my Math 821 class!
Example 7.3.2. Let $\Delta_1$ and $\Delta_2$ be the complexes on the previous page. Abbreviating $w, x, y, z = x_1, x_2, x_3, x_4$, the Stanley-Reisner ideal of $\Delta_1$ is

$$I_{\Delta_1} = \langle wxyz, wxy, wyz, xyz, wy \rangle = \langle xyz, wy \rangle.$$  

Note that the minimal generators of $I_\Delta$ are the minimal nonfaces of $\Delta$. Similarly,

$$I_{\Delta_2} = \langle wxz, xyz, wy \rangle.$$ 

If $\Delta$ is the simplex on $[n]$ then it has no nonfaces, so $I_\Delta$ is the zero ideal and $k[\Delta] = k[x_1, \ldots, x_n]$. In general, the more faces $\Delta$ has, the bigger its Stanley-Reisner ring is.

Since $\Delta$ is a simplicial complex, the monomials in $I_\Delta$ are exactly those whose support is not a face of $\Delta$. Therefore, the monomials supported on a face of $\Delta$ are a natural vector space basis for the graded ring $k[\Delta]$. Its Hilbert series can be calculated by counting these monomials:

$$\text{Hilb}(k[\Delta], q) \overset{\text{def}}{=} \sum_{i \geq 0} q^i \dim_k (k[\Delta])_i = \sum_{\sigma \in \Delta, \mu: \supp \mu = \sigma} q^{\abs{\mu}}$$

$$= \sum_{\sigma \in \Delta} \left( \frac{q}{1 - q} \right)^{\abs{\sigma}} = \sum_{i=0}^{d} f_{i-1} \left( \frac{q}{1 - q} \right)^i = \frac{\sum_{i=0}^{d} f_{i-1} q^i (1 - q)^{d-i}}{(1 - q)^d} = \sum_{i=0}^{d} h_i q^i.$$

The numerator of this rational expression is a polynomial in $q$, called the $h$-polynomial of $\Delta$, and its list of coefficients $(h_0, h_1, \ldots)$ is called the $h$-vector of $\Delta$. Clearing denominators and applying the binomial theorem yields a formula for the $h$-numbers in terms of the $f$-numbers:

$$\sum_{i=0}^{d} h_i q^i = \sum_{i=0}^{d} f_{i-1} q^i (1 - q)^{d-i} = \sum_{i=0}^{d} f_{i-1} q^i \sum_{j=0}^{d-i} \binom{d-i}{j} (-1)^j q^j$$

$$= \sum_{i=0}^{d} \sum_{j=0}^{d-i} \binom{d-i}{j} (-1)^j q^{i+j} f_{i-1}$$

and now extracting the $q^k$ coefficient (i.e., the summand in the second sum with $j = k - i$) yields

$$h_k = \sum_{i=0}^{k} \binom{d-i}{k-i} (-1)^{k-i} f_{i-1}. \quad (7.3.1)$$

(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.

$$f_{i-1} = \sum_{k=0}^{i} \binom{d-k}{i-k} h_k. \quad (7.3.2)$$

So the $f$-vector and $h$-vector contain equivalent information about a complex. On the level of generating functions, the conversions look like this [BH93, p. 213]:

$$\sum_i h_i q^i = \sum_i f_{i-1} q^i (1 - q)^{d-i}, \quad (7.3.3)$$

$$\sum_i f_i q^i = \sum_i h_i q^{i-1} (1 + q)^{d-i}. \quad (7.3.4)$$

108
The equalities (7.3.1) and (7.3.2) can be obtained by applying the binomial theorem to the right-hand sides of (7.3.3) and (7.3.4) 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: (7.3.1) 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, (7.3.2) gives

$$f_{d-1} = \sum_{k=0}^{d} h_k.$$

### 7.4 Shellable and Cohen-Macaulay simplicial complexes

For certain complexes, the $h$-numbers themselves have a direct combinatorial interpretation. The last formula suggests that they should enumerate facets of a pure complex in some way. Here is an important special class of complexes where they do.

**Definition 7.4.1.** A pure simplicial complex $\Delta$ of dimension $d - 1$ 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 $\langle F_i \rangle \setminus \langle F_1, \ldots, F_{i-1} \rangle$ has a unique minimal element, usually denoted $R_i$.
2. For every $i > 1$, the complex $\langle F_i \rangle \cap \langle F_1, \ldots, F_{i-1} \rangle$ is pure of dimension $d - 2$.

The proof of equivalence is left as an exercise.

For example, consider the **bipyramid**, which is a 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. This complex has many shelling orders, one of which is

$$234, \ 124, \ 134, \ 235, \ 125, \ 135.$$  

Here is a picture of the whole complex and the decomposition:

In the right-hand figure, the new edges created upon adding each triangle are indicated in bold. The corresponding decomposition of the face poset is

$$\emptyset, 234 \cup [1, 124] \cup [13, 134] \cup [5, 235] \cup [15, 125] \cup [135, 135]$$
which can be visualized as follows (each face is color-coded according to the interval $[R_i, F_i]$ that contains it):

![Diagram of a simplicial complex]

Here is 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 7.4.2.** Let $\Delta$ be shellable of dimension $d - 1$, with $h$-vector $(h_0, \ldots, h_d)$. Then

$$h_j = \# \{ F_i \mid \# R_i = j \} = \# \{ F_i \mid \langle F_i \rangle \cap \langle F_1, \ldots, F_{i-1} \rangle \text{ has } d - i \text{ faces of dimension } d - 2 \}.$$  

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 nonnegative, 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!

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. Define the **link** of a face $\sigma \in \Delta$ by

$$\text{lk}_{\Delta}(\sigma) = \{ \tau \in \Delta \mid \tau \cap \sigma = \emptyset, \tau \cup \sigma \in \Delta \}.$$  

The link of $\sigma$ is a subcomplex of $\Delta$, and can be thought of as “what you see if you stand at $\sigma$ and look outward.” For example, if $|\Delta|$ is a manifold of dimension $d$, then the link of any $k$-face is a simplicial $(d - k - 1)$-sphere, so in particular every link has zero homology in non-top dimension.

**Proposition 7.4.3** (Reisner’s theorem). A simplicial complex $\Delta$ is Cohen-Macaulay over $R$ iff (a) $\Delta$ is pure (so that $\dim \text{lk}_{\Delta}(\sigma) = \dim \Delta - \dim \sigma - 1$ for all $\sigma$) and (b) for every face $\sigma \in \Delta$, one has

$$\hat{H}_k(\text{lk}_{\Delta}(\sigma); R) = 0 \quad \forall k < \dim \Delta - \dim \sigma - 1.$$  

In practice, 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**
Figure 7.1: A step-by-step shelling of the octahedron with vertices a,b,c,d,e,f. Facets are labeled 1...8 in shelling order. Enumerating the sets $R_i$ by cardinality gives the $h$-vector $(1, 3, 3, 1)$.
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 (7.2.1) the wedge sum of $n$ copies of $S^d$ has reduced homology $R^n$ in dimension $d$, and 0 in other dimensions.)

A Cohen-Macaulay complex $\Delta$ is Gorenstein (over $R$) if in addition $\tilde{H}_{\dim \Delta - \dim \sigma - 1}(\text{lk}_\Delta(\sigma); R) \cong R$ for all $\sigma$. 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.)

MORE THINGS TO DEFINE/SKETCH: star, deletion, induced subcomplex, matroid complexes (this may be in the matroid section?), Laplacian

7.5 Exercises

Exercise 7.1. Let $\Delta$ be a simplicial complex on vertex set $V$, and let $v_0 \notin V$. The cone over $\Delta$ is the simplicial complex $C\Delta$ generated by all faces $\sigma \cup \{v_0\}$ 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 $\Delta$ is shellable if and only if $C\Delta$ is shellable.

Exercise 7.2. Let $\Delta$ 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 $R$.

Exercise 7.3. Construct two simplicial complexes with the same $f$-vector such that one is shellable and one isn’t.

Exercise 7.4. Prove that conditions (1) and (2) in the definition of shellability (Defn. 7.4.1) are equivalent.

Exercise 7.5. Prove Proposition 7.4.2.

Exercise 7.6. 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

$$\text{lk}_{X \cup Y}(\sigma) = \text{lk}_X(\sigma) \cup \text{lk}_Y(\sigma) \quad \text{and} \quad \text{lk}_{X \cap Y}(\sigma) = \text{lk}_X(\sigma) \cap \text{lk}_Y(\sigma).$$

Exercise 7.7. (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.)

Exercise 7.8. (Requires less experience with homological algebra than you might think.) The reduced Euler characteristic of a simplicial complex $\Delta$ is the alternating sum of its $f$-numbers:

$$\tilde{\chi}(\Delta) = \sum_{\sigma \in \Delta} (-1)^{\dim \sigma} = -f_{-1} + f_0 - f_1 + f_2 - \cdots = -1 + f(\Delta, -1).$$

(Topologists usually work with the unreduced Euler characteristic $\chi(\Delta) = f(\Delta, -1)$, which corresponds to ignoring the empty face.) Prove the Euler-Poincaré formula:

$$\tilde{\chi}(\Delta) = \sum_{k \geq -1} (-1)^k \dim_k \tilde{H}_k(\Delta; k).$$

(The choice of ground field $k$ is immaterial, but you can take it to be $\mathbb{R}$ if you want.)
Chapter 8

Polytopes

This section is under construction.

8.1 The basics

Polytopes are familiar objects: cubes, pyramids, Platonic solids, etc. Polytopes can be described either by their vertices or by their facets (maximal faces); the Fundamental Theorem of Polytopes says that the two descriptions are in fact equivalent.

One reference for this material is chapter 2 of Schrijver’s notes [Sch13].

First, a couple of 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 $\text{conv}(T)$. Explicitly, one can show (exercise; not too hard) that

$$\text{conv}(x_1, \ldots, x_r) = \left\{ c_1x_1 + \cdots + c_rx_r \mid 0 \leq c_i \leq 1 \text{ for all } i \text{ and } \sum_{i=1}^r c_i = 1 \right\}. \quad (8.1.1)$$

These points are called convex linear combinations of the $x_i$. A related definition is the affine hull of a point set:

$$\text{aff}(x_1, \ldots, x_r) = \left\{ c_1x_1 + \cdots + c_rx_r \mid \sum_{i=1}^r c_i = 1 \right\}. \quad (8.1.2)$$

This is the smallest affine space containing all $x_i$. (“Affine space” means “translate of a vector subspace of $\mathbb{R}^n$.”) The interior of $S$ as a subspace of its affine span is called the relative interior of $S$, denoted $\text{relint} S$.

Clearly $\text{conv}(T) \subseteq \text{aff}(T)$. For example, if $x_1, x_2, x_3$ are three non-collinear points in $\mathbb{R}^3$, then $\text{conv}(x_1, x_2, x_3)$ is the triangle having the $x_i$ as vertices, while $\text{aff}(x_1, x_2, x_3)$ is the unique plane containing all three points.

Definition 8.1.1. A polyhedron $P$ is a nonempty intersection of finitely many closed half-spaces in $\mathbb{R}^n$. Equivalently,

$$P = \{ x \in \mathbb{R}^n \mid a_{i1}x_1 + \cdots + a_{in}x_n \geq b_i \text{ } \forall i \in [m] \}$$

where $a_{ij}, b_i \in \mathbb{R}$. These equations are often written as a single matrix equation $Ax \geq b$, where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^n$. 

**Definition 8.1.2.** A **polytope** in $\mathbb{R}^n$ is defined as either (i) a bounded polyhedron; or (ii) the convex hull of a finite set of points.

**Theorem 8.1.3** (The Fundamental Theorem of Polytopes). The two definitions of “polytope” in Definition 8.1.2 are equivalent.

We have some work to do before getting to the proof of Theorem 8.1.3.

**Definition 8.1.4.** A point $x$ in a polyhedron $P$ is a **vertex** of $P$ if $x \notin \text{conv}(P \setminus \{x\})$.

**Definition 8.1.5.** 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 $x \in F$, $y \in P$. The face is **proper** if $\ell$ 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 $\partial P$ (proof left as an exercise).

As a concrete example, 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 nonzero vector $v$, the extreme points of $P$ in that direction are by definition the maxima of the linear functional $x \mapsto x \cdot v$, and the set of those points is a face.

If you pick a linear functional “at random”, then with probability 1, the face it determines will be a vertex of $P$. Higher-dimensional faces correspond to more special directions.

**Proposition 8.1.6.** 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. For each $x \in P$, there is a unique minimal face $F_x \subseteq P$ containing $x$.
3. $x \in \text{relint } F_x$ for every $x \in P$.
4. The vertices of a polytope are exactly its 0-dimensional faces.

**Proof.** (1) Each face is defined by adding a single linear inequality to the list of inequalities defining $P$.

(2) We will show that if $F', F''$ are two faces containing $x$, then $F' \cap F''$ is also a face. Let $\ell', \ell''$ be the linear functionals maximized on $F', F''$ respectively. Let $F$ be the face corresponding to the linear function $\ell = \ell' + \ell''$ (or indeed any positive linear combination of them). Then $x$ is a global maximum of $F$ on $P$, and since $x$ also maximizes both $F'$ and $F''$, it follows that $F = F' \cap F''$, as desired. We can therefore define $F_x$ to be the intersection of all faces containing $x$.

(3) If $x \in \partial F_x$ then $F_x$ has a face $G$ containing $x$, but $G$ is also a face of $P$ by (1), which contradicts the definition of $F_x$.

(4) Suppose that $x$ is a 0-dimensional face, i.e., the unique maximum of some linear functional $\ell$. If $x$ is a convex linear combination $\sum c_i y_i$ of points $y_i \in P$, then $\ell(x) \geq \sum c_i \ell(y_i)$, with equality only if $\ell(y_i) = \ell(x)$ for all $i$, but then $y_i = x$ for all $i$. Therefore $x$ is not in the convex hull of $P \setminus \{x\}$, hence is a vertex.

OTOH, if $x$ is a point of $P$ that is not an 0-dimensional face, then by (3) $x \in \text{relint } F_x$, hence $x$ is a convex combination of two other points (such as $x \pm \epsilon v$ for any vector $v$ and some sufficiently small $\epsilon$).

**Sketch of proof of Theorem 8.1.3.** Some details are left as an exercise. A full proof appears in [Sch13, §2.2].
First, let $P$ be a intersection of finitely many half-spaces, i.e., $P = \{x \in \mathbb{R}^n \mid Ax \leq b\}$, where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. By projecting onto the orthogonal complement of the rowspace of $A$, we can assume WLOG that $\operatorname{rank} A = n$. For each point $z \in P$, let $A_z$ be the submatrix of $A$ consisting of rows $a_i$ for which $a_i \cdot z = b_i$. **One must show** that

$$z \in P \iff \operatorname{rank} A_z = n. \quad (8.1.3)$$

(More generally, $\operatorname{rank} A_z = n - \dim F_z$.)

It follows that the vertices are all of the form $A^{-1}b_R$, 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 there are only finitely many vertices $v_1, \ldots, v_k$ (specifically, $k \leq \binom{n}{m}$). So far, this argument applies to any polyhedron. In the next step, **one must show** that

$$P = \text{conv}\{v_1, \ldots, v_k\} \quad (8.1.4)$$

using in addition the assumption that $P$ is bounded.

Second, let $P = \text{conv}(z_1, \ldots, z_r) \subseteq \mathbb{R}^n$. Assume without loss of generality that $\text{aff}(P) = \mathbb{R}^n$ (otherwise, replace $\mathbb{R}^n$ with the affine hull) and that the origin is in the interior of $P$ (translating if necessary). Define

$$P^* := \{y \in \mathbb{R}^n \mid x \cdot y \leq 1 \ \forall x \in P\}. \quad (8.1.5)$$

This is called the (polar) dual of $P$. **One must show** that in fact

$$P^* = \{y \in \mathbb{R}^n \mid z_i \cdot y \leq 1 \ \forall i \in [r]\} \quad (8.1.6)$$

which means that $P^*$ is an intersection of finitely many half-spaces. So, by the first part of the theorem, $P^* = \text{conv}(y_1, \ldots, y_s)$ for some $y_1, \ldots, y_s$. Meanwhile, the double dual $P^{**} = (P^*)^*$ is defined by

$$P^{**} = \{x \in \mathbb{R}^n \mid x \cdot y \leq 1 \ \forall y \in P^*\} = \{x \in \mathbb{R}^n \mid x \cdot y_j \leq 1 \ \forall j \in [s]\} \quad (8.1.7)$$

where the second equality comes from applying the previous observation. Now **one must show** that

$$P = P^{**} \quad (8.1.8)$$

so that (8.1.7) expresses $P$ as the intersection of finitely many half-spaces.

**Example 8.1.7.** Let $P$ be the polytope shown on the left of Figure 8.1. It can be expressed as an intersection of hyperplanes (i.e., solution set to a system of linear inequalities) as follows:

$$P = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 : \begin{bmatrix} -1 & 0 \\ 0 & -1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \leq \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right\}.$$

If we number the rows $R_0, \ldots, R_4$, every pair of rows other than $\{R_0, R_2\}$ and $\{R_1, R_3\}$ is of full rank. The points corresponding to the other eight pairs of rows are:

<table>
<thead>
<tr>
<th>Rows</th>
<th>Point</th>
<th>Type</th>
<th>Constraint on $P^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0, 1</td>
<td>$(-1, -1)$</td>
<td>vertex</td>
<td>$-x - y \leq 1$</td>
</tr>
<tr>
<td>0, 3</td>
<td>$(-1, 1)$</td>
<td>vertex</td>
<td>$-x + y \leq 1$</td>
</tr>
<tr>
<td>0, 4</td>
<td>$(-1, 2)$</td>
<td>not in $P$</td>
<td></td>
</tr>
<tr>
<td>1, 2</td>
<td>$(1, -1)$</td>
<td>vertex</td>
<td>$x - y \leq 1$</td>
</tr>
<tr>
<td>1, 4</td>
<td>$(2, -1)$</td>
<td>not in $P$</td>
<td></td>
</tr>
<tr>
<td>2, 3</td>
<td>$(1, 1)$</td>
<td>not in $P$</td>
<td></td>
</tr>
<tr>
<td>2, 4</td>
<td>$(1, 0)$</td>
<td>vertex</td>
<td>$x \leq 1$</td>
</tr>
<tr>
<td>3, 4</td>
<td>$(0, 1)$</td>
<td>vertex</td>
<td>$y \leq 1$</td>
</tr>
</tbody>
</table>

Thus the vertices of $P$ correspond to the bounding hyperplanes (i.e., lines) of $P^*$, and vice versa.
Definition 8.1.8. 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.

A polytope 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.

A polytope is simple if every vertex belongs to exactly $n$ faces. (In fact no vertex can belong to fewer than $n$ faces.)

The face poset of $P$ is the poset of all faces, ordered by inclusion. This poset is ranked by dimension; one can decide whether to include the empty face as $\hat{0}$ and all of $P$ as $\hat{1}$. For instance, the face poset of a simplex is a Boolean algebra. Two polytopes are combinatorially isomorphic if their face posets are isomorphic.

Proposition 8.1.9. A polytope $P$ is simple if and only if its dual $P^*$ is simplicial. In this case the face poset of $P$ is the dual of the face poset of $P^*$.

8.2 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 in this area.

Theorem 8.2.1. Let $\Delta$ be the boundary sphere of a convex simplicial polytope $P \subseteq \mathbb{R}^d$. Then $\Delta$ is shellable, and its $h$-vector is a palindrome, i.e., $h_i = h_{d-i}$ for all $i$.

Palindromicity of the $h$-vector is known as 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 $\mathcal{H}$ be the collection of hyperplanes spanned by facets of $P$. Let $\ell$ be a line that passes through the interior of $P$ and meets each hyperplane in $\mathcal{H}$ in a distinct point. (Note that almost any line will do.) Imagine walking along this line, starting just outside $P$ so that only one facet is visible. Call that facet $F_1$. As you continue to walk, more and more facets become visible. Label the facets $F_2,\ldots, F_m$ in the order in which they appear (equivalently, order them in the order in which the line $\ell$ meets their affine spans). When you get to infinity, come back the other way (so that all of a sudden “invisible” and “visible” switch meanings) and continue to label the facets $F_{m+1},\ldots, F_n$ in the order in which they disappear.

In fact $F_1,\ldots, F_n$ is a shelling order (called a line shelling), because

$$\langle F_j \rangle \cap \langle F_1,\ldots, F_{j-1} \rangle = \begin{cases} \{ \text{ridges of } F_j \text{ that appear upon crossing } \text{aff}(F_j) \} & \text{for } 1 \leq j \leq m, \\ \{ \text{ridges of } F_j \text{ that disappear upon crossing } \text{aff}(F_j) \} & \text{for } m+1 \leq j \leq n. \end{cases}$$

(This assertion does need to be checked.) Moreover, each facet $F_i$ contributes to $h_k(P)$, where

$$k = k(F_i) = \# \{ j < i \mid F_i, F_j \text{ have a common ridge} \}.$$  

On the other hand, the reversal of $<$ is another instance of this construction, hence is also a shelling order. Since each facet shares a ridge with exactly $n$ other facets (because $P$ is simplicial!), the previous formula says that if facet $F$ contributes to $h_i$ with respect to the first shelling order then it contributes to $h_{n-i}$ in its reversal. Since the $h$-vector is an invariant of $P$, it follows that $h_i = h_{n-i}$ 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).

### 8.3 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?**

**Definition 8.3.1.** 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 $\partial P$, 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 $\varepsilon$ (its intersection with $\text{aff}(P)$) contains both points of $P$ and points not in $P$. The **relative interior** of $P$, int $P$, is $P \setminus \partial P$.

For a polytope $P \subseteq \mathbb{R}^N$ define sequences

$$i(P, n) = |nP \cap \mathbb{Z}^N|$$

$$i^*(P, n) = |n(\text{int } P) \cap \mathbb{Z}^N|$$

$i(P, n)$ 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

$$P = \text{conv}\{(0, 0, 0), (1, 1, 0), (1, 0, 1), (0, 1, 1)\} \in \mathbb{R}^3.$$ 

Then

$$nP = \text{conv}\{(0, 0, 0), (n, n, 0), (n, 0, n), (0, n, n)\}.$$ 

Each point in $nP$ can be written as $\beta(n, n, 0) + \beta_2(n, 0, n) + \beta_3(0, n, n) + \beta_4(0, 0, 0)$, 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 $\alpha_i$ 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 $\alpha_i$ to be 1/2. If any one of the $\alpha_i$ has fractional part 1/2, the others must be also. Writing $\gamma_i = \alpha_i - 1/2$, we get points of the form

$$(\gamma_1 + 1/2)(1, 1, 0) + (\gamma_2 + 1/2)(1, 0, 1) + (\gamma_3 + 1/2)(0, 1, 1)$$

$$= \gamma_1(1, 1, 0) + \gamma_2(1, 0, 1) + \gamma_3(0, 1, 1) + (1, 1, 1).$$

Note here that $\sum \gamma_i = (\sum \alpha_i) - 3/2 \leq n - 3/2$. Since the $\gamma_i$ 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 degree $n - 4$; 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 $\bar{P} = \{(x, 1) \in \mathbb{R}^{N+1} : x \in P\}$, and let $C$ be the simplicial cone generated by $\bar{P}$:

$$C = C(\bar{P}) = \{ry : y \in \bar{P}, r \in \mathbb{R}, r \geq 0\}.$$
The boundary and interior of $C$ are $\partial C = \{ ry : y \in \partial \hat{P} \}$ and $\text{int} C = C \setminus \partial C$. 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

$$i(P, n) = |C \cap \{(z, n) \in \mathbb{R}^{N+1} : z \in \mathbb{Z}^N\}|$$

$$i^*(P, n) = |\text{int} C \cap \{(z, n) \in \mathbb{R}^{N+1} : z \in \mathbb{Z}^N\}|.$$

We can represent all points in the cone in terms of the vertices of $P$.

**Proposition 8.3.2.** Let $P$ be a rational $N$-simplex in $\mathbb{R}^N$, with vertices $v_0, v_1, \ldots, v_N$, and let $C = C(\hat{P})$. 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

$$Q = \left\{ \sum_{i=0}^{N} r_i(v_i, 1) \mid 0 \leq r_i < 1 \ \forall i \right\}.$$

Thus $Q$ is a half-open parallelepiped containing 0 and $\hat{P}$.

**Proposition 8.3.3.** Let $P$ be an integral $N$-simplex in $\mathbb{R}^N$, with vertices $v_0, v_1, \ldots, v_N$, and let $C = C(\hat{P})$. A point $z \in \mathbb{Z}^{N+1}$ is an integer point in $C$ if and only if $z = y + \sum_{i=0}^{N} r_i(v_i, 1)$ 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 $r_i$, $0 \leq r_i < 1$, so if $\deg z$ is an integer, $0 \leq \deg z \leq N$.

**Theorem 8.3.4.** Let $P$ be an integral $N$-simplex in $\mathbb{R}^N$, with vertices $v_0, v_1, \ldots, v_N$, let $C = C(\hat{P})$, and let $Q = \{ \sum_{i=0}^{N} r_i(v_i, 1) : \text{for each } i, 0 \leq r_i < 1 \}$. Let $\delta_j$ 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 8.3.5.** For $P$ an integral $N$-simplex, $i(P, n)$ is a polynomial in $n$.

**Proof.**

$$\sum_{n=0}^{\infty} i(P, n) \lambda^n = (\delta_0 + \delta_1 \lambda + \cdots + \delta_N \lambda^N)(1 + \lambda + \lambda^2 + \cdots)^{N+1}$$

$$= (\delta_0 + \delta_1 \lambda + \cdots + \delta_N \lambda^N) \left( \sum_{k=0}^{\infty} \binom{k+N}{N} \lambda^k \right).$$

The coefficient of $\lambda^n$ on the right hand side is $\sum_{j=0}^{N} \delta_j \binom{n-j+N}{N}$. 

119
For the interior of $P$ (and of $C$) we use an analogous construction, but with the opposite half-open parallelepiped. Let
\[ Q^* = \left\{ \sum_{i=0}^{N} r_i(v_i, 1) \mid 0 < r_i \leq 1 \forall i \right\}. \]

**Proposition 8.3.6.** Let $P$ be an integral $N$-simplex in $\mathbb{R}^N$, with vertices $v_0, v_1, \ldots, v_N$, and let $C = C(\tilde{P})$. A point $z \in \mathbb{Z}^{N+1}$ is an integer point in $\text{int} C$ if and only if $z = y + \sum_{i=0}^{N} c_i(v_i, 1)$ 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 $\text{int} C$ (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 $z \in Q^*$, $0 < \deg z \leq N + 1$.

**Theorem 8.3.7.** Let $P$ be an integral $N$-simplex in $\mathbb{R}^N$, with vertices $v_0, v_1, \ldots, v_N$, let $C = C(\tilde{P})$, and let $Q^* = \{\sum_{i=0}^{N} r_i(v_i, 1) : \text{for each } i, 0 < r_i \leq 1\}$. Let $\delta_j^*$ be the number of points of degree $j$ in $Q^* \cap \mathbb{Z}^{N+1}$. Then
\[ \sum_{n=0}^{\infty} i^*(P, n) \lambda^n = \frac{\delta_0^* + \delta_1^* \lambda^2 + \cdots + \delta_{N+1}^* \lambda^{N+1}}{(1 - \lambda)^{N+1}}. \]

**Corollary 8.3.8.** 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 $\delta_i$ and the $\delta_i^*$. Note that
\[ Q^* = \left\{ \sum_{i=0}^{N} r_i(v_i, 1) : \text{for each } i, 0 < r_i \leq 1\right\} = \left\{ \sum_{i=0}^{N} (1 - t_i)(v_i, 1) : \text{for each } i, 0 \leq t_i < 1\right\} = \left\{ \sum_{i=0}^{N} (v_i, 1) - \sum_{i=0}^{N} t_i(v_i, 1) : \text{for each } i, 0 \leq t_i < 1\right\} = \sum_{i=0}^{N} (v_i, 1) - Q = \left( \sum_{i=0}^{N} v_i, N + 1 \right) - Q \]

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 8.3.9.** 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**.

120
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 8.3.10.** 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 \( \lambda \) 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 \( \delta_j \)?

\[
\delta_0 = i(P,0) = 1, \text{ since this is the number of integer points in the polytope } 0P = \{0\}.
\]

\[
\delta_1 + (N+1)\delta_0 = i(P,1), \text{ so } \delta_1 = |P \cap \mathbb{Z}^N| - (N+1).
\]

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 \( \text{vol}(nP) = n^N \text{vol}(P) \) (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,

\[
0 = \lim_{n \to \infty} \frac{i(P,n) - \text{vol}(nP)}{n^N} = \lim_{n \to \infty} \frac{i(P,n)}{n^N} - \text{vol}(P).
\]

So \( C = \lim_{n \to \infty} \frac{i(P,n)}{n^N} = \text{vol}(P) \).

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.,

\[
i(P,n) = \begin{cases} (n+1)^2 & \text{if } n \text{ even} \\ n^2 & \text{if } n \text{ odd} \end{cases} = (n + \gcd(n,2) - 1)^2.
\]

### 8.4 Exercises

**Exercise 8.1.** Prove that the topological boundary of a polyhedron is the union of its proper faces.

**Exercise 8.2.** Prove that the convex hull of a finite point set is the set of convex linear combinations of it.

**Exercise 8.3.** Fill in the details in the proof of Theorem 8.1.3 by proving all the assertions of the form “One must show”, i.e., (8.1.3), (8.1.4), (8.1.6), and (8.1.8).
If necessary, you may use Minkowski's Hyperplane Separation Theorem, which states that if $S \subseteq \mathbb{R}^n$ is a convex set and $y \not\in S$, then there exists a hyperplane separating $S$ from $y$ — or equivalently a linear functional $\ell : \mathbb{R}^n \to \mathbb{R}$ such that $\ell(y) > 0$ and $\ell(x) < 0$ for all $x \in S$. 
Chapter 9

Group Representations

9.1 Basic definitions

Definition 9.1.1. Let $G$ be a group (typically finite) and let $V \cong k^n$ be a finite-dimensional vector space over a field $k$. A representation of $G$ on $V$ is a group homomorphism $\rho : G \to GL(V)$, where $GL(V) = GL_n(k)$ is the group of linear automorphisms of $V$, or equivalently the group of invertible $n \times n$ matrices over $k$. That is, for each $g \in G$ there is an invertible $n \times n$ matrix $\rho(g)$, satisfying

$$\rho(g)\rho(h) = \rho(gh) \quad \forall g, h \in G.$$

(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 or degree of the representation.

Some remarks:

- $\rho$ 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 $gv$ instead of the bulkier $\rho(g)v$.
- It is common to say that $\rho$ is a representation, or that $V$ is a representation, or that the pair $(\rho, V)$ is a representation.
- $\rho$ is faithful if it is injective as a group homomorphism.

Example 9.1.2. Let $G$ be any group. The trivial representation is the map

$$\rho_{\text{triv}} : G \to GL_1(k) \cong k^\times$$

sending $g \mapsto 1$ for all $g \in G$.

Example 9.1.3. Let $kG$ be the vector space of formal $k$-linear combinations of elements of $G$: that is, $kG = \sum_{h \in G} a_h h | a_h \in k$. The regular representation of $G$ is the map $\rho_{\text{reg}} : G \to GL(kG)$ 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 $kG$ according to the group multiplication law. The regular representation has dimension $|G|$.
The vector space \( kG \) is a ring, with multiplication given by multiplication in \( G \) and extended \( k \)-linearly. In this context it is called the **group algebra** of \( G \) over \( k \).

**Remark 9.1.4.** A representation of \( G \) is equivalent to a (left) **module** over the group algebra \( kG \). Technically “representation” refers to the way \( G \) acts and “module” refers to the way in which it acts, but the two terms really carry the same information.

**Example 9.1.5.** Let \( G = \mathfrak{S}_n \), the symmetric group on \( n \) elements. The **defining representation** \( \rho_{\text{def}} \) of \( G \) on \( k^n \) 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.

**Example 9.1.6.** More generally, let \( G \) act on a finite set \( X \). Then there is an associated **permutation representation** on \( k^X \), the vector space with basis \( X \), given by

\[
g \left( \sum_{x \in X} a_x x \right) = \sum_{x \in X} a_x (g \cdot x).
\]

For short, we might specify the action of \( G \) and \( X \) and say that it “extends linearly” to \( k^X \). 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 9.1.7.** Let \( G = \mathbb{Z}/k\mathbb{Z} \) be the cyclic group of order \( k \), and let \( \zeta \) be a \( k^{th} \) root of unity (not necessarily primitive). Then \( G \) has a 1-dimensional representation given by \( \rho(x) = \zeta^x \).

**Example 9.1.8.** 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 \( D_n \).

1. Regarding \( s \) as a reflection and \( r \) as a rotation in \( \mathbb{R}^2 \) gives a faithful 2-dimensional representation (the “**geometric representation**”)

\[
\rho_{\text{geo}}(s) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad \rho_{\text{geo}}(r) = \begin{bmatrix} \cos(2\pi/n) & \sin(2\pi/n) \\ -\sin(2\pi/n) & \cos(2\pi/n) \end{bmatrix}.
\]

2. The permutation representations of \( D_n \) on vertices or on edges are faithful \( n \)-dimensional representations.

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 \( r^{n/2} \) acts by rotation by \( 180^\circ \) and fixes all diameters.

**Example 9.1.9.** The symmetric group \( \mathfrak{S}_n \) has a nontrivial 1-dimensional representation, the **sign representation**, given by

\[
\rho_{\text{sign}}(\sigma) = \begin{cases} 1 & \text{if } \sigma \text{ is even}, \\ -1 & \text{if } \sigma \text{ is odd}. \end{cases}
\]

Note that \( \rho_{\text{sign}}(g) = \det \rho_{\text{def}}(g) \), where \( \rho_{\text{def}} \) is the defining representation of \( \mathfrak{S}_n \). In general, if \( \rho \) is any representation, then \( \det \rho \) is a 1-dimensional representation.

**Example 9.1.10.** Let \( (\rho, V) \) and \( (\rho', V') \) be representations of \( G \), where \( V \cong k^n \), \( V' \cong k^m \). 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:

\[
\begin{bmatrix}
\rho(g) & 0 \\
0 & \rho'(g)
\end{bmatrix}.
\]
9.2 Isomorphisms and homomorphisms

When are two representations the same? More generally, what is a map between representations?

**Definition 9.2.1.** Let \((\rho, V)\) and \((\rho', V')\) be representations of \(G\). A linear transformation \(\phi : V \to V'\) is **\(G\)-equivariant**, or a **homomorphism**, if \(\rho'(g) \cdot \phi(v) = \phi(\rho(g) \cdot v)\) for all \(g \in G\) and \(v \in V\). More concisely, \(g\pi = \phi g\) for all \(g \in G\), i.e., the following diagram commutes

\[
\begin{array}{ccc}
V & \xrightarrow{\phi} & V' \\
\rho(g) \downarrow & & \downarrow \rho'(g) \\
V & \xrightarrow{\phi} & V'
\end{array}
\] (9.2.1)

We sometimes use the notation \(\phi : \rho \to \rho'\). An **isomorphism** of representations is a \(G\)-equivariant map that is a vector space isomorphism. In the language of modules, a \(G\)-equivariant transformation is the same thing as a \(kG\)-module homomorphism.

**Example 9.2.2.** It is easy to see that the identity map \(V \to V\) is an automorphism for any group action.

**Example 9.2.3.** Let \(G = S_n\) act on \(\mathbb{R}^n\) by the defining representation, and on \(\mathbb{R}\) by the trivial representation. Consider the map \(\mathbb{R}^n \to \mathbb{R}\) given by

\[
\phi \left( \sum_{i=1}^n a_i e_i \right) = \sum_{i=1}^n a_i.
\]

This map is \(G\)-equivariant because permuting the coordinates of a vector does not change their sum.

**Example 9.2.4.** Let \(n\) be odd, and consider the dihedral group \(D_n\) acting on a regular \(n\)-gon. Label the vertices \(1, \ldots, n\) in cyclic order. Label each edge the same as its opposite vertex, as in the figure on the left. Then the permutation action \(\rho\) on vertices is identical to the action \(\rho'\) 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 corresponding basis vector for an edge.

**Example 9.2.5.** Let \(v_1, \ldots, v_n\) be the points of a regular \(n\)-gon in \(\mathbb{R}^2\) centered at the origin, e.g., \(v_j = (\cos \frac{2\pi j}{n}), \sin \frac{2\pi j}{n})\). Then the map \(\mathbb{R}^n \to \mathbb{R}^2\) sending the \(j^{th}\) standard basis vector to \(v_j\) is \(D_n\)-equivariant, where \(D_n\) acts on \(\mathbb{R}^n\) by permutation and on \(\mathbb{R}^2\) via the geometric representation.

**Example 9.2.6.** 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 \(\mathcal{S}_4\) on the vertices of \(K_4\), which
induces a representation $\rho_V$ on the space $V = k\langle v_1, \ldots, v_4 \rangle \cong k^4$. This action naturally determines an action on the six edges of $K_4$, which in turn induces a permutation representation $\rho_E$ on $E = k\langle e_{12}, \ldots, e_{34} \rangle \cong k^6$. The relation between the two actions can be described by a $G$-equivariant transformation — but be careful: it is not a map $V \to E$ but a map $E \to V$, namely

$$\phi(e_{ij}) = v_i + v_j$$

so that $\rho_V \circ \phi(g) = \phi \circ \rho_E(g)$ for all $g$, i.e., the following diagram commutes:

$$
\begin{array}{ccc}
E & \xrightarrow{\phi} & V \\
\rho_E(g) \downarrow & & \downarrow \rho_V(g) \\
E & \xrightarrow{\phi} & V
\end{array}
$$

**Example 9.2.7.** Consider the representations $\rho_{\text{triv}}$ and $\rho_{\text{sign}}$ of $\mathfrak{S}_2$. Recall that these are the representations on $V = k$ given by

$$\rho_{\text{triv}}(\text{id}) = 1, \quad \rho_{\text{sign}}(\text{id}) = 1, \quad \rho_{\text{sign}}((1 2)) = -1.$$ 

If $\rho = \rho' \in \{\rho_{\text{triv}}, \rho_{\text{sign}}\}$, then any linear transformation $\phi : k \to k$ (i.e., any map $\phi(x) = cx$ for some $c \in k$) satisfies the commutative diagram (9.2.1), and if $c = 0$ then $\phi$ is an isomorphism. So in these cases the set of $G$-equivariant homomorphisms is actually isomorphic to $k$. On the other hand, if $\phi : \rho_{\text{triv}} \to \rho_{\text{sign}}$ is $G$-equivariant, then for every $x \in k$,

$$\rho_{\text{sign}}(\phi(v)) = -\phi(v) = -cv, \quad \phi(\rho_{\text{triv}}(v)) = \phi(v) = cv,$$

and if $\text{char } k \neq 2$ then these things are only equal if $c = 0$, so there is no nontrivial $G$-homomorphism $\rho_{\text{triv}} \to \rho_{\text{sign}}$.

This example is the tip of an iceberg: we can use the set of $G$-homomorphisms $\phi : \rho \to \rho'$ to measure how similar $\rho$ and $\rho'$ are. It will help that the set is in fact a $G$-representation in its own right.

**Example 9.2.8.** Let $k$ be a field of characteristic $\neq 2$, and let $V = k^2$, with standard basis $\{e_1, e_2\}$. Let $G = \mathfrak{S}_2 = \{12, 21\}$. The defining representation $\rho = \rho_{\text{def}}$ of $G$ on $V$ is given by

$$\rho(12) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \rho(21) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$ 

On the other hand, the representation $\sigma = \rho_{\text{triv}} \oplus \rho_{\text{sign}}$ on $V$ is given by

$$\sigma(12) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \sigma(21) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$ 

These two representations are in fact isomorphic. Indeed, $\rho$ acts trivially on $k\langle e_1 + e_2 \rangle$ and acts by $-1$ on $k\langle e_1 - e_2 \rangle$. These two vectors form a basis of $V$, and one can check that the change-of-basis map $\phi$ is an isomorphism $\rho \to \sigma$, i.e.,

$$\phi = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix}.$$
9.3 Irreducibility, indecomposability and Maschke’s theorem

Definition 9.3.1. Let \((\rho, V)\) be a representation of \(G\).

- A vector subspace \(W \subseteq V\) is \(G\)-invariant if \(gW \subseteq W\) for every \(g \in G\). Equivalently, \(gW = W\) (because \(g\) is invertible, hence has no kernel) for every \(g \in G\), so that \((\rho|_W, W)\) is a representation of \(G\) (the restriction of \(\rho\) to \(W\)). In module language, \(W\) is a \(kG\)-submodule of \(V\).
- \(V\) is decomposable if there are nontrivial \(G\)-invariant subspaces \(W, W' \unlhd V\) with \(W \cap W' = 0\) and \(W + W' = V\). Here \(W'\) is called the complement of \(W\); 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 9.3.2. As in Example 9.2.8, let \(V = \{e_1, e_2\}\) be the standard basis for \(k^2\), where \(\text{char} k \neq 2\). Recall that the defining representation of \(S_2 = \{12, 21\}\) is given by
\[
\rho_{\text{def}}(12) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \rho_{\text{def}}(21) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}
\]
and the change-of-basis map
\[
\phi = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix}
\]
is a \(G\)-equivariant isomorphism \(\rho_{\text{def}} \rightarrow \rho_{\text{triv}} \oplus \rho_{\text{sign}}\). On the other hand, if \(\text{char} k = 2\), then the matrix \(\phi\) is not invertible and this argument breaks down. For instance, if \(k = \mathbb{Z}/2\mathbb{Z}\), then \(W\) is the only \(G\)-invariant subspace of \(V\), and consequently \(\rho_{\text{def}}\) is not semisimple.

Fortunately, we can rule out this kind of pathology most of the time.

Theorem 9.3.3 (Maschke’s Theorem). Let \(G\) be a finite group, let \(k\) be a field whose characteristic does not divide \(|G|\), and let \((\rho, V)\) be a representation of \(G\) over \(k\). Then every \(G\)-invariant subspace has a \(G\)-invariant complement. In particular, \((\rho, V)\) is semisimple.

Proof. If \(\rho\) is irreducible, then there is nothing to prove. Otherwise, let \(W\) be a \(G\)-invariant subspace, and let \(\pi : V \rightarrow W\) be a projection, i.e., a linear surjection that fixes the elements of \(W\) pointwise. (Such a map \(\pi\) can be constructed as follows: choose a basis for \(W\), extend it to a basis for \(V\), and let \(\pi\) fix all the basis elements in \(W\) and kill all the ones in \(V \setminus W\).)

The map \(\pi\) is \(k\)-linear, but not necessarily \(G\)-equivariant. However, we can turn \(\pi\) into a \(G\)-equivariant projection by “averaging over \(G\)”. Specifically, define a map \(\tilde{\pi} : V \rightarrow W\) by
\[
\tilde{\pi}(v) \overset{\text{def}}{=} \frac{1}{|G|} \sum_{g \in G} g\pi(g^{-1}v).
\]
(9.3.1)

First, \(\tilde{\pi}(v)\) does indeed belong to \(W\) for all \(v \in V\), because \(\pi(g^{-1}v) \in W\) and \(W\) is \(G\)-invariant.
Second, we show that \( \tilde{\pi} \) fixes \( W \) pointwise. Indeed, if \( w \in W \), then \( g^{-1}w \in W \), so \( \pi(g^{-1}w) = g^{-1}w \) and

\[
\tilde{\pi}(w) = \frac{1}{|G|} \sum_{g \in G} gg^{-1}w = w.
\]

These first and second facts imply that \( \tilde{\pi} \) is also a projection \( V \to W \).

Third, we check that \( \tilde{\pi} \) is \( G \)-equivariant, i.e., \( \tilde{\pi}(hv) = h\tilde{\pi}(v) \) for all \( h \in G \) and \( v \in V \). Indeed,

\[
\tilde{\pi}(hv) = \frac{1}{|G|} \sum_{g \in G} g\pi(g^{-1}hv) = \frac{1}{|G|} \sum_{k \in G: hk = g} (hk)\pi((hk)^{-1}hv) = \frac{1}{|G|} h \sum_{k \in G} k\pi(k^{-1}v) = h\tilde{\pi}(v).
\]

Now that we have a \( G \)-equivariant projection, we can define \( W^\perp = \ker \tilde{\pi} \). Certainly \( V \cong W \oplus W^\perp \) as vector spaces. By \( G \)-equivariance, if \( v \in W^\perp \) and \( g \in G \), then \( \tilde{\pi}(gv) = g\tilde{\pi}(v) = 0 \), i.e., \( gv \in W^\perp \). That is, \( W^\perp \) is \( G \)-invariant.

Note that if \( \text{char} \, k \) does divide \( |G| \), then the proof breaks down because the map \( v \mapsto \sum_{g \in G} g\pi(g^{-1}v) = |G|v \) will kill everything in \( W \) instead of preserving it.

Maschke’s Theorem implies that, when the conditions hold, a representation \( \rho \) is determined up to isomorphism by the multiplicity of each irreducible representation in \( \rho \). Accordingly, to understand representations of \( G \), we should first study irreps. By the way, implicit in the proof is the following useful fact (which does not require any assumption on \( \text{char} \, k \)):

**Proposition 9.3.4.** Any \( G \)-equivariant map has \( G \)-equivariant kernel and \( G \)-equivariant image.

This is again a familiar fact from the module-theoretic standpoint — every kernel or image of a module homomorphism is also a module.

**Example 9.3.5.** Let \( k \) have characteristic 0 (for simplicity), and \( G = S_n \). The defining representation of \( G \) on \( k^n \) is not simple, because it has an invariant subspace, namely the span of the all-1’s vector, a 1-dimensional subspace \( L \) that is fixed pointwise by every \( \sigma \in S_n \) and therefore carries the trivial representations.\(^1\) 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 \( k^n \), namely the space of all vectors whose coordinates sum to 0. This is called (a little confusingly) the **standard representation** of \( S_n \), denoted \( \rho_{\text{std}} \). 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 \( \rho_{\text{std}} \) is irreducible (Exercise 9.2).

### 9.4 Characters

The first miracle of representation theory is that we can detect the isomorphism type of a representation \( \rho \) without knowing every coordinate of every matrix \( \rho(g) \): it turns out that all we need to know is the *traces* of the \( \rho(g) \).

\(^1\)For the same reason, every permutation representation of every group has a trivial summand.
Definition 9.4.1. Let \((\rho, V)\) be a representation of \(G\) over \(k\). Its **character** is the function \(\chi_\rho : G \to k\) given by

\[
\chi_\rho(g) = \text{tr} \rho(g).
\]

**Example 9.4.2.** Some simple facts and some characters we’ve seen before:

- A one-dimensional representation is its own character.
- For any representation \(\rho\), we have \(\chi_\rho(\text{Id}_G) = \dim \rho\), because \(\rho(\text{Id}_G)\) is the \(n \times n\) identity matrix.
- The defining representation \(\rho_{\text{def}}\) of \(S_n\) has character \(\chi_{\text{def}}(\sigma) = \text{number of fixed points of } \sigma\).
  - Indeed, this is true for any permutation representation of any group.
- The regular representation \(\rho_{\text{reg}}\) has character \(\chi_{\text{reg}}(\sigma) = \begin{cases} |G| & \text{if } \sigma = \text{Id}_G \\ 0 & \text{otherwise.} \end{cases}\)

**Example 9.4.3.** Consider the geometric representation \(\rho_{\text{geo}}\) of the dihedral group \(D_n = \langle r, s \mid r^n = s^2 = 0, srs = r^{-1} \rangle\) by rotations and reflections:

\[
\rho_{\text{geo}}(s) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad \rho_{\text{geo}}(r) = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}
\]

The character of \(\rho_{\text{geo}}\) is

\[
\chi_{\rho_{\text{geo}}}(r^j) = 2 \cos \theta \quad (0 \leq j < n), \quad \chi_{\rho_{\text{geo}}}(sr^j) = 0 \quad (0 \leq j < n).
\]

On the other hand, if \(\rho'\) 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 9.4.4.** 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.** Recall from linear algebra that \(\text{tr}(ABA^{-1}) = \text{tr}(B)\) in general. Therefore,

\[
\text{tr} \left( \rho(hgh^{-1}) \right) = \text{tr} \left( \rho(h)\rho(g)\rho(h^{-1}) \right) = \text{tr} \left( \rho(h)\rho(g)\rho(h)\rho(h)^{-1} \right) = \text{tr} \rho(g).
\]

For the second assertion, let \(\phi : \rho \to \rho'\) be an isomorphism, i.e., \(\phi \cdot \rho(g) = \rho'(g) \cdot \phi\) for all \(g \in G\) (treating \(\phi\) as a matrix in this notation). Since \(\phi\) is invertible, we have therefore \(\phi \cdot \rho(g) \cdot \phi^{-1} = \rho'(g)\). Now take traces.
It turns out that the converse of the second assertion is also true: a representation is determined by its character. In fact, much, much more is true.

### 9.5 New characters from old

The basic vector space functors of direct sum, duality, tensor product and Hom carry over naturally to representations, and behave well on their characters. Throughout this section, let \((\rho, V)\) and \((\rho', V')\) be finite-dimensional representations of \(G\) over \(\mathbb{C}\), with \(V \cap V' = 0\).

1. **Direct sum.** To construct a basis for \(V \oplus V'\), we can take the union of a basis for \(V\) and a basis for \(V'\). Equivalently, we can write the vectors in \(V \oplus V'\) as column block vectors:

\[
V \oplus V' = \left\{ \begin{bmatrix} v \\ v' \end{bmatrix} \mid v \in V, \ v' \in V' \right\}.
\]

Accordingly, we can define the direct sum \((\rho \oplus \rho', V \oplus V')\) by

\[
(\rho \oplus \rho')(h) = \begin{bmatrix} \rho(h) & 0 \\ 0 & \rho'(h) \end{bmatrix}.
\]

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 \(k\)-linear transformations \(\phi : V \to k\). A \(G\)-representation \((\rho, V)\) gives rise to a representation \((\rho^*, V^*)\) given by

\[
(\rho^*)(h)(v) = \phi(h^{-1}v)
\]

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 \(\rho^*\) to be a homomorphism (i.e., for \(\rho^*(gh) = \rho^*(g)\rho^*(h)\) rather than \(\rho^*(h)\rho^*(g)\)). The representation \(\rho^*\) is called the **dual representation** (or **contragredient**) of \(\rho\).

**Proposition 9.5.1.** For every \(h \in G\),

\[
\chi_{\rho^*}(h) = \overline{\chi_\rho(h)}.
\]

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 \(\chi_{\rho(h)} = \text{tr} J\). 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 \(\text{tr} J^{-1} = \chi_{\rho^*}(h) = \overline{\chi_\rho(h)}\).

3. **Tensor product.** Let \(V = k\langle e_1, \ldots, e_n \rangle\) and \(V' = k\langle e'_1, \ldots, e'_m \rangle\). As a vector space, we define\(^2\)

\[
V \otimes V' = k \langle e_i \otimes e'_j \mid 1 \leq i \leq n, \ 1 \leq j \leq m \rangle,
\]

equipped with a multilinear action of \(k\) (that is, \(c(x \otimes y) = cx \otimes y = x \otimes cy\) for \(c \in k\)). In particular, \(\dim(V \otimes V') = (\dim V)(\dim V')\). We can accordingly define a representation \((\rho \otimes \rho', V \otimes V')\) by

\[
(\rho \otimes \rho')(h)(v \otimes v') = \rho(h)(v) \otimes v' + v \otimes \rho'(h)(v')
\]

\(^2\)The “official” definition of the tensor product is much more functorial and can be made basis-free, but this definition is equivalent and is more convenient for our present purposes.

130
or more concisely
\[ h \cdot (v \otimes v') = hv \otimes v' + v \otimes hv' \]
extended bilinearly to all of \( V \otimes V' \).

In terms of matrices, \( (\rho \otimes \rho')(h) \) is represented by the \( nm \times nm \) matrix in block form
\[
\begin{bmatrix}
a_{11} \rho'(h) & \cdots & a_{1n} \rho'(h) \\
\vdots & \ddots & \vdots \\
a_{n1} \rho'(h) & \cdots & a_{nn} \rho'(h)
\end{bmatrix}
\]
where \( \rho(h) = [a_{ij}]_{i,j=1,...,n} \). For example, if \( \rho(h) = \begin{bmatrix} 1 & 1 \\ \omega & -\omega^2 \end{bmatrix} \) and \( \rho'(h) = \begin{bmatrix} \omega & 0 & 0 \\ 0 & 0 & 1 \\ 0 & \omega^2 & 0 \end{bmatrix} \) (where \( \omega = e^{2\pi i/3} \)),
then
\[
(\rho \otimes \rho')(h) = \begin{bmatrix}
\omega & 0 & 0 & \omega & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 \\
0 & \omega^2 & 0 & 0 & \omega^2 & 0 \\
\omega^2 & 0 & 0 & -1 & 0 & 0 \\
0 & 0 & \omega & 0 & 0 & -\omega^2 \\
0 & 1 & 0 & 0 & -\omega & 0
\end{bmatrix}
\]
Taking traces, we see that
\[
\chi_{\rho \otimes \rho'}(h) = \chi_{\rho}(h) \chi_{\rho'}(h).
\] (9.5.2)

4. **Hom.** There are two kinds of Hom. Let \( V \) and \( W \) be representations of \( G \). Then the vector space
\[
\text{Hom}_\mathbb{C}(V,W) = \{ \mathbb{C}\text{-linear transformations } \phi : V \to W \}
\]
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
\[
(h \cdot \phi)(v) = h(\phi(h^{-1}v)) = \rho'(h)\left( \phi\left( \rho(h^{-1})(v) \right) \right).
\] (9.5.3)
for \( h \in G, \phi \in \text{Hom}_\mathbb{C}(V,W), v \in V \). It is straightforward to verify that this is a genuine group action, i.e., that \( (hh') \cdot \phi = h \cdot (h' \cdot \phi) \). Moreover, \( \text{Hom}_\mathbb{C}(V,W) \cong V^* \otimes W \) as vector spaces, and this isomorphism is \( G \)-equivariant, so
\[
\chi_{\text{Hom}(\rho,\rho')}(h) = \chi_{\rho}(h) \chi_{\rho'}(h).
\] (9.5.4)

The other kind of Hom is
\[
\text{Hom}_G(V,W) = \{ \text{\( G \)-equivariant linear transformations } \phi : V \to W \}.
\]
Evidently \( \text{Hom}_G(V,W) \subseteq \text{Hom}_\mathbb{C}(V,W) \), but equality need not hold. For example, if \( V \) and \( W \) are the trivial and sign representations of \( S_n \) (for \( n \geq 2 \)), then \( \text{Hom}_\mathbb{C}(V,W) \cong \mathbb{C} \) but \( \text{Hom}_G(V,W) = 0 \). See Example 9.2.7.

When is a linear transformation
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
\[
V^G = \{ v \in V \mid hv = h \quad \forall h \in G \}.
\]
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 \( \text{Hom}_C(V,W) \) as above. (The proof of this fact is left to the reader; it is nearly immediate from the definition of that action.) That is,

\[
\text{Hom}_G(V,W) = \text{Hom}_C(V,W)^G. \tag{9.5.5}
\]

To summarize:

\[
\begin{aligned}
\chi_{\rho \oplus \rho'} &= \chi_\rho + \chi_{\rho'}, \\
\chi_{\rho \otimes \rho'} &= \chi_\rho \cdot \chi_{\rho'}, \\
\chi_{\text{Hom}_C(\rho, \rho')} &= \overline{\chi_\rho} \cdot \chi_{\rho'}.
\end{aligned}
\]

However, we still need a way of computing the character of \( \text{Hom}_G(\rho, \rho') \).

### 9.6 The inner product on class functions; Schur’s Lemma

Recall that a class function is a function \( \chi : G \to \mathbb{C} \) that is constant on conjugacy classes of \( G \). Define an inner product on the vector space \( C\ell(G) \) of \( \mathbb{C} \)-valued class functions by

\[
\langle \chi, \psi \rangle_G = \frac{1}{|G|} \sum_{h \in G} \chi(h)\overline{\psi}(h).
\]

It is clear that this 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 9.6.1.** Let \( (\rho, V) \) be a representation of \( G \). Then

\[
\dim \mathbb{C} V^G = \frac{1}{|G|} \sum_{h \in G} \chi_\rho(h) = \langle \chi_{\text{triv}}, \chi_\rho \rangle_G.
\]

**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, \( \pi \) is a projection from \( V \to V^G \). Choose a basis for \( V^G \) and extend it to a basis for \( V \), so that \( \pi \) can be represented by the block matrix

\[
\begin{bmatrix}
I & * \\
0 & 0
\end{bmatrix}
\]

where the first and second column blocks (resp., row blocks) correspond to \( V^G \) and \( (V^G)^\perp \) respectively. Therefore

\[
\dim \mathbb{C} V^G = \text{tr}(\pi) = \frac{1}{|G|} \sum_{h \in G} \chi_\rho(h).
\]

132
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 9.6.2.** Suppose that $\rho$ is a permutation representation. Then $V^G$ is the space of functions that are constant on the orbits, and its dimension is the number of orbits. Therefore, the formula of Proposition 9.6.1 becomes

$$\# \text{ orbits} = \frac{1}{|G|} \sum_{h \in G} \# \text{ fixed points of } h$$

which is Burnside’s Lemma from basic abstract algebra.

**Proposition 9.6.3.** For any two representations $\rho, \rho'$ of $G$, we have $\langle \chi_\rho, \chi_{\rho'} \rangle_G = \dim \text{Hom}_G(\rho, \rho')$.

**Proof.** Calculate $\langle \chi_\rho, \chi_{\rho'} \rangle_G$ as

$$\frac{1}{|G|} \sum_{h \in G} \chi_\rho(h) \chi_{\rho'}(h) = \frac{1}{|G|} \sum_{h \in G} \chi_{\text{Hom}(\rho, \rho')}(h) \quad \text{(by (9.5.4))}$$

$$= \dim \text{Hom}(\rho, \rho')^G \quad \text{(by Proposition 9.6.1)}$$

$$= \dim \text{Hom}_G(\rho, \rho') \quad \text{(by (9.5.5)).}$$

One intriguing observation is that this expression is symmetric in $\rho$ and $\rho'$, which is not necessarily obvious from the definition. (For vector spaces the equality is clearer; in fact $\text{Hom}_k(V, W) \cong \text{Hom}_k(V, W)^*$ — the isomorphism is matrix transpose.)

The inner product is the tool that lets us prove the following useful fact, known as Schur’s lemma.

**Proposition 9.6.4 (Schur’s Lemma).** Let $G$ be a group, and let $(\rho, V)$ and $(\rho', V')$ be finite-dimensional irreps of $G$ over a field $k$.

1. Every $G$-equivariant map $\phi : V \to V'$ is either zero or an isomorphism.
2. If in addition $k$ is algebraically closed, then

$$\text{Hom}_G(V, V') \cong \begin{cases} k & \text{if } \rho \cong \rho' \\ 0 & \text{otherwise.} \end{cases}$$

In particular, the only $G$-equivariant maps from an $G$-irrep to itself are multiplication by a scalar.

**Proof.** For (1), recall from Proposition 9.3.4 that $\ker \phi$ and $\im \phi$ are $G$-invariant subspaces. But since $\rho, \rho'$ are simple, there are not many possibilities. Either $\ker \phi = 0$ and $\im \phi = W$, when $\phi$ is an isomorphism. Otherwise, $\ker \phi = V$ or $\im \phi = 0$, either of which implies that $\phi = 0$.

For (2), let $\phi \in \text{Hom}_G(V, V')$. If $\rho \not\cong \rho'$ then $\phi = 0$ by (1) and we’re done. Otherwise, we may as well assume that $V = V'$.

Since $k$ is algebraically closed, $\phi$ has an eigenvalue $\lambda$. Then $\phi - \lambda I$ is $G$-equivariant and singular, hence zero by (1). So $\phi = \lambda I$. We’ve just shown that the only $G$-equivariant maps from $V$ to itself are multiplication by $\lambda$. 

\[133\]
9.7 The fundamental theorem of character theory for finite groups

We are ready to prove the following omnibus theorem, which says that all information about a representation of a finite group is contained in its character.

**Theorem 9.7.1.** Let $(\rho, V)$ and $(\rho', V')$ be finite-dimensional representations of $G$ over $\mathbb{C}$.

1. If $\rho$ and $\rho'$ are irreducible, then

$$
\langle \chi_\rho, \chi_{\rho'} \rangle_G = \begin{cases} 
1 & \text{if } \rho \cong \rho', \\
0 & \text{otherwise}. 
\end{cases}
$$

In particular, there are only finitely many isomorphism classes of irreps of $G$.

2. If $\rho_1, \ldots, \rho_n$ are distinct irreducible representations and

$$
\rho = \bigoplus_{i=1}^n (\rho_i \oplus \cdots \oplus \rho_i) = \bigoplus_{i=1}^n \rho_i^\oplus_{m_i},
$$

then

$$
\langle \chi_\rho, \chi_{\rho_i} \rangle_G = m_i \quad \text{and} \quad \langle \chi_\rho, \chi_\rho \rangle_G = \sum_{i=1}^n m_i^2.
$$

In particular, $\langle \chi_\rho, \chi_\rho \rangle_G = 1$ if and only if $\rho$ is irreducible.

3. If $\chi_\rho = \chi_{\rho'}$, then $\rho \cong \rho'$.

4. If $\rho_1, \ldots, \rho_n$ is a complete list of irreps of $G$, then

$$
\rho_{\text{reg}} \cong \bigoplus_{i=1}^n \rho_i^\oplus_{\dim \rho_i},
$$

and consequently

$$
\sum_{i=1}^n (\dim \rho_i)^2 = |G|. \tag{9.7.1}
$$

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.** Assertion (1) follows from part (2) of Schur’s Lemma together with Proposition 9.6.3. It follows that the characters of isomorphism classes of irreps form a linearly independent subset of the finite-dimensional space $C\ell(G)$, so there can be only finitely many of them.

Assertion (2) follows because the inner product is bilinear on direct sums. That is, for any class function $\psi$, we have

$$
\langle \chi_\rho \oplus \chi_{\rho'}, \psi \rangle_G = \langle \chi_\rho + \chi_{\rho'}, \psi \rangle_G = \langle \chi_\rho, \psi \rangle_G + \langle \chi_{\rho'}, \psi \rangle_G.
$$

For (3), Maschke’s Theorem says that every complex representation $\rho$ can be written as a direct sum of irreducibles. Their multiplicities determine $\rho$ up to isomorphism, and can be recovered from $\chi_\rho$ by (2).

For (4), recall that $\chi_{\text{reg}}(\text{Id}_G) = |G|$ and $\chi_{\text{reg}}(g) = 0$ for $g \neq \text{Id}_G$. Therefore

$$
\langle \chi_{\text{reg}}, \rho_i \rangle_G = \frac{1}{|G|} \sum_{g \in G} \chi_{\text{reg}}(g) \rho_i(g) = \frac{1}{|G|} |G| \rho_i(\text{Id}_G) = \dim \rho_i
$$
so \( \rho_i \) appears in \( \rho_{reg} \) with multiplicity equal to its dimension.

For (5), the irreducible characters are orthonormal (hence linearly independent in \( \text{Cl}(G) \)), by Schur’s Lemma together with assertion (3). The trickier part is to show that they in fact span \( \text{Cl}(G) \). Let \( Y \) be the subspace of \( \text{Cl}(G) \) spanned by the irreducible characters, and let

\[
Z = \{ \phi \in \text{Cl}(G) : \langle \phi, \chi_{\rho} \rangle_G = 0 \text{ for every irreducible character } \rho \}.
\]

That is, \( Z = Y^\perp \) with respect to \( \langle \cdot, \cdot \rangle_G \). We will show that in fact \( Z = 0 \).

Let \( \phi \in Z \). For any representation \((\rho, V)\), define a map \( T_\rho = T_{\rho, \phi} : V \to V \) by

\[
T_\rho = \frac{1}{|G|} \sum_{g \in G} \phi(g) \rho(g)
\]

or equivalently

\[
T_\rho(v) = \frac{1}{|G|} \sum_{g \in G} \phi(g) gv
\]

where \( g \) acts on \( v \in V \) by \( \rho(g) \); note that \( \phi(g) \) is a number.

**Claim:** \( T_\rho \) is the zero map (in disguise).

First we show that \( T_\rho \) is \( G \)-equivariant. Let \( h \in G \); then

\[
T_\rho(hv) = \frac{1}{|G|} \sum_{g \in G} \phi(g) ghv = \frac{1}{|G|} h \sum_{g \in G} \phi(g) h^{-1} ghv
\]

\[
= h \frac{1}{|G|} \sum_{k \in G} \phi(k) h^{-1} (hkh^{-1}) (hv) = \frac{1}{|G|} \sum_{k \in G} \phi(k) (kv)
\]

(setting \( k = h^{-1} gh, hkh^{-1} = g \))

\[
= h \frac{1}{|G|} \sum_{k \in G} \phi(k) (kv)
\]

(because \( \phi \in \text{Cl}(G) \))

\[
= h T_\rho(v).
\]

Second, we prove that \( T_\rho = 0 \) if \( \rho \) is irreducible. By Schur’s Lemma, \( T_\rho \) is multiplication by a scalar (possibly zero). On the other hand

\[
\text{tr}(T_\rho) = \frac{1}{|G|} \sum_{g \in G} \phi(g) \chi_{\rho}(g) = \langle \phi, \chi_{\rho} \rangle_G = 0
\]

by the assumption \( \phi \in Z \). Since \( T_\rho \) has trace zero and is multiplication by a scalar, it is the zero map.

Third, the definition of \( T \) implies that it is additive on direct sums (that is, \( T_{\rho \oplus \rho'} = T_\rho + T_{\rho'} \)), so by Maschke’s Theorem, \( T_\rho = 0 \) for every representation \( \rho \), proving the claim.

Now take \( \rho = \rho_{reg} \). Then

\[
0 = T_{\rho_{reg}}(\text{Id}_G) = \frac{1}{|G|} \sum_{g \in G} \phi(g) g.
\]

This is an equation in the group algebra \( CG \) (the underlying space for the regular representation). Equating coefficients of \( g \) on either side implies that \( \phi(g) = 0 \) for every \( g \in G \), so \( \phi = 0 \). We have now shown that \( Y \) has trivial orthogonal complement as a subspace of \( \text{Cl}(G) \), so \( Y = \text{Cl}(G) \), completing the proof. \( \square \)
9.8 Computing character tables

Theorem 9.7.1 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 $\text{Cl}(G)$. 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 defining characters.
- 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.

In the following examples, we will notate a character $\chi$ 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 9.8.1.** The group $G = \mathfrak{S}_3$ has three conjugacy classes, determined by cycle shapes:

- $C_{111} = \{\text{Id}_G\}$,
- $C_{21} = \{(12), (13), (23)\}$,
- $C_3 = \{(123), (132)\}$.

We already know two irreducible 1-dimensional characters of $\mathfrak{S}_3$, namely the trivial character and sign characters. Also, we always have the regular character $\chi_{\text{reg}} = [6, 0, 0]$. So we begin with the following character table:

<table>
<thead>
<tr>
<th>Size</th>
<th>Conj. class</th>
<th>$C_{111}$</th>
<th>$C_{21}$</th>
<th>$C_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\chi_1 = \chi_{\text{triv}}$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>$\chi_2 = \chi_{\text{sign}}$</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>$\chi_{\text{reg}}$</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

Equation (9.7.1) says that $\mathfrak{S}_3$ has three irreps, the squares of whose dimensions add up to $|\mathfrak{S}_3| = 6$. So we are looking for one more irreducible character $\chi_{\text{other}}$ of dimension 2. By (e) of Theorem 9.7.1, we have

$$\chi_{\text{reg}} = \chi_{\text{triv}} + \chi_{\text{sign}} + 2\chi_{\text{other}}$$

from it which is easy to obtain

$$\chi_{\text{other}} = [2, 0, -1].$$

One can check that $\chi_{\text{other}}$ is irreducible by confirming that its scalar product with itself is 1. By the way, the defining representation of $\mathfrak{S}_3$ is $\chi_{\text{def}} = \chi_{\text{triv}} \oplus \chi_{\text{other}}$.

**Example 9.8.2.** We calculate all the irreducible characters of $\mathfrak{S}_4$. There are five conjugacy classes, corresponding to the cycle-shapes 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 start by writing down some characters that we know.
Of course $\chi_{\text{triv}}$ and $\chi_{\text{sign}}$ are irreducible, since they are 1-dimensional. On the other hand, $\chi_{\text{def}}$ can’t be irreducible because $\mathfrak{S}_4$ doesn’t have a 4-dimensional irrep. Indeed,

$$\langle \chi_{\text{def}}, \chi_{\text{def}} \rangle_G = 2$$

which means that $\rho_{\text{def}}$ must be a direct sum of two distinct irreps. (If it were the direct sum of two copies of the unique 2-dimensional irrep, then $\langle \chi_{\text{def}}, \chi_{\text{def}} \rangle_G$ would be 4, not 2, by (ii) of Theorem 9.7.1.) We calculate

$$\langle \chi_{\text{def}}, \chi_{\text{triv}} \rangle_G = 1, \quad \langle \chi_{\text{def}}, \chi_{\text{sign}} \rangle_G = 0.$$ 

Therefore $\chi_3 = \chi_{\text{def}} - \chi_{\text{triv}}$ is an irreducible character.

Another 3-dimensional character is $\chi_4 = \chi_3 \otimes \chi_{\text{sign}}$. It is easy to check that $\langle \chi_4, \chi_4 \rangle_G = 1$, so $\chi_4$ is irreducible.

The other irreducible character $\chi_5$ 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

$$\chi_5 = \frac{\chi_{\text{reg}} - \chi_1 - \chi_2 - 3\chi_3 - 3\chi_4}{2}.$$ 

and so the complete character table of $\mathfrak{S}_4$ is as follows.

<table>
<thead>
<tr>
<th>Size</th>
<th>Conjugacy Class</th>
<th>$\chi_1$</th>
<th>$\chi_2$</th>
<th>$\chi_3$</th>
<th>$\chi_4$</th>
<th>$\chi_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$C_{111}$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>$C_{211}$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>$C_{22}$</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td></td>
<td>$C_{31}$</td>
<td>3</td>
<td>1</td>
<td>-1</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td></td>
<td>$C_{4}$</td>
<td>3</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

9.9 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 characters, i.e., $\text{Ch}(G) = \text{Hom}(G, \mathbb{C}^\times)$, is an abelian group under tensor product (equivalently, pointwise multiplication), with identity $\chi_{\text{triv}}$.

**Definition 9.9.1.** 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 $G$.  

137
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 \( \chi([a, b]) = \text{Id}_G \) for all \( \chi \in \text{Ch}(G) \) and \( a, b \in G \).

(In fact, this condition characterizes the elements of the commutator subgroup, as will be shown soon.) Therefore, when studying characters on \( G \), you might as well assume \( G \) is abelian (i.e., \( \text{Ch}(G) \cong \text{Ch}(G^{ab}) \)).

Accordingly, let \( G \) be an abelian group of finite order \( n \). The conjugacy classes of \( G \) are all singleton sets (since \( g h g^{-1} = h \) for all \( g, h \in G \)), so there are \( n \) distinct irreducible representations of \( G \). On the other hand,

\[
\sum_{\chi \text{ irreducible}} (\dim \chi)^2 = n
\]

by Theorem 9.7.1 (iv), 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 \( \chi \in \text{Ch}(G) \) is determined by its value on \( g \), which must be some \( n \)th 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 a primitive \( n \)th root of unity. So \( \text{Hom}(G, \mathbb{C}^\times) \cong G \) (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 \( i \)th factor, and let \( \zeta_i \) be a primitive \( (n_i) \)th root of unity. Then each character \( \chi \) 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 \( \text{Hom}(G, \mathbb{C}^\times) \cong G \), an isomorphism known as Pontryagin duality. More generally, for any finite group \( G \) (not necessarily abelian), there is an isomorphism

\[
\text{Hom}(G, \mathbb{C}^\times) \cong G^{ab}.
\]

(9.9.1)

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 description of characters of abelian groups implies that if \( G \) is abelian and \( g \neq \text{Id}_G \), then \( \chi(g) \neq 1 \) for at least one character \( \chi \). Therefore, for every group \( G \), we have

\[
[G, G] = \{ g \in G \mid \chi(g) = 1 \text{ for all 1-dimensional characters } \chi \}
\]

since \([G, G]\) is the identity element of \( G^{ab} \). This

**Example 9.9.2.** Suppose that \( G \) is a group of order 24 with 8 conjugacy classes.\(^3\) 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 1,1,1,1,1,3,3. 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.

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.

\(^3\)According to Group Properties Wiki (10/29/18), there happens to be exactly one such group, namely \( \mathfrak{A}_4 \times \mathbb{Z}_2 \).
Example 9.9.3. Consider the case $G = S_n$. Certainly $[S_n, S_n] \subseteq A_n$, 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 $S_n$ (multiplying left to right) shows that $[S_n, S_n]$ contains every 3-cycle, and it is not hard to show that the 3-cycles generate the full alternating group. Therefore (9.9.1) gives
\[
\text{Hom}(S_n, \mathbb{C}^\times) \cong \mathbb{S}_n / A_n \cong \mathbb{Z}/2\mathbb{Z}.
\]

It follows that $\chi_{\text{triv}}$ and $\chi_{\text{sign}}$ are the only one-dimensional characters of $S_n$. 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 $S_n$, hence determine the character completely.

For instance, suppose we want to compute the character table of $S_5$ (Exercise 9.5), which has seven conjugacy classes. There are 21 lists of seven positive integers whose squares add up to $|S_5| = 5! = 120$, but only four of them that contain exactly two 1’s:
\[
1, 1, 2, 2, 5, 9, \quad 1, 1, 2, 2, 5, 6, 7, \quad 1, 1, 2, 3, 4, 5, 8, \quad 1, 1, 4, 4, 5, 5, 6.
\]

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.

### 9.10 Restriction, induction, and Frobenius reciprocity

**Definition 9.10.1.** Let $H \subseteq G$ be finite groups, and let $\rho : G \to GL(V)$ be a representation of $G$. Then the restriction of $\rho$ to $H$ is a representation of $G$, denoted $\text{Res}_{G}^{H}(\rho)$. Alternative notations include $\text{Res}(\rho)$ (if the groups $G$ and $H$ are clear from context) and $\rho^{\downarrow H}$. Likewise, the restriction of $\chi = \chi_{\rho}$ to $H$ is a character of $H$ denoted by $\text{Res}_{G}^{H}(\chi)$.

Restricting a representation does not change its character. On the other hand, the restriction of an irreducible representation is not always irreducible.

**Example 9.10.2.** Let $C_{\lambda}$ denote the conjugacy class in $S_n$ of permutations of cycle-shape $\lambda$. Recall that $G = S_3$ has an irrep, namely the Specht module $Sp_{(2,1)}$, whose character $\psi$ is given by
\[
\psi(C_{111}) = 2, \quad \psi(C_{21}) = 0, \quad \psi(C_3) = -1.
\]

Let $H = A_3 \subseteq S_3$. This is an abelian group (isomorphic to $\mathbb{Z}/3\mathbb{Z}$), so the two-dimensional representation $\text{Res}(\rho)$ is not irreducible. Indeed,
\[
\langle \psi, \psi \rangle_G = \frac{2^2 + 3 \cdot 0^2 + 2 \cdot (-1)^2}{6} = \frac{6}{6} = 1, \quad \langle \psi, \psi \rangle_H = \frac{6}{3} = 2.
\]

Let $\omega = e^{2\pi i/3}$. Then the table of irreducible characters of $A_3$ is as follows:

<table>
<thead>
<tr>
<th>Character</th>
<th>$A_3$</th>
<th>$S_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\chi_{\text{triv}}$</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$\chi_1$</td>
<td>1</td>
<td>$\omega$</td>
</tr>
<tr>
<td>$\chi_2$</td>
<td>1</td>
<td>$\omega^2$</td>
</tr>
</tbody>
</table>

(9.10.1)

Now it is evident that $\text{Res}(\psi) = [2, -1, -1] = \chi_1 + \chi_2$. 

139
With a little more work, it is also possible to construct a representation of $G$ from a representation $(\rho, W)$ of a subgroup $H \subseteq G$.

The idea is as follows. Let $V$ be a direct sum of copies of $W$, one for each (left) coset of $H$ in $G$. For the sake of bookkeeping, we pick a set of coset representatives $B = \{b_1, \ldots, b_n\}$, so that $G = b_1 H \cup \cdots \cup b_n H$, and we think of $V$ as a sum of copies of $W$, indexed by $B$. Technically, $V$ is the tensor product $\mathbb{C}B \otimes W = (b_1 \otimes W) \oplus \cdots \oplus (b_n \otimes W)$.

To say how an element $g \in G$ acts on the space $b_i \otimes W$, we want to write $gb_i$ in the form $b_k h$, where $k \in [n]$ and $h \in H$. We then make $g$ act by

$$g(b_i \otimes w) = b_k \otimes \rho(h)(w),$$

extended linearly to all of $V$. Heuristically, this formula is justified by the equation

$$g(b_i \otimes w) = gb_i \otimes w = b_k h \otimes w = b_k \otimes hw = b_k \otimes \rho(h)(w).$$

In other words, $g$ sends the $i$th copy of $W$ to the $k$th copy, acting by $h$ along the way. Thus we have a map $\text{Ind}^G_H(\rho)$ that sends each $g \in G$ to the linear transformation $V \to V$ just defined. Alternative notations for $\text{Ind}^G_H(\rho)$ include $\text{Ind}(\rho)$ (if $G$ and $H$ are clear from context) and $\rho \uparrow^G_H$.

**Example 9.10.3.** Let $G = \mathfrak{S}_3$ and $H = \mathfrak{A}_3 = \{\text{Id}, (1 \ 2 \ 3), (1 \ 3 \ 2)\}$, and let $(\rho, W)$ be a representation of $H$, where $W = (e_1, \ldots, e_n)$. Let $B = \{b_1 = \text{Id}, b_2 = (1 \ 2)\}$, so that $V = \mathbb{C}B \otimes W = \mathbb{C}(b_1 \otimes e_i, b_2 \otimes e)$. To define the induced representation, we need to solve the equations $gb_i = b_k h$:

<table>
<thead>
<tr>
<th>$g$</th>
<th>$gb_1 = b_k$</th>
<th>$h$</th>
<th>$gb_2 = b_k$</th>
<th>$h$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{Id}$</td>
<td>$\text{Id}$</td>
<td>$b_1 \text{Id}$</td>
<td>$b_2 \text{Id}$</td>
<td></td>
</tr>
<tr>
<td>$(1 \ 2 \ 3)$</td>
<td>$(1 \ 2 \ 3)$</td>
<td>$b_1 (1 \ 2 \ 3)$</td>
<td>$b_2 (1 \ 3 \ 2)$</td>
<td></td>
</tr>
<tr>
<td>$(1 \ 3 \ 2)$</td>
<td>$(1 \ 3 \ 2)$</td>
<td>$b_1 (1 \ 3 \ 2)$</td>
<td>$b_2 (1 \ 2 \ 3)$</td>
<td></td>
</tr>
<tr>
<td>$(1 \ 2)$</td>
<td>$(1 \ 2)$</td>
<td>$b_2 \text{Id}$</td>
<td>$b_1 \text{Id}$</td>
<td></td>
</tr>
<tr>
<td>$(1 \ 3)$</td>
<td>$(1 \ 3)$</td>
<td>$b_2 (1 \ 2 \ 3)$</td>
<td>$b_1 (1 \ 2 \ 3)$</td>
<td></td>
</tr>
<tr>
<td>$(2 \ 3)$</td>
<td>$(2 \ 3)$</td>
<td>$b_2 (1 \ 3 \ 2)$</td>
<td>$b_1 (1 \ 3 \ 2)$</td>
<td></td>
</tr>
</tbody>
</table>

Therefore, the representation $\text{Ind}^G_H(\rho)$ 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 $b_2 \otimes W$.

$$\text{Id} \mapsto \begin{bmatrix} \rho(\text{Id}) & 0 \\ 0 & \rho(\text{Id}) \end{bmatrix} \quad (1 \ 2 \ 3) \mapsto \begin{bmatrix} \rho(1 \ 2 \ 3) & 0 \\ 0 & \rho(1 \ 3 \ 2) \end{bmatrix} \quad (1 \ 3 \ 2) \mapsto \begin{bmatrix} \rho(1 \ 3 \ 2) & 0 \\ 0 & \rho(1 \ 2 \ 3) \end{bmatrix}$$

$$\begin{bmatrix} 0 & \rho(\text{Id}) \\ \rho(\text{Id}) & 0 \end{bmatrix} \quad (1 \ 2) \mapsto \begin{bmatrix} 0 & \rho(1 \ 2 \ 3) \\ \rho(1 \ 2 \ 3) & 0 \end{bmatrix} \quad (1 \ 3) \mapsto \begin{bmatrix} 0 & \rho(1 \ 3 \ 2) \\ \rho(1 \ 3 \ 2) & 0 \end{bmatrix} \quad (2 \ 3) \mapsto \begin{bmatrix} 0 & \rho(1 \ 3 \ 2) \\ \rho(1 \ 3 \ 2) & 0 \end{bmatrix}$$

For instance, if $\rho$ is the 1-dimensional representation (= character) $\chi_1$ of (9.10.1), then the character of $\text{Ind}(\rho)$ is given on conjugacy classes in $\mathfrak{S}_3$ by

$$\chi_{\text{Ind}(\rho)}(C_{111}) = 2, \quad \chi_{\text{Ind}(\rho)}(C_{21}) = \omega + \omega^2 = -1, \quad \chi_{\text{Ind}(\rho)}(C_3) = 0,$$

which implies that $\text{Ind}^G_H(\rho)$ is the nontrivial summand of the defining representation of $\mathfrak{S}_3$.

In fact, $\text{Ind}(\rho)$ is a representation, and there is a general combinatorial formula for its character.
Proposition 9.10.4. Let $H$ be a subgroup of $G$ and let $(\rho, W)$ be a representation of $H$ with character $\chi$. Then $\text{Ind}^G_H(\rho)$ is a representation of $G$, with character defined on $g \in G$ by

$$\text{Ind}^G_H(\chi)(g) = \frac{1}{|H|} \sum_{k \in G: k^{-1}gk \in H} \chi(k^{-1}gk).$$

In particular, $\text{Ind}(\rho)$ is independent (up to isomorphism) of the choice of $B$.

Proof. First, we verify that $\text{Ind}(\rho)$ 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

$$gb_k = b_kh$$

and in turn there is a unique $b_{k'} \in B$ and $h' \in H$ such that

$$g'gb_k = b_kh'^{1}.$$

We need to verify that $g' \cdot (g \cdot (b_i \otimes w)) = (g'g) \cdot (b_i \otimes w)$. Indeed,

$$\left(g' \cdot (g \cdot (b_i \otimes w)) = g' \cdot (b_k \otimes hw) \right) = b_{k'} \otimes h'hw.$$

On the other hand, by (9.10.3) and (9.10.4), $gb_k = b_kh'b_k^{1}$ and $g' = b_kh'b_k^{1}$, so

$$(g'g) \cdot (b_i \otimes w) = (b_kh'b_k^{1}) \cdot (b_i \otimes w) = b_{k'} \otimes h'hw$$

as desired. Note by the way that

$$\dim \text{Ind}(\rho) = \frac{|G|}{|H|} \dim \rho.$$

Now that we know that $\text{Ind}(\rho)$ is a representation of $G$ on $V$, we calculate its character on an arbitrary element $g \in G$. With respect to the basis $\{b_i \otimes e_j\}$, we can write $\text{Ind}(\rho)(g)$ as a $r \times r$ block matrix:

$$\text{Ind}(\rho)(g) = [B_{i,j}]_{i,j \in [r]}, \quad B_{i,j} = \begin{cases} \rho(b_k^{-1}gb_i) & \text{if } b_k^{-1}gb_i \in H, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, the induced character is

$$\text{Ind}(\chi)(g) = \sum_{i \in [r]: b_i^{-1}gb_i \in H} \chi(b_i^{-1}gb_i)$$

$$= \sum_{i \in [r]: b_i^{-1}gb_i \in H} \frac{1}{|H|} \sum_{h \in H} \chi(h^{-1}b_i^{-1}gb_i h)$$

because $\chi_\rho$ is constant on conjugacy classes of $H$. Setting $k = b_i h$, we obtain

$$\text{Ind}^G_H(\chi)(g) = \frac{1}{|H|} \sum_{k \in G: k^{-1}gk \in H} \chi(k^{-1}gk).$$

Note that $k$ runs over all elements of $G$ as $i$ and $h$ run over $[r]$ and $H$ respectively. Also, $k^{-1}gk = h^{-1}b_i^{-1}gb_i h \in H$ if and only if $b_i^{-1}gb_i \in H$, simply because $H$ is a group. Since $\text{Ind}(\chi)$ is independent of the choice of $B$, so is the isomorphism type of $\text{Ind}(\rho)$. \qed
Corollary 9.10.5. Suppose $H$ is a normal subgroup of $G$. Then
\[
\text{Ind}_H^G(\chi)(g) = \begin{cases} 
\frac{|G|}{|H|} \chi(g) & \text{if } g \in H, \\
0 & \text{otherwise.}
\end{cases}
\]

Proof. Normality implies that $k^{-1}gk \in H$ if and only if $g \in H$, independent of $K$. Alternatively, normality implies that left cosets and right cosets coincide, so the blocks in the block matrix $\text{Ind}(\rho)(g)$ will all be on the main diagonal (and equal to $\rho(g)$) when $g \in H$, and off the main diagonal otherwise. \hfill \square

Corollary 9.10.6. Let $H \subseteq G$ and let $\rho$ be the trivial representation of $H$. Then
\[
\text{Ind}_H^G(\chi_{\text{triv}})(g) = \frac{\#\{k \in G : k^{-1}gk \in H\}}{|H|}.
\] (9.10.5)

Example 9.10.7. The character $\psi = \text{Ind}_{\mathfrak{S}_3}^{\mathfrak{A}_3}(\chi_{\text{triv}})$ is defined by $\psi(g) = 2$ for $g \in \mathfrak{A}_3$, $\psi(g) = 0$ for $g \notin \mathfrak{A}_3$. Thus $\psi = \chi_{\text{triv}} + \chi_{\text{sign}}$.

Example 9.10.8. Let $G = \mathfrak{S}_4$ and let $H$ be the subgroup \{id, (1 2), (3 4), (1 2)(3 4)\}. Note that $H$ is not a normal subgroup of $G$. Let $\rho$ be the trivial representation of $G$ and $\chi$ its character. We can calculate $\psi = \text{Ind}_H^G \chi$ using (9.10.5)
\[
\psi(C_{1111}) = 6, \quad \psi(C_{211}) = 2, \quad \psi(C_{22}) = 2, \quad \psi(C_{31}) = 0, \quad \psi(C_4) = 0.
\]
where, as usual, $C_\lambda$ denotes the conjugacy class in $\mathfrak{S}_4$ of permutations with cycle-shape $\lambda$. In the notation of Example 9.8.2, the decomposition into irreducible characters is $\chi_1 + \chi_2 + 2\chi_5$.

Restriction and induction are related by the following useful formula.

Theorem 9.10.9 (Frobenius Reciprocity). Let $H \subseteq G$ be groups, let $\chi$ be a character of $H$, and let $\psi$ be a character of $G$. Then
\[
\langle \text{Ind}_H^G(\chi), \psi \rangle_G = \langle \chi, \text{Res}_H^G(\psi) \rangle_H.
\]

Proof.
\[
\langle \text{Ind}(\chi), \psi \rangle_G = \frac{1}{|G|} \sum_{g \in G} \text{Ind}(\chi)(g) \cdot \psi(g)
\]
\[
= \frac{1}{|G|} \sum_{g \in G} \frac{1}{|H|} \sum_{k \in G, k^{-1}gk \in H} \chi(k^{-1}gk) \cdot \psi(g)
\]
(by Prop. 9.10.4)
\[
= \frac{1}{|G||H|} \sum_{h \in H} \sum_{g \in G} \sum_{k \in G, k^{-1}gk = h} \chi(h) \cdot \psi(k^{-1}gk)
\]
\[
= \frac{1}{|G||H|} \sum_{h \in H} \sum_{k \in G} \chi(h) \cdot \psi(h)
\]
(i.e., $g = khk^{-1}$)
\[
= \frac{1}{|H|} \sum_{h \in H} \chi(h) \cdot \psi(h) = \langle \chi, \text{Res}(\psi) \rangle_H.
\]

This will be useful later; for now, here is a quick application.
Example 9.10.10. Sometimes, Frobenius reciprocity suffices to calculate the isomorphism type of an induced representation. Let $G = S_3$ and $H = A_3$, and let $\psi, \chi_1$ and $\chi_2$ be as in Example 9.10.2. We would like to compute $\text{Ind}(\chi_1)$. By Frobenius reciprocity

$$\langle \text{Ind}(\chi_1), \psi \rangle_G = \langle \chi_1, \text{Res}(\psi) \rangle_H = 1.$$ 

But $\psi$ is irreducible. Therefore, it must be the case that $\text{Ind}(\chi_1) = \psi$, and the corresponding representations are isomorphic. The same is true if we replace $\chi_1$ with $\chi_2$.

9.11 Characters of the symmetric group

We have worked out the irreducible characters of $S_3$, $S_4$ and $S_5$ ad hoc (the last as an exercise). In fact, we can do so in general, exploiting a vast connection to the combinatorics of partitions and tableaux.

Recall that a partition of $n$ is a sequence $\lambda = (\lambda_1, \ldots, \lambda_\ell)$ of weakly decreasing positive integers whose sum is $n$. We write $\lambda \vdash n$ or $|\lambda| = n$ to indicate that $\lambda$ is a partition of $n$. The number $\ell = \ell(\lambda)$ is the length of $\lambda$. The set of all partitions of $n$ is $\text{Par}(n)$, and the number of partitions of $n$ is $p(n) = |\text{Par}(n)|$.

For each $\lambda \vdash n$, let $C_\lambda$ be the conjugacy class in $S_n$ consisting of all permutations with cycle shape $\lambda$. Since the conjugacy classes are in bijection with $\text{Par}(n)$, it makes sense to look for a set of representations indexed by partitions.

Definition 9.11.1. Let $\mu = (\mu_1, \ldots, \mu_\ell) \vdash n$.

- The Ferrers diagram of shape $\mu$ is the top- and left-justified array of boxes with $\mu_i$ boxes in the $i^{th}$ row.
- A (Young) tableau$^4$ of shape $\mu$ 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 $\mu$ 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 $\mu$ 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 $\mu$.

\[
\begin{array}{ccc}
\begin{array}{ccc}
& & \\
& & \\
& & \\
\end{array}
&& \begin{array}{ccc}
1 & 3 & 6 \\
2 & 7 \\
4 & 5
\end{array}
& & \begin{array}{ccc}
1 & 3 & 6 \\
2 & 7 \\
4 & 5
\end{array}
\end{array}
\]

Ferrers diagram Young tableau Young tabloid

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 $\mu$ is

$$\binom{n}{\mu} = \frac{n!}{\mu_1! \cdots \mu_\ell!}.$$

$^4$Terminology of tableaux is not consistent: some authors reserve the term “Young tableau” for a tableau in which the numbers increase downward and leftward. 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: there are $n!$ tableaux of shape $\mu$ for any $\mu \vdash n$. 

143
The symmetric group $S_n$ acts on tabloids by permuting the numbers. This action gives rise to a permutation representation $(\rho_\mu, V_\mu)$ of $S_n$, the $\mu$-tabloid representation of $S_n$. Here, as usual, $V_\mu$ is the vector space of all formal $\mathbb{C}$-linear combinations of tabloids of shape $\mu$. The character of $\rho_\mu$ is denoted $\chi_\mu$.

**Example 9.11.2.** For $n = 3$, the characters of the tabloid representations $\rho_\mu$ are as follows.

<table>
<thead>
<tr>
<th>Characters</th>
<th>$C_{111}$</th>
<th>$C_{21}$</th>
<th>$C_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\chi_3$</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$\chi_{21}$</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>$\chi_{111}$</td>
<td>6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$</td>
<td>C_\mu</td>
<td>$</td>
<td>1</td>
</tr>
</tbody>
</table>

Many familiar representations of $S_n$ can be expressed in this form.

- There is a unique tabloid of shape $\mu = (n)$. Every permutation fixes it, so $\rho(n) \cong \rho_{\text{triv}}$.
- The tabloids of shape $\mu = (1,1,\ldots,1)$ are just the permutations of $[n]$. Therefore $\rho(1,1,\ldots,1) \cong \rho_{\text{reg}}$.
- A tabloid of shape $\mu = (n-1,1)$ is determined by its singleton part. So the representation $\rho_\mu$ is isomorphic to the action on the singleton by permutation, i.e., $\rho(n-1,1) \cong \rho_{\text{def}}$.

For $n = 3$, the table in (9.11.1) is a triangular matrix. In particular, the characters $\chi_\mu$ are linearly independent, hence a basis, in the vector space of class functions on $S_3$. In fact, we will prove that this is the case for all $n$. We first need to define two orders on the set $\text{Par}(n)$.

**Definition 9.11.3.** Let $\lambda, \mu \in \text{Par}(n)$. **Lexicographic order** on $\text{Par}(n)$ is defined by $\lambda < \mu$ if for some $k > 0$ we have

$$\lambda_1 = \mu_1, \quad \lambda_2 = \mu_2, \quad \ldots, \quad \lambda_k = \mu_k, \quad \lambda_k < \mu_k.$$  

Meanwhile, **dominance order** on $\text{Par}(n)$ is defined as follows: $\mu$ dominates $\lambda$, written $\lambda \triangleleft \mu$, if $\lambda \neq \mu$ and

$$\sum_{i=1}^{k} \lambda_i \leq \sum_{i=1}^{k} \mu_i \quad \forall k.$$  

Lex order is a total order on $\text{Par}(n)$. For instance, if $n = 5$, we have

$$(5) > (4,1) > (3,2) > (3,1,1) > (2,2,1) > (2,1,1,1) > (1,1,1,1,1).$$

(“Lex-greater partitions are short and wide; lex-smaller ones are tall and skinny.”)

Dominance is a partial order on $\text{Par}(n)$. It first fails to be a total order for $n = 6$ (neither of 33 and 411 dominates the other). Lexicographic order is a linear extension of dominance order: if $\lambda \triangleleft \mu$ then $\lambda < \mu$.

Since the tabloid representations $\rho_\mu$ are permutation representations, we can calculate $\chi_\mu$ by counting fixed points. That is, for any permutation $w \in C_\lambda$, we have

$$\chi_\mu(C_\lambda) = \chi_\mu(w) = \# \{\text{tabloids } T \mid \text{sh}(T) = \mu, \text{ } w(T) = T\}. \quad (9.11.2)$$
Proposition 9.11.4. Let $\lambda, \mu \vdash n$. Then:

1. $\chi_\mu(C_\mu) \neq 0$.
2. $\chi_\mu(C_\lambda) \neq 0$ only if $\lambda \leq \mu$ (thus, only if $\lambda \preceq \mu$ in lexicographic 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 $w = (1 3 6)(2 7)(4 5) \in S_7$, then $T$ can be either of the following two tabloids:

\[
\begin{array}{ccc}
1 & 3 & 6 \\
2 & 7 \\
4 & 5 \\
\end{array} \quad \begin{array}{ccc}
1 & 3 & 6 \\
4 & 5 \\
2 & 7 \\
\end{array}
\]

It follows from (9.11.2) that $\chi_\mu(C_\mu) \neq 0$. (In fact $\chi_\mu(C_\mu) = \prod_j r_j!$, where $r_j$ is the number of occurrences of $j$ in $\mu$.)

For the second assertion, observe that $w \in S_n$ fixes a tabloid $T$ of shape $\mu$ 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 $\lambda \preceq \mu$. \hfill $\Box$

Corollary 9.11.5. The characters $\{\chi_\mu \mid \mu \vdash n\}$ form a basis for $\Cl(G)$.

Proof. Make the characters into a $p(n) \times p(n)$ matrix $X = [\chi_\mu(C_\lambda)]_{\mu,\lambda \vdash n}$ with rows and columns ordered by lex order on Par($n$). By Proposition 9.11.4, $X$ is a triangular matrix with nonzero entries on the diagonal, so it is nonsingular. \hfill $\Box$

We can transform the rows of the matrix $X$ into a list of irreducible characters of $S_n$ by applying the Gram-Schmidt process with respect to the inner product $\langle \cdot, \cdot \rangle_{S_n}$. Triangularity of $X$ means that we can label the irreducible characters $\{\sigma_\nu : \nu \vdash n\}$ in such a way that

\[
\begin{align*}
\langle \sigma_\nu, \chi_\mu \rangle_G &\neq 0, \\
\langle \sigma_\nu, \chi_\mu \rangle_G &\neq 0 & \text{if } \nu < \mu.
\end{align*}
\]

(9.11.3)

On the level of representations, this corresponds to decomposing the tabloid representation $V_\mu$ into its irreducible $G$-invariant subspaces $Sp_\lambda$, which are called Specht modules. The multiplicities of the Specht modules in the tabloid representations are called the Kostka numbers $K_{\lambda,\mu}$, which can be computed on the level of characters:

\[
K_{\lambda,\mu} = \langle \sigma_\lambda, \chi_\mu \rangle_G, \quad V_\mu = \bigoplus_\lambda (Sp_\lambda)^{\oplus K_{\lambda,\mu}}, \quad \chi_\mu = \sum_\lambda K_{\lambda,\mu} \sigma_\lambda.
\]

The observations of (9.11.3) imply that $K_{\lambda,\mu} = 0$ unless $\lambda \leq \mu$.

Example 9.11.6. Recall the table of characters (9.11.1) of the tabloid representations for $n = 3$. We will use this to reproduce the table of irreducible characters from Example 9.8.1.

First, $\chi_3 = [1,1,1] = \chi_{\text{triv}}$ is irreducible and is the character $\sigma_3$ of the Specht module $Sp_3$.

Second, for the character $\chi_{21}$, observe that $\langle \chi_{21}, \sigma_3 \rangle_G = 1$. Applying Gram-Schmidt produces a character orthonormal to $\sigma_3$, namely

\[
\sigma_{21} = \chi_{21} - \sigma_3 = [2,0,-1],
\]

145
which is irreducible.

Third, for the character \( \chi_{111} \), we have

\[
\langle \chi_{111}, \sigma_3 \rangle_G = 1, \quad \langle \chi_{111}, \sigma_{21} \rangle_G = 2.
\]

Accordingly, we apply Gram-Schmidt to obtain the character

\[
\sigma_{111} = \chi_{111} - \sigma_3 - 2\sigma_{21} = [1, -1, 1]
\]

which is 1-dimensional, hence irreducible. In summary, the complete list of irreducible characters, labeled so as to satisfy (9.11.3), is as follows:

<table>
<thead>
<tr>
<th>conjugacy class</th>
<th>( \sigma_3 )</th>
<th>( \sigma_{21} )</th>
<th>( \sigma_{111} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>111</td>
<td>1, 1, 1</td>
<td>2, 0, -1</td>
<td>1, -1, 1</td>
</tr>
<tr>
<td>21</td>
<td>( \chi_{\text{triv}} )</td>
<td>( \chi_{\text{def}} - \chi_{\text{triv}} )</td>
<td>( \chi_{\text{sign}} )</td>
</tr>
</tbody>
</table>

To summarize our calculation, we have

\[
\begin{bmatrix}
\chi_3 \\
\chi_{21} \\
\chi_{111}
\end{bmatrix} =
\begin{bmatrix}
1 & 1 & 1 \\
3 & 1 & 0 \\
6 & 0 & 0
\end{bmatrix}
= \begin{bmatrix}
1 & 0 & 0 \\
1 & 1 & 0 \\
1 & 2 & 1
\end{bmatrix}
\begin{bmatrix}
\sigma_3 \\
\sigma_{21} \\
\sigma_{111}
\end{bmatrix}
\]

where \( K \) is the matrix of Kostka numbers. The Kostka numbers will also play a key role in the study of symmetric functions and tableaux, coming in the next chapter.

### 9.12 Exercises

In all exercises, unless otherwise specified, \( G \) is a finite group and \((\rho, V)\) and \((\rho', V')\) are finite-dimensional representations of \( G \) over \( \mathbb{C} \).

**Exercise 9.1.** Let \( \chi \) be an irreducible character of \( G \) and let \( \psi \) be a one-dimensional character. Prove that \( \omega := \chi \otimes \psi \) is an irreducible character.

**Exercise 9.2.** Let \( n \geq 2 \). Prove that the standard representation \( \rho_{\text{std}} \) of \( S_n \) is irreducible. (Hint: Compute \( \langle \chi_{\text{def}}, \chi_{\text{def}} \rangle \), using what is essentially linearity of expectation.)

**Exercise 9.3.** 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.)

**Exercise 9.4.** Consider the permutation action of the symmetric group \( G = S_4 \) on the vertices of the complete graph \( K_4 \), whose corresponding representation is the defining representation \( \rho_{\text{def}} \). Let \( \rho \) be the 3-dimensional representation corresponding to the action of \( S_4 \) on pairs of opposite edges of \( K_4 \).

(a) Compute the character of \( \rho \).
(b) Show that \( \rho \) is the direct sum of the trivial representation and one other irrep.
(c) Explicitly describe all \( G \)-equivariant linear transformations \( \phi : \rho \rightarrow \rho_{\text{def}} \). (Hint: Schur’s lemma should be useful.)
Exercise 9.5. Work out the character table of $S_5$ without using any of the material in Section 9.11. (Hint: To construct another irreducible character, start by considering the action of $S_5$ on the edges of the complete graph $K_5$ induced by the usual permutation action on the vertices.)

Exercise 9.6. Work out the character table of the quaternion group $Q$. (Recall that $Q$ is the group of order 8 whose elements are \{±1, ±i, ±j, ±k\} with relations $i^2 = j^2 = k^2 = -1$, $ij = ki$, $jk = jk$.)

Exercise 9.7. Work out the characters $\sigma_{\lambda}$ of the Specht modules $Sp_{\lambda}$ for all $\lambda \vdash 4$. (Start with the characters of the tabloid representations, then do linear algebra. Feel free to use a computer algebra system to automate the tedious parts.) Compare your result to the character table of $S_4$ calculated ad hoc in Example 9.8.2. Make as many observations or conjectures as you can about how the partition $\lambda$ is related to the values of the character $\sigma_{\lambda}$.

Exercise 9.8. Recall that the alternating group $A_n$ consists of the $n!/2$ even permutations in $S_n$, that is, those with an even number of even-length cycles.

(a) Show that the conjugacy classes in $A_4$ are not simply the conjugacy classes in $S_4$. (Hint: Consider the possibilities for the dimensions of the irreducible characters of $A_4$.)

(b) Determine the conjugacy classes in $A_4$, and the complete list of irreducible characters.

(c) Use this information to determine $[A_4, A_4]$ without actually computing any commutators.
Chapter 10

Symmetric Functions

10.1 Prelude: Symmetric polynomials

Definition 10.1.1. Let $R$ be a commutative ring, 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.

Note that the symmetric polynomials that are homogeneous of degree $d$ form an $R$-module.

For example, if $n = 3$, then up to scalar multiplication, the only symmetric polynomial of degree 1 in $x_1, x_2, x_3$ is $x_1 + x_2 + x_3$.

In degree 2, here are two:

$$x_1^2 + x_2^2 + x_3^2, \quad x_1x_2 + x_1x_3 + x_2x_3.$$  

Every other symmetric polynomial that is homogeneous of degree 2 is a $R$-linear combination of these two, because the coefficients of $x_1^2$ and $x_1x_2$ determine the coefficients of all other monomials. Note that the set of all degree-2 symmetric polynomial forms an $R$-module.

Here is a basis for the space of degree 3 symmetric polynomials:

$$x_1^3 + x_2^3 + x_3^3,$$
$$x_1^2x_2 + x_1x_2^2 + x_1^2x_3 + x_1x_3^2 + x_2^2x_3 + x_2x_3^2,$$
$$x_1x_2x_3.$$

Each member of this basis is a sum of the monomials in a single orbit under the action of $S_3$. Accordingly, we can index them by the partition whose parts are the exponents of one of its monomials. That is,

$$m_3(x_1, x_2, x_3) = x_1^3 + x_2^3 + x_3^3,$$
$$m_{21}(x_1, x_2, x_3) = x_1^2x_2 + x_1x_2^2 + x_1^2x_3 + x_1x_3^2 + x_2^2x_3 + x_2x_3^2,$$
$$m_{111}(x_1, x_2, x_3) = x_1x_2x_3.$$

In general, for $\lambda = (\lambda_1, \ldots, \lambda_\ell)$, we define

$$m_\lambda(x_1, \ldots, x_n) = \sum_{\{a_1, \ldots, a_\ell\} \subseteq [n]} x_1^{\lambda_{a_1}}x_2^{\lambda_{a_2}}\cdots x_n^{\lambda_{a_\ell}}.$$
But unfortunately, this is zero if $\ell > n$. So $n$ variables is not enough! In other words, 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...

## 10.2 Formal power series

Let $R$ be an integral domain (typically $\mathbb{Z}$ or a field), and let $x = \{x_i : i \in I\}$ be a set of commuting indeterminates indexed by $I$. (Typically $I = \mathbb{P}$, but we may as well make the definition general.) A monomial is an object $x^\alpha = \prod_{i \in I} x_i^{\alpha_i}$, where $\alpha = (\alpha_i)_{i \in I} \in \mathbb{N}^I$ and $\sum_{i \in I} \alpha_i$ is finite (equivalently, all but finitely many of the $\alpha_i$ are zero). A formal power series is an object of the form

$$\sum_{\alpha \in \mathbb{N}^I} c_\alpha x^\alpha$$

with $c_\alpha \in R$ for all $\alpha$. The set $R[[x]]$ of all formal power series is easily seen to be an abelian group under addition, and in fact an $R$-module. (Formally, it can be regarded as the set of functions from monomials to $R$, so it is a direct product of countably infinitely many copies of $R$.) In fact, it is a ring as well, with multiplication given by

$$\left( \sum_{\alpha \in \mathbb{N}^I} c_\alpha x^\alpha \right) \left( \sum_{\beta \in \mathbb{N}^I} d_\beta x^\beta \right) = \sum_{\gamma \in \mathbb{N}^I} \left( \sum_{(\alpha, \beta) : \alpha + \beta = \gamma} c_\alpha d_\beta \right) x^\gamma,$$

because the inner sum on the right-hand side has only finitely many terms for each $\gamma$, 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 only rarely need to plug in real values for the indeterminates $x_i$ (and when we do, 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. (But 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.)

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 Exercise 10.3 for an example.

## 10.3 Symmetric functions

We can now define symmetric functions properly, as elements of the ring of formal power series $\mathbb{C}[[x]] = \mathbb{C}[[x_1, x_2, \ldots]]$.

**Definition 10.3.1.** Let $\lambda \vdash n$. The monomial symmetric function $m_\lambda$ is the power series

$$m_\lambda = \sum_{\{a_1, \ldots, a_\ell\} \subseteq \mathbb{P}} x_{a_1}^{\lambda_1} x_{a_2}^{\lambda_2} \cdots x_{a_\ell}^{\lambda_\ell}.$$
That is, $m_\lambda$ is the sum of all monomials whose exponents are the parts of $\lambda$. Equivalently,

$$m_\lambda = \sum_\alpha x^\alpha$$

where the sum ranges over all rearrangements $\alpha$ of $\lambda$ (regarding $\lambda$ as a countably infinite sequence in which all but finitely many terms are 0).

We then define

$$\Lambda_d = \Lambda_{R,d}(x) = \{\text{degree-}d \text{ symmetric functions in indeterminates } x \text{ with coefficients in } R\},$$

$$\Lambda = \Lambda_R(x) = \bigoplus_{d \geq 0} \Lambda_d.$$

Each $\Lambda_d$ is a finitely generated free $R$-module, with basis $\{m_\lambda \mid \lambda \vdash d\}$, and their direct sum $\Lambda$ is a graded $R$-algebra. If we let $S_\infty$ be the group whose members are the permutations of $\{x_1, x_2, \ldots\}$ with only finitely many non-fixed points (equivalently, $S_\infty = \bigcup_{n=1}^{\infty} S_n$), then $\Lambda$ is the ring of formal power series in $\{x_1, x_2, \ldots\}$ of bounded degree that are invariant under the action of $S_\infty$.

The eventual punchline will be that we will construct a graded isomorphism

$$\Lambda \xrightarrow{E} \bigoplus_{n \geq 0} C\ell(S_n)$$

called the Frobenius characteristic. Thus will allow us to translate symmetric function identities into statements about representations and characters of $S_n$, and vice versa. Many of these statements are best stated in terms of bases for $\Lambda$ other than the monomial symmetric functions, so we now consider several important families.

### 10.4 Elementary symmetric functions

For $k \in \mathbb{N}$ we define

$$e_k = \sum_{S \subseteq \mathbb{P} \atop |S|=k} \prod_{s \in S} x_s = \sum_{0 < i_1 < i_2 < \cdots < i_k} x_{i_1} x_{i_2} \cdots x_{i_k} = m_{11 \ldots k}$$

where there are $k$ 1’s in the last expression. (In particular $e_0 = 1$.) We then define

$$e_\lambda = e_{\lambda_1} \cdots e_{\lambda_\ell}.$$

For example, $e_{11} = (x_1 + x_2 + x_3 + \cdots)^2 = (x_1^2 + x_2^2 + \cdots) + 2(x_1 x_2 + x_1 x_3 + x_2 x_3 + x_1 x_4 + \cdots) = m_2 + 2m_{11}$. In degree 3, we have

$$e_3 = \sum_{i < j < k} x_i x_j x_k = m_{111},$$

$$e_{21} = (x_1 + x_2 + x_3 + \cdots)(x_1 x_2 + x_1 x_3 + x_2 x_3 + x_1 x_4 + \cdots) = m_{21} + 3m_{111},$$

$$e_{111} = (x_1 + x_2 + x_3 + \cdots)^3 = m_3 + 3m_{21} + 6m_{111}.$$
We can conveniently express all the $e$’s together as a generating function. Observe that

$$E(t) := \prod_{i \geq 1} (1 + tx_i) = \sum_{k \geq 0} t^k e_k$$

(10.4.1)

by expanding $E(t)$ as a power series in $t$ whose coefficients are power series in $\{x_i\}$. Note that there are no issues of convergence: we are working in the ring of formal power series $R[[t, x_1, x_2, \ldots]]$.

Recall the dominance partial order $\unlhd$ on partitions from Definition 9.11.3.

**Theorem 10.4.1.** Let $b_{\lambda, \mu}$ be the coefficient of $e_{\lambda}$ 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 $\tilde{\lambda} \unlhd \mu$. In particular $\{e_{\lambda} \mid \lambda \vdash n\}$ is an $R$-module basis for $\Lambda_n$.

**Proof.** It suffices to figure out the coefficient of the monomial $x^\mu = \prod_i x_i^{\mu_i}$ in the expansion of $e_{\lambda}$. That coefficient will be the number of ways to factorize $x^\mu$ into squarefree pieces whose degrees are the parts of $\lambda$.

I claim that $b_{\lambda, \mu}$ is the number of ways to put balls in boxes as follows. Suppose that we have $\lambda_1$ balls labeled 1, $\lambda_2$ balls labeled 2, etc. We also have countably infinitely many boxes. Place the balls in the boxes so that box $i$ receives $\mu_i$ balls, all with different labels.

For example, let $\mu = (3, 2, 2)$ and $\lambda = (3, 2, 1, 1)$, so that we are trying to count the ways of factoring $x^\mu = x_1^3 x_2^2 x_3^2$ into four squarefree monomials of degrees 3, 2, 1, 1. One such factorization is

$$x_1^3 x_2^2 x_3^2 = x_1 x_2 x_3 \cdot x_1 x_3 \cdot x_2 \cdot x_1$$

which corresponds to the balls-and-boxes picture

<table>
<thead>
<tr>
<th>Box 1</th>
<th>Box 2</th>
<th>Box 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 4</td>
<td>1 3</td>
<td>1 2</td>
</tr>
</tbody>
</table>

where a ball labeled $j$ in box $i$ indicates that the $j^{th}$ factor included $x_i$.

Now observe that the first $k$ boxes collectively can contain no more than $k$ balls with each label. That is,

$$\mu_1 + \cdots + \mu_k \leq \tilde{\lambda}_1 + \cdots + \tilde{\lambda}_k \quad (\forall k)$$

which is precisely the statement that $b_{\lambda, \mu} = 0$ unless $\tilde{\lambda} \unlhd \mu$. Moreover, if $\tilde{\lambda} = \mu$ then there is exactly one way to do this: place balls labeled 1, \ldots, $\tilde{\lambda}_i$ in the $i^{th}$ box. Hence $b_{\lambda, \tilde{\lambda}} = 1$.

Therefore, if we order partitions of $n$ by any linear extension of dominance (such as the lexicographic order), then the matrix $[b_{\lambda, \mu}]$ will be seen to be upper unitriangular. In particular it is invertible, and it follows that the $R$-module spanned by the $e_\lambda$’s is the same as that spanned by the $m_\mu$’s, namely $\Lambda_{R,n}$.

**Corollary 10.4.2** ("Fundamental Theorem of Symmetric Functions"). The elementary symmetric functions $e_1, e_2, \ldots$ are algebraically independent. Therefore, $\Lambda = R[e_1, e_2, \ldots]$ 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 $e_\lambda$’s, which does not exist.
10.5 Complete homogeneous symmetric functions.

For \( k \in \mathbb{N} \), we define \( h_k \) to be the sum of all monomials of degree \( k \):

\[
    h_k = \sum_{0 < i_1 \leq i_2 \leq \cdots \leq i_k} x_{i_1}x_{i_2} \cdots x_{i_k} = \sum_{\lambda \vdash k} m_{\lambda}.
\]

(So \( h_0 = 1 \).) We then define \( h_\lambda = h_{\lambda_1} \cdots h_{\lambda_\ell} \).

For example, \( h_{11} = e_{11} \) and \( h_2 = m_{11} + m_2 \). In degree 3, we have

\[
    h_{111} = m_1^3 = (x_1 + x_2 + x_3 + \cdots)^3 = 6m_{111} + 3m_{21} + m_3, \\
    h_{21} = h_1h_2 = e_1(m_{11} + m_2) = e_1(e_{11} - e_2) = e_{11} - e_{21} = m_3 + 2m_{21} + 3m_{11}, \\
    h_3 = m_{111} + m_{21} + m_3.
\]

The analogue of (10.4.1) for the homogeneous symmetric functions is

\[
    H(t) := \prod_{i \geq 1} \frac{1}{1 - tx_i} = \sum_{k \geq 0} t^k h_k
\]

because each factor in the infinite product is a geometric series \( 1 + tx_i + t^2x_i^2 + \cdots \), so when we expand and collect like powers of \( t \), the coefficient of \( t^k \) will be the sum of all possible ways to build a monomial of degree \( k \). It is immediate from the algebra that

\[
    H(t)E(-t) = 1
\]

as formal power series. Extracting the coefficients of positive powers of \( t \) gives the Jacobi-Trudi relations:

\[
    \sum_{k=0}^{n} (-1)^k e_k h_{n-k} = 0
\]

for all \( n > 0 \), where we put \( e_0 = h_0 = 1 \). Explicitly,

\[
    h_1 - e_1 = 0, \quad h_2 - e_1 h_1 + e_2 = 0, \quad h_3 - e_1 h_2 + e_2 h_1 - e_3 = 0, \cdots
\]

These equations can be used (iteratively) to solve for the \( h_k \) as polynomials in the \( e_k \):

\[
    h_1 = e_1, \\
    h_2 = e_1 h_1 - e_2 = e_1^2 - e_2, \\
    h_3 = e_1 h_2 - e_2 h_1 + e_3 = e_1(e_1^2 - e_2) - e_2 e_1 + e_3,
\]

etc. So again \( \{ h_\lambda \mid \lambda \vdash n \} \) is an \( R \)-module basis for \( \Lambda_n \), and \( h_1, h_2, \ldots \) generate \( \Lambda \) as a \( \mathbb{C} \)-algebra.

Here is another way to see that the \( h \)'s are a \( R \)-basis. Define a ring endomorphism \( \omega : \Lambda \to \Lambda \) by

\[
    \omega(e_i) = h_i
\]

for all \( i \), so that \( \omega(e_\lambda) = h_\lambda \). This is well-defined since the elementary symmetric functions are algebraically independent (recall that \( \Lambda \cong R[e_1, e_2, \ldots] \)).

**Proposition 10.5.1.** \( \omega(\omega(f)) = f \) for all \( f \in \Lambda \). In particular, the map \( \omega \) is a ring automorphism.
**Proof.** Applying $\omega$ to the Jacobi-Trudi relations \((10.5.2)\), we see that for every $n \geq 1$,

$$0 = \sum_{k=0}^{n} (-1)^{n-k} \omega(e_k) \omega(h_{n-k})$$

$$= \sum_{k=0}^{n} (-1)^{n-k} h_k \omega(h_{n-k})$$

$$= \sum_{k=0}^{n} (-1)^{k} h_{n-k} \omega(h_k)$$

$$= (-1)^{n} \sum_{k=0}^{n} (-1)^{n-k} h_{n-k} \omega(h_k)$$

and comparing this last expression with the original Jacobi-Trudi relations gives $\omega(h_k) = e_k$. 

\[\square\]

**Corollary 10.5.2.** \(\{h_\lambda\}\) is a graded $\mathbb{Z}$-basis for $\Lambda_\mathbb{Z}$. Moreover, $\Lambda_\mathbb{R} \cong R[h_1, h_2, \ldots]$ as rings.

### 10.6 Power-sum symmetric functions

Define

$$p_k = x_1^k + x_2^k + \cdots = m_k,$$

$$p_\lambda = p_{\lambda_1} \cdots p_{\lambda_\ell}.$$  

For example, in degree 2,

$$p_2 = m_2,$$

$$p_{11} = (x_1 + x_2 + \cdots)^2 = m_2 + 2m_{11}.$$  

While the $p_\lambda$ are a vector space basis for $\Lambda_\mathbb{Q}$ (the proof is left as an exercise), they are not a $\mathbb{Z}$-module basis for $\Lambda_\mathbb{Z}$. In other words, not every symmetric function with integer coefficients can be expressed as an integer combination of the power-sums; for example, $m_{11} = (p_{11} - p_2)/2$.

Expanding the power-sum symmetric functions in the monomial basis will provide the first example of the deep connection between representation theory and symmetric functions.

**Proposition 10.6.1.** For $\lambda \vdash n$, we have

$$p_\lambda = \sum_{\mu \vdash n} \chi_\mu(C_\lambda)m_\mu$$

where $\chi_\mu(C_\lambda)$ means the character of the tabloid representation of shape $\mu$ on the conjugacy class $C_\lambda$ of cycle-shape $\lambda$ (see Section 9.11).

The proof is left as an exercise (Exercise 10.2).

The elementary, complete homogeneous, and power-sum symmetric functions are **multiplicative** in the sense that

$$a_\lambda = \prod_{i=1}^{\ell(\lambda)} a_{\lambda_i}, \quad \forall a \in \{e, h, p\}.$$  

The monomial symmetric functions are not multiplicative, nor are the family we are about to describe.
10.7 Schur functions

The definition of these symmetric power series is very different from the preceding ones. While it looks weird at first, 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 $\Lambda$.

**Definition 10.7.1.** A column-strict tableau $T$ of shape $\lambda$, or $\lambda$-CST for short, is a labeling of the boxes of the Ferrers diagram of $\lambda$ with integers (not necessarily distinct) that is

- weakly increasing across every row; and
- strictly increasing down every column.

The partition $\lambda$ is called the shape of $T$, and the set of all column-strict tableaux of shape $\lambda$ is denoted $\text{CST}(\lambda)$. The content of a CST is the sequence $\alpha = (\alpha_1, \alpha_2, \ldots)$, where $\alpha_i$ is the number of boxes labelled $i$, and the weight of $T$ is the monomial $x^T = x^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \cdots$. For example:

\[
\begin{array}{ccc}
1 & 1 & 3 \\
2 & 3 \\
\end{array}
\quad \quad
\begin{array}{ccc}
1 & 1 & 1 \\
4 & 8 \\
\end{array}
\quad \quad
\begin{array}{ccc}
1 & 2 & 3 \\
1 & 4 \\
\end{array}
\]

\[x_1^2 x_2 x_3^2 \quad x_1^3 x_4 x_8 \quad \text{Not a CST}\]

**Definition 10.7.2.** The Schur function corresponding to a partition $\lambda$ is

\[s_\lambda = \sum_{T \in \text{CST}(\lambda)} x^T.\]

It is far from obvious that $s_\lambda$ is symmetric, but in fact it is! We will prove this shortly.

**Example 10.7.3.** 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

\[s_{(n)} = h_n = \sum_{\lambda \vdash n} m_\lambda. \quad (10.7.1)\]

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 $n$ distinct labels, which can be arbitrary. Therefore

\[s_{(1,1,\ldots,1)} = e_n = m_{(1,1,\ldots,1)}. \quad (10.7.2)\]

As another simple example, let $\lambda = (2, 1)$. We will express $s_\lambda$ as a sum of the monomial symmetric functions $m_3, m_{21}, m_{111}$.

First, no tableau of shape $\lambda$ 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.

\[
\begin{array}{c}
a \\
b \\
c
\end{array}
\quad \quad
\begin{array}{c}
a \\
b \\
c
\end{array}
\]

154
Therefore, the coefficient of $m_{111}$ is 2.

Third, for every $a \neq b \in \mathbb{P}$, there is one tableau of shape $\lambda$ and weight $x_a^2x_b$: the one on the left if $a < b$, or the one on the right if $a > b$.

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>a</td>
</tr>
</tbody>
</table>

Therefore, $s_{(2,1)} = 2m_{111} + m_{21}$.

**Proposition 10.7.4.** $s_\lambda$ is a symmetric function for all $\lambda$.

**Proof.** Let $\alpha$ be a monomial of total degree $|\lambda|$. First, observe that the number of standard tableaux of shape $\lambda$ and weight $\alpha$ depends only on the ordered sequence of nonzero exponents of $\alpha$.\(^1\) For instance, for any $\lambda \vdash 8$, there are the same number of $\lambda$-CST’s with weights

$$x_1^1x_2^3x_4^1x_9^1 \quad \text{and} \quad x_2^1x_3^2x_5^1x_7^1$$

because there is a bijection between them given by changing the labels 1, 3, 4, 9 to 2, 3, 5, 7, respectively, or vice versa.

To complete the proof that $s_\lambda$ is symmetric, it suffices to show that for every $k$, swapping the powers of $x_k$ and $x_{k+1}$ in a weight monomial $\alpha$ does not change the number of standard tableaux of shape $\lambda$ and weight $\alpha$. That will imply that $s_\lambda$ is invariant under the transposition $(k \ k+1)$, and all these transpositions together generate the group $S_\infty$.

We will prove this by a bijection, which is easiest to show by example. Let $\lambda = (9, 7, 4, 3, 2)$. We would like to show that there are the same number of $\lambda$-CST’s with weights

$$x_1^3x_2^2x_3^3x_4^3x_5^4x_6^4x_7^4 \quad \text{and} \quad x_1^3x_2^2x_3^3x_4^3x_5^4x_6^4x_7^4$$

Let $T$ be the following $\lambda$-CST:

<p>| | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>5</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Observe that the occurrences of 5 and of 6 each form “snakes” from southwest to northeast.

\(^1\)This is precisely the statement that $s_\lambda$ is a **quasisymmetric** function.
Now we will construct a new tableau 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 (like column 3 in this tableau). Consider each group of relevant 5’s and 6’s in the same row as a single block.

Now swap the numbers of 5’s and 6’s in each block. (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.)

This construction

- preserves column-strictness;
- is an involution (because the ignored columns do not change);
- preserves the exponents on all other variables;
- swaps the exponents on $x_k$ and $x_{k+1}$,

as desired.

**Theorem 10.7.5.** The Schur functions $\{s_\lambda \mid \lambda \vdash n\}$ are a $\mathbb{Z}$-basis for $\Lambda_\mathbb{Z}$.

**Proof.** By the definition of Schur functions, we have for every $\lambda$

$$s_\lambda = \sum_{\mu \vdash n} K_{\lambda\mu} m_\mu$$

where $K_{\lambda\mu}$ is the number of column-strict tableaux $T$ with shape $\lambda$ and content $\mu$. The $K_{\lambda\mu}$ are called Kostka numbers.

First, suppose that $\lambda = \mu$. Then there is exactly one possibility for $T$: for each $i$, fill the $i^{th}$ row full of $i$’s. Therefore

$$\forall \lambda \vdash n : \quad K_{\lambda\lambda} = 1. \quad (10.7.3)$$

Second, observe that if $T$ is a CST of shape $\lambda$ and content $\mu$ (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 $i$ in $T$ must appear in one of the first $i$ rows;
and therefore $\sum_{i=1}^{k} \mu_i \leq \sum_{i=1}^{k} \lambda_i$, which is just the statement that $\lambda \succeq \mu$ (recall the definition of dominance order, Defn. 9.11.3). But that means that 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 10.4.1.

10.8 The Cauchy kernel and the Hall inner product

The next step in studying the ring of symmetric functions $\Lambda$ will be to define an inner product structure on it. These will come from considering the Cauchy kernel and the dual Cauchy kernel, which are the formal power series

$$\Omega = \prod_{i,j \geq 1} (1 - x_i y_j)^{-1}, \quad \Omega^* = \prod_{i,j \geq 1} (1 + x_i y_j).$$

These series are symmetric with respect to each of the variable sets $x = \{x_1, x_2, \ldots\}$ and $y = \{y_1, y_2, \ldots\}$. As we will see, the Cauchy kernel can be expanded in many different ways in terms of symmetric functions in the variable sets $x$ and $y$.

For a partition $\lambda \vdash n$, let $^3 r_i$ be the number of $i$’s in $\lambda$. Define

$$z_\lambda = r_1! r_2! \cdots r_n! \quad \text{and} \quad \varepsilon_\lambda = (-1)^{r_2 + r_4 + \cdots}.$$  \hfill (10.8.1)

For example, if $\lambda = (3, 3, 2, 1, 1)$ then $z_\lambda = (1^3 3!)(2^1 1!)(3^2 2!) = 216$ and $\varepsilon_\lambda = -1$. The number $\varepsilon_\lambda$ is the sign of each permutation in the conjugacy class $C_\lambda$, and the size of that conjugacy class is $|C_\lambda| = n! / z_\lambda$ (proof left as an exercise).

**Proposition 10.8.1.** We have

$$\Omega = \sum_{\lambda} h_\lambda(x) m_\lambda(y) = \sum_{\lambda} \frac{p_\lambda(x)p_\lambda(y)}{z_\lambda} \quad \text{and} \quad (10.8.2)$$

$$\Omega^* = \sum_{\lambda} \varepsilon_\lambda(x) m_\lambda(y) = \sum_{\lambda} \frac{\varepsilon_\lambda(x)p_\lambda(y)}{z_\lambda} \quad (10.8.3)$$

where the sums run over all partitions $\lambda$.

---

2Symmetry in each set of variables ought to say that $\Omega$ and $\Omega^*$ belong to the tensor product $\Lambda(x) \otimes \Lambda(y)$. Technically, they really live in the completion of that ring, since the degrees of terms in $\Omega$ and $\Omega^*$ are unbounded. If you are not familiar with the algebraic notion of completion, don’t worry about this.

3In [Sta99], Stanley uses $m_i$ where I use $r_i$. I want to avoid conflict with the notation for monomial symmetric functions.
Proof. For the first equality in (10.8.2), we expand the Cauchy kernel as follows:

\[
\prod_{i,j \geq 1} (1-x_iy_j)^{-1} = \prod_{j \geq 1} \left( \prod_{i \geq 1} (1-x_it)^{-1} \bigg|_{t=y_j} \right)
\]

\[
= \prod_{j \geq 1} \left( \sum_{k \geq 0} h_k(x)t^k \bigg|_{t=y_j} \right) \quad \text{(using (10.5.1))}
\]

\[
= \prod_{j \geq 1} \sum_{k \geq 0} h_k(x)y_j^k
\]

\[
= \sum \, y_1^{r_1} \cdots y_\ell^{r_\ell} h_{r_1}(x) \cdots h_{r_\ell}(x)
\]

\[
= \sum_{\lambda} h_\lambda(x)m_\lambda(y)
\]

as desired.

For the second equality in (10.8.2), recall the standard power series expansions

\[
\log(1 + q) = \sum_{n \geq 1} (-1)^{n+1} \frac{q^n}{n}, \quad \exp(q) = \sum_{n \geq 0} \frac{q^n}{n!}.
\]

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 \(\exp \log(q) = q\). In particular,

\[
\log \Omega = \log \prod_{i,j \geq 1} (1-x_iy_j)^{-1} = -\log \prod_{i,j \geq 1} (1-x_iy_j) = -\sum_{i,j \geq 1} \log(1 + (-x_iy_j))
\]

\[
= \sum_{i,j \geq 1} \sum_{n \geq 1} \frac{x_i^n y_j^n}{n} = \sum_{n \geq 1} \frac{1}{n} \left( \sum_{i \geq 1} x_i^n \right) \left( \sum_{j \geq 1} y_j^n \right)
\]

\[
= \sum_{n \geq 1} \frac{p_n(x)p_n(y)}{n}
\]

and now exponentiating both sides and applying the power series expansion for \(\exp\), we get

\[
\Omega = \exp \left( \sum_{n \geq 1} \frac{p_n(x)p_n(y)}{n} \right) = \sum_{k \geq 0} \frac{1}{k!} \left( \sum_{n \geq 1} \frac{p_n(x)p_n(y)}{n} \right)^k
\]

\[
= \sum_{k \geq 0} \frac{1}{k!} \left[ \sum_{\lambda: \ell(\lambda) = k} \left( \sum_{r_1, r_2} \cdots \right) \left( \frac{p_1(x)p_1(y)}{1} \right)^{r_1(\lambda)} \left( \frac{p_2(x)p_2(y)}{2} \right)^{r_2(\lambda)} \cdots \right]
\]

\[
= \sum_{\lambda} \prod_{i=1}^\infty \left( \prod_{r_i} \frac{p_i(x)p_i(y)}{r_i!} \right)^{r_i(\lambda)}
\]

The proofs of the identities for the dual Cauchy kernel are analogous, and are left as an exercise.

**Corollary 10.8.2.** 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 \( \omega \) is the involution of 10.5.4).

Proof. (1) Start with the identity of (10.8.2):
\[
\sum_{\lambda} h_{\lambda}(x)m_{\lambda}(y) = \sum_{\lambda} \frac{p_{\lambda}(x)p_{\lambda}(y)}{z_{\lambda}}.
\]
Set \( y_1 = t \), and \( y_k = 0 \) for all \( k > 1 \). This kills all terms on the left side for which \( \lambda \) has more than one part, so we get
\[
\sum_{\lambda=(n)} h_{\lambda}(x)t^n = \sum_{\lambda} \frac{p_{\lambda}(x)t^{\lambda}}{z_{\lambda}}
\]
and extracting the coefficient of \( t^n \) gives the desired expression for \( h_n \).

(2) Start with (10.8.3) and do the same thing.

(3) Let \( \omega \) act on symmetric functions in \( x \) while fixing those in \( y \). Using (10.8.2) and (10.8.3), we obtain
\[
\sum_{\lambda} \frac{p_{\lambda}(x)p_{\lambda}(y)}{z_{\lambda}} = \omega \left( \sum_{\lambda} \frac{p_{\lambda}(x)m_{\lambda}(y)}{z_{\lambda}} \right) = \omega \left( \sum_{\lambda} \varepsilon_{\lambda} \frac{p_{\lambda}(x)p_{\lambda}(y)}{z_{\lambda}} \right)
\]
and equating the red coefficients of \( p_{\lambda}(y)/z_{\lambda} \) yields the desired result. \( \square \)

Definition 10.8.3. The Hall inner product on symmetric functions is defined by
\[
\langle h_{\lambda}, m_{\mu} \rangle_{\Lambda} = \delta_{\lambda\mu}
\]
and extending by linearity to all of \( \Lambda \).

By linear algebra, if \( \{u_{\lambda}\} \) and \( \{v_{\mu}\} \) are graded bases for \( \Lambda \) indexed by partitions such that \( \Omega = \sum_{\lambda} u_{\lambda}(x)v_{\lambda}(y) \), then they are orthogonal to each other with respect to the Hall inner product; i.e., \( \langle u_{\lambda}, v_{\mu} \rangle = \delta_{\lambda\mu} \). Thus \( \{p_{\lambda}\} \) and \( \{p_{\lambda}/z_{\lambda}\} \) are orthogonal, and \( \{p_{\lambda}/\sqrt{z_{\lambda}}\} \) is an orthonormal basis for \( \Lambda_R \). (This, by the way, shows that \( \langle \cdot, \cdot \rangle \) is a genuine inner product in the sense of being a nondegenerate bilinear form.)

The involution \( \omega \) is an isometry with respect to the Hall inner product, i.e.,
\[
\langle a, b \rangle = \langle \omega(a), \omega(b) \rangle.
\]

The easiest way to see this is in terms of the power-sum basis: by (3) of Corollary 10.8.2, we have
\[
\langle \omega p_{\lambda}, \omega p_{\mu} \rangle = \langle \varepsilon_{\lambda} p_{\lambda}, \varepsilon_{\lambda} p_{\mu} \rangle = \varepsilon_{\lambda}^2 \langle p_{\lambda}, p_{\mu} \rangle = \langle p_{\lambda}, p_{\mu} \rangle
\]
because \( \varepsilon_{\lambda} \in \{1, -1\} \) for all \( \lambda \).

The basis \( \{p_{\lambda}/\sqrt{z_{\lambda}}\} \) is orthonormal, but it is not at all nice from a combinatorial point of view, because it involves irrational coefficients. There is a much better orthonormal basis, namely the Schur functions. The next goal will be to prove that
\[
\Omega = \prod_{i,j \geq 1} \frac{1}{1 - x_i y_j} = \sum_{\lambda} s_{\lambda}(x)s_{\lambda}(y).
\]

(10.8.5)
The proof requires a marvelous bijection called the RSK correspondence (for Robinson, Schensted and Knuth).
10.9 The RSK Correspondence

**Definition 10.9.1.** Let $\lambda \vdash n$. A standard [Young] tableau of shape $\lambda$ is a filling of the Ferrers diagram of $\lambda$ with the numbers $1, 2, \ldots, n$ that is increasing left-to-right and top-to-bottom. We write $\text{SYT}(\lambda)$ for the set of all standard tableaux of shape $\lambda$, and set $f^\lambda = |\text{SYT}(\lambda)|$.

For example, if $\lambda = (3, 3)$, then $f^\lambda = 5$; the members of $\text{SYT}(\lambda)$ are as follows:

\[
\begin{array}{ccc}
1 & 3 & 5 \\
2 & 4 & 6
\end{array}
\quad
\begin{array}{ccc}
1 & 3 & 4 \\
2 & 5 & 6
\end{array}
\quad
\begin{array}{ccc}
1 & 2 & 5 \\
3 & 4 & 6
\end{array}
\quad
\begin{array}{ccc}
1 & 2 & 4 \\
3 & 5 & 6
\end{array}
\quad
\begin{array}{ccc}
1 & 2 & 3 \\
4 & 5 & 6
\end{array}
\]

Each Young tableau of shape $\lambda$ corresponds to a saturated chain in the interval $[\emptyset, \lambda]$ of Young’s lattice, namely

$$\emptyset = \lambda_{(0)} \subseteq \lambda_{(1)} \subseteq \cdots \subseteq \lambda_{(n)} = \lambda$$

where $\lambda_{(k)}$ denotes the subtableau consisting only of the boxes filled with the numbers $1, \ldots, k$. This correspondence between Young tableaux and saturated chains in $[\emptyset, \lambda]$ is a bijection, and is of fundamental importance.

The RSK correspondence (for Robinson-Schensted-Knuth) constructs, for every permutation $w \in S_n$, a pair $\text{RSK}(w) = (P, Q)$ of standard tableaux of the same shape $\lambda \vdash n$, using the following row-insertion operation defined as follows.

**Definition 10.9.2.** Let $T$ be a column-strict tableau and let $x \in P$. The row-insertion $T \leftarrow x$ is defined as follows:

- If $T = \emptyset$, then $T \leftarrow x = [x]$.
- 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. (For short, “$x$ bumps $u$.”)
- Repeat until the bumping stops.

Got that? Now, for $w = w_1w_2 \cdots w_n \in S_n$, let $P$ be the tableau $((\emptyset \leftarrow w_1) \leftarrow w_2) \leftarrow \cdots \leftarrow w_n \in S_n$. Let $Q$ be the standard tableau of the same shape as $P$ that records which box appears at which step of the insertion. 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 10.9.3.** Let $w = 57214836 \in S_8$. We 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 $Q$.

\[
P = \begin{bmatrix} 5 \end{bmatrix} \quad Q = \begin{bmatrix} 1 \end{bmatrix}
\]

(10.9.1a)

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 $P$.

\[
P = \begin{bmatrix} 5 & 7 \end{bmatrix} \quad Q = \begin{bmatrix} 1 & 2 \end{bmatrix}
\]

(10.9.1b)
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 $Q$.

\[
P = \begin{array}{cc}
2 & 7 \\
5 & \\
\end{array}
\quad Q = \begin{array}{cc}
1 & 2 \\
3 & \\
\end{array}
\]  

(10.9.1c)

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.

\[
P = \begin{array}{cc}
1 & 7 \\
2 & \\
5 & \\
\end{array}
\quad Q = \begin{array}{cc}
1 & 2 \\
3 & \\
4 & \\
\end{array}
\]  

(10.9.1d)

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.

\[
P = \begin{array}{cc}
1 & 4 \\
2 & 7 \\
5 & \\
\end{array}
\quad Q = \begin{array}{cc}
1 & 2 \\
3 & 5 \\
4 & \\
\end{array}
\]  

(10.9.1e)

Step 6: Row-insert $w_6 = 8$ into $P$. The 8 just goes at the end of the first row.

\[
P = \begin{array}{cc}
1 & 4 & 8 \\
2 & 7 & \\
5 & \\
\end{array}
\quad Q = \begin{array}{cc}
1 & 2 & 6 \\
3 & 5 & \\
4 & \\
\end{array}
\]  

(10.9.1f)

Step 7: Row-insert $w_7 = 3$ into $P$. 3 bumps 4, and then 4 bumps 7.

\[
P = \begin{array}{cc}
1 & 3 & 8 \\
2 & 4 & \\
5 & 7 & \\
\end{array}
\quad Q = \begin{array}{cc}
1 & 2 & 6 \\
3 & 5 & \\
4 & 7 & \\
\end{array}
\]  

(10.9.1g)

Step 8: Row-insert $w_8 = 6$ into $P$. 6 bumps 8 into the second row.

\[
P = \begin{array}{cc}
1 & 3 & 6 \\
2 & 4 & 8 \\
5 & 7 & \\
\end{array}
\quad Q = \begin{array}{cc}
1 & 2 & 6 \\
3 & 5 & 8 \\
4 & 7 & \\
\end{array}
\]  

(10.9.1h)
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 10.9.4.** Suppose that we were given the pair of tableaux in (10.9.1h). 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 (10.9.1g).

**Example 10.9.5.** Suppose \(P\) is the standard tableau with 18 boxes shown on the left. Suppose in addition 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 boldface on the right. (That is, the 16 was bumped by the 15, which was bumped by the 13, and so on.) To find the previous tableau in the algorithm, we push every number in the bumping path up and toss out the top one.

That is, we must have obtained the original tableau by row-inserting 10 into the tableau on the right.

Iterating this “de-insertion” allows us to recover \(w\) from the pair \((P, Q)\). We have proved the following fact:

**Theorem 10.9.6.** The RSK correspondence is a bijection

\[
\mathfrak{S}_n \xrightarrow{\text{RSK}} \bigcup_{\lambda \vdash n} \text{SYT}(\lambda) \times \text{SYT}(\lambda).
\]

**Corollary 10.9.7.** \(\sum_{\lambda \vdash n} (f^{\lambda})^2 = n!\).

**Example 10.9.8.** The SYT’s with \(n = 3\) boxes are as follows:

\[
\begin{align*}
&\begin{array}{ccc}
1 & 2 & 3 \\
1 & 3 & 2 \\
3 & 2 & 1
\end{array} & & \begin{array}{ccc}
1 & 2 & 3 \\
1 & 3 & 2 \\
2 & 3 & 1
\end{array} & & \begin{array}{ccc}
1 & 2 & 3 \\
2 & 1 & 3 \\
3 & 1 & 2
\end{array}
\end{align*}
\]

Note that \(f^{(3)} = f^{(1,1,1)} = 1\) and \(f^{(2,1)} = 2\), and \(1^2 + 2^2 + 1^2 = 6 = 3!\).

**Example 10.9.9.** The SYT’s with \(n = 4\) boxes are as follows:

\[
\begin{align*}
&\begin{array}{cccc}
1 & 2 & 3 & 4 \\
1 & 2 & 3 & 4 \\
1 & 2 & 3 & 4 \\
1 & 2 & 3 & 4
\end{array} & & \begin{array}{cccc}
1 & 2 & 3 & 4 \\
1 & 2 & 3 & 4 \\
1 & 2 & 3 & 4 \\
2 & 3 & 4 & 1
\end{array} & & \begin{array}{cccc}
1 & 2 & 3 & 4 \\
1 & 2 & 3 & 4 \\
2 & 3 & 4 & 1 \\
3 & 4 & 1 & 2
\end{array}
\end{align*}
\]

162
The numbers $f^{\lambda}$ are therefore as follows:

<table>
<thead>
<tr>
<th>Shape $\lambda$</th>
<th>$f^{\lambda}$</th>
<th>$(f^{\lambda})^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(4)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(1, 1, 1, 1)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(3, 1)</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>(2, 1, 1)</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>(2, 2)</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>24</td>
</tr>
</tbody>
</table>

These numbers ought to look familiar! Another neat fact about the RSK correspondence is this:

**Proposition 10.9.10.** Let $w \in \mathfrak{S}_n$. If $RSK(w) = (P, Q)$, then $RSK(w^{-1}) = (Q, P)$. In particular, the number of involutions in $\mathfrak{S}_n$ is $\sum_{\lambda\vdash n} f^{\lambda}$.

This is hard to see 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 is an alternate description of RSK [Sta99, §7.13] from which the symmetry is more apparent; see

The RSK correspondence can be extended to more general tableaux, and this turns out to be the key to expanding the Cauchy kernel in terms of Schur functions.

**Definition 10.9.11.** A generalized permutation of length $n$ is a $2 \times n$ matrix

$$w = \begin{pmatrix} q \\ p \end{pmatrix} = \begin{pmatrix} q_1 & q_2 & \cdots & q_n \\ p_1 & p_2 & \cdots & p_n \end{pmatrix}$$

where $q = (q_1, \ldots, q_n), p = (p_1, \ldots, p_n) \in [n]^n$, and the ordered pairs $(q_1, p_1), \ldots, (q_n, p_n)$ are in lexicographic order, i.e., $q_1 \leq \cdots \leq q_n$, and if $q_i = q_{i+1}$ then $p_i \leq p_{i+1}$. The weight of $w$ is the monomial $x^P y^Q = x_{p_1} \cdots x_{p_n} y_{q_1} \cdots y_{q_n}$.

The set of all generalized permutations of length $n$ will be denoted $GP(n)$. If $q_i = i$ for all $i$ and the $p_i$’s are pairwise distinct, then $w = \begin{pmatrix} q \\ p \end{pmatrix}$ can be regarded as an ordinary permutation $w \in \mathfrak{S}_n$ by reading off the $p$’s, and evidently every element of $\mathfrak{S}_n$ arises in this way. (This is just the standard two-line notation for $w$.)

We can run the RSK algorithm on generalized permutations just as for plain vanilla permutations. At the $i^{th}$ 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 10.9.12.** Consider the generalized permutation

$$w = \begin{pmatrix} q \\ p \end{pmatrix} = \begin{pmatrix} 1 & 1 & 2 & 4 & 4 & 4 & 5 & 5 \\ 2 & 4 & 1 & 1 & 3 & 3 & 2 & 4 \end{pmatrix} \in GP(9).$$

The RSK algorithm proceeds as follows:
The tableaux $P, Q$ we get in this way 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 $q$ occur consecutively (say as $q_{i}, \ldots, q_{j}$), and the corresponding entries of $p$ are weakly increasing, so none of them will bump any other, 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 $w$. 

\begin{table}
\centering
\begin{tabular}{ccc}
Step # & $P$ & $Q$
\hline
1 & 2 & 1 \\
2 & 2 4 & 1 1 \\
3 & 1 4 & 1 1 \\
4 & 1 1 & 1 1 \\
5 & 1 1 3 & 1 1 4 \\
6 & 1 1 3 3 & 1 1 4 4 \\
7 & 1 1 2 3 & 1 1 4 4 \\
8 & 1 1 2 2 & 1 1 4 4 \\
9 & 1 1 2 2 4 & 1 1 4 4 5 \\
\end{tabular}
\end{table}

164
Accordingly, we have a bijection
\[
GP(n) \xrightarrow{\text{RSK}} \bigcup_{\lambda \vdash n} \{(P,Q) \mid P,Q \in \text{CST}(\lambda)\}.
\] (10.9.3)
in which a generalized permutation \((q,p)\) maps to a pair of tableaux \(P,Q\) with weight monomials \(x^P\) and \(y^Q\).

On the other hand, a generalized permutation \(w = (q,p) \in GP(n)\) can also be specified by an infinite matrix \(M = [m_{ij}]\) with finitely many nonzero entries, in which \(m_{ij}\) is the number of occurrences of \((q_i, p_j)\) in \(w\) (so \(n = \sum m_{ij}\)). For example, the generalized permutation \(w \in GP(9)\) of Example 10.9.12 corresponds to the integer matrix
\[
\begin{bmatrix}
0 & 1 & 0 & 1 & 0 & \cdots \\
1 & 0 & 0 & 0 & 0 & \cdots \\
0 & 0 & 0 & 0 & 0 & \cdots \\
1 & 0 & 2 & 0 & 0 & \cdots \\
0 & 2 & 0 & 1 & 0 & \cdots \\
0 & 0 & 0 & 0 & 0 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots \\
\end{bmatrix}
\]

Under this bijection, the weight monomial \(x^P y^Q\) equals \(\prod_{i,j} (x^i y^j)^{m_{ij}}\). Any given weight monomial arises from only finitely many matrices, so the generating function for matrices by weights is a well-defined power series, namely
\[
\sum_{M = [m_{ij}]} \prod_{i,j} (x^i y^j)^{m_{ij}} = \prod_{i,j \geq 1} \sum_{m_{ij} = 0}^{\infty} (x^i y^j)^{m_{ij}} = \prod_{i,j \geq 1} \frac{1}{1 - x^i y^j}
\]
which is precisely the Cauchy kernel \(\Omega\). On the other hand,
\[
\Omega = \sum_{M = [m_{ij}]} \prod_{i,j} (x^i y^j)^{m_{ij}}
\]
\[
= \sum_{n \in \mathbb{N}} \sum_{(q,p) \in GP(n)} x_{p_1} \cdots x_{p_n} y_{q_1} \cdots y_{q_n} \quad \text{(by the bijection between matrices and GP’s)}
\]
\[
= \sum_{\lambda} \sum_{P,Q \in \text{CST}(\lambda)} x^P y^Q \quad \text{(by the RSK correspondence)}
\]
\[
= \sum_{\lambda} \left( \sum_{P \in \text{CST}(\lambda)} x^P \right) \left( \sum_{Q \in \text{CST}(\lambda)} y^Q \right)
\]
\[
= \sum_{\lambda} s_\lambda(x)s_\lambda(y).
\]

We have proven:

**Theorem 10.9.13.** \(\Omega = \sum_\lambda s_\lambda(x)s_\lambda(y)\). Consequently, the Schur functions form an orthonormal \(\mathbb{Z}\)-basis for \(\Lambda\) under the Hall inner product.

### 10.10 Knuth equivalence and jeu de taquin

**Definition 10.10.1.** Let \(b, b'\) be finite ordered lists of positive integers (or “words in the alphabet \(\mathbb{P}\)”). We say that \(b, b'\) are **Knuth equivalent**, written \(b \sim b'\), if one can be obtained from the other by a sequence of transpositions as follows:
1. If $x \leq y < z$, then $\cdots xzy \cdots \sim \cdots zxy \cdots$.
2. If $x < y \leq z$, 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, $212231 \sim 212231$ by Rule 1, and $212231 \sim 212231$ by Rule 2 (applied in reverse).

This definition looks completely unmotivated at first, but hold that thought!

**Definition 10.10.2.** Let $\lambda, \mu$ be partitions with $\mu \subseteq \lambda$. The skew (Ferrers) shape $\lambda/\mu$ is defined by removing from $\lambda$ the boxes in $\mu$.

For example, if $\lambda = (4, 4, 2, 1)$, $\mu = (3, 2)$, and $\mu' = (3, 3)$, then $\nu = \lambda/\mu$ and $\nu' = \lambda/\mu'$ are as follows:

\[
\begin{array}{cccc}
\lambda & = & \begin{array}{cccc}
\cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
\end{array} & & \nu & = & \begin{array}{cccc}
\cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
\end{array} & & \nu' & = & \begin{array}{cccc}
\cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
\end{array} \\
\end{array}
\]

(where the $\times$’s mean “delete this box”). Note that there is no requirement that a skew shape be connected.

**Definition 10.10.3.** Let $\nu = \lambda/\mu$ be a skew shape. A column-strict (skew) tableau of shape $\nu$ is a filling of the boxes of $\nu$ with positive integers such that each row is weakly increasing eastward and each column is strictly increasing southward. (Note that if $\mu = \emptyset$, this is just a CST; see Definition 10.7.1.) For example, here are two column-strict skew tableaux:

\[
\begin{array}{cccc}
1 & 4 & & \\
2 & 3 & 4 & \\
1 & 2 & & \\
\end{array} & & \begin{array}{cccc}
1 & & & \\
3 & & & \\
\end{array} & & \begin{array}{cccc}
3 & & & \\
2 & 4 & & \\
\end{array} & & \begin{array}{cccc}
& 1 & & \\
& 2 & & \\
& & & \\
\end{array}
\]

Again, there is no requirement that a skew tableau be connected.

We now define an equivalence relation on column-strict skew tableaux, called *jeu de taquin*\(^4\). The rule is as follows:

\[
\begin{array}{cccc}
\circ & y & x & y \\
x & & \rightarrow & \\
\end{array} & & \begin{array}{cccc}
\bullet & y & x & y \\
x & \rightarrow & \\
\end{array} & & \begin{array}{cccc}
\circ & y & x & y \\
x & \rightarrow & \\
\end{array}
\]

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 $x \leq y$) 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 $\bullet$ denote the inner corner that is being slid into.

---

\(^4\)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 two skew tableaux $T, T'$ can be obtained from each other by such slides (or by their reverses), we say that they are \textit{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 \textit{rectification} of $T$); see, e.g., the example (10.10.1) above. In fact, the rectification is unique; the order in which we choose inner corners does not matter.

\textbf{Definition 10.10.4.} Let $T$ be a column-strict skew tableau. The \textit{row-reading word} of $T$, denoted row($T$), is obtained by reading the rows left to right, bottom to top.

For example, the reading words of the skew tableaux in (10.10.1) are

$$2341214, 2342114, 2342114, 2324114, 3224114, 3224114, 3224114, 3224114.$$ 

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 row($T$). For skew tableaux, this is not the case. Note that some of the slides in (10.10.1) do not change the row reading word; as a simpler example, the following skew tableaux both have reading word 122:

$$122 \quad 212 \quad 212.$$ 

On the other hand, it’s not hard to see that rectifying the second or third tableau will yield the first; therefore, they are all jeu de taquin equivalent.

For a word $b$ on the alphabet $P$, let $P(b)$ denote its insertion tableau under the RSK algorithm. (That is, construct a generalized permutation ($q$) in which $q$ is any word; run RSK; and remember only the tableau $P$, so that the choice of $q$ does not matter.)

\textbf{Theorem 10.10.5.} (Knuth–Schützenberger) For two words $b, b'$, the following are equivalent:

1. $P(b) = P(b')$.
2. $b \sim b'$.
3. $T \approx T'$, for any (or all) column-strict skew tableaux $T, T'$ with row-reading words $b, b'$ respectively.

This is sometimes referred to (e.g., in [Ful97]) as the equivalence of “bumping” (the RSK algorithm as presented in Section 10.9) and “sliding” (jeu de taquin).

\section{10.11 Yet another version of RSK}

Fix $w \in S_n$. 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 \( \lambda, \mu, \nu \) respectively, label the NE corner \( \rho \) 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 \( \lambda \subseteq \mu = \nu \) and the box doesn’t contain an X, then it must be the case that \( \mu_i = \lambda_i + 1 \) for some \( i \). Obtain \( \rho \) from \( \mu \) by incrementing \( \mu_i + 1 \).

**Rule 3:** If \( \mu \neq \nu \), then set \( \rho = \mu \lor \nu \) (where \( \lor \) means the join in Young’s lattice: i.e., take the componentwise maximum of the elements of \( \mu \) and \( \nu \)).

**Rule X:** If there is an X in the box, then it must be the case that \( \lambda = \mu = \nu \). Obtain \( \rho \) from \( \lambda \) by incrementing \( \lambda_1 \).

Note that the underlined assertions need to be proved; this can be done by induction.

**Example 10.11.1.** Let \( n = 8 \) and \( w = 57214836 \). In Example 10.9.3, we found that \( \text{RSK}(w) = (P, Q) \), where

\[
P = \begin{bmatrix}
1 & 3 & 6 \\
2 & 4 & 8 \\
5 & 7
\end{bmatrix}
\quad \text{and} \quad
Q = \begin{bmatrix}
1 & 2 & 6 \\
3 & 5 & 8 \\
4 & 7
\end{bmatrix}.
\]

The following extremely impressive figure shows what happens when we run the alternate RSK algorithm on \( w \). The partitions \( \lambda \) are shown in red. The numbers in parentheses indicate which rules were used.
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 $k^{th}$ row, and the northernmost partition $\mu(k)$ in the $k^{th}$ 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!

### 10.12 The Frobenius characteristic

As in Section 9.6, denote by $\mathcal{C}(\mathfrak{S}_n)$ the vector space of $\mathbb{C}$-valued class functions on the symmetric group $\mathfrak{S}_n$. Define a graded vector space

$$\mathcal{C}(\mathfrak{S}) = \bigoplus_{n \geq 0} \mathcal{C}(\mathfrak{S}_n)$$

(where $\mathcal{C}(\mathfrak{S}_0)$ is just a copy of $\mathbb{C}$).

We make $\mathcal{C}(\mathfrak{S})$ into a graded $\mathbb{C}$-algebra as follows. First, we declare that the elements of $\mathcal{C}(\mathfrak{S}_0)$ behave like scalars. Second, for $n_1, n_2 \in \mathbb{P}$, set $\mathfrak{S}_{n_1,n_2} = \mathfrak{S}_{n_1} \times \mathfrak{S}_{n_2}$. For $f_1 \in \mathcal{C}(\mathfrak{S}_{n_1})$, define $f_1 \times f_2 \in \mathcal{C}(\mathfrak{S}_{n_1,n_2})$
by
\[(f_1 \times f_2)(w_1, w_2) = f_1(w_1)f_2(w_2)\].

There is a natural inclusion of groups \(\mathfrak{S}_{n_1, n_2} \to \mathfrak{S}_n\), where \(n = n_1 + n_2\). Let \(\mathfrak{S}_{n_1}\) act on \(\{1, 2, \ldots, n_1\}\) and let \(\mathfrak{S}_{n_2}\) act on \(\{n_1 + 1, n_1 + 2, \ldots, n_1 + n_2\}\). Thus we can define \(f_1f_2 \in \mathcal{C}\ell(\mathfrak{S}_n)\) by means of the induced “character”, applying Proposition 9.10.4:
\[
f_1f_2 = \text{Ind}_{\mathfrak{S}_{n_1, n_2}}^{\mathfrak{S}_n} (f_1 \times f_2) = \frac{1}{n_1! n_2!} \sum_{g \in \mathfrak{S}_n: g^{-1}wg \in \mathfrak{S}_{n_1, n_2}} (f_1 \times f_2)(g^{-1}wg).
\]

This product makes \(\mathcal{C}\ell(\mathfrak{S})\) into a commutative graded \(\mathbb{C}\)-algebra. (We omit the proof; one has to check properties like associativity.)

For a partition \(\lambda \vdash n\), let \(1_\lambda \in \mathcal{C}\ell(\mathfrak{S}_n)\) be the indicator function on the conjugacy class \(C_\lambda \subseteq \mathfrak{S}_n\), and let \(\mathfrak{S}_\lambda = \mathfrak{S}_{\{1, \ldots, \lambda_1\}} \times \mathfrak{S}_{\{\lambda_1 + 1, \ldots, \lambda_1 + \lambda_2\}} \times \cdots \times \mathfrak{S}_{\{n - \lambda_1 + 1, \ldots, n\}} \subseteq \mathfrak{S}_n\).

For \(w \in \mathfrak{S}_n\), let \(\text{sh}(w)\) denote the cycle-shape of \(w\), expressed as a partition. Define a function \(\psi: \mathfrak{S}_n \to \Lambda^n\) by
\[
\psi(w) = p_{\text{sh}(w)}.
\]

Note that \(\psi\) is a class function (albeit with values in \(\Lambda\) rather than in \(\mathbb{C}\)).

**Definition 10.12.1.** The *Frobenius characteristic* is the map
\[
\text{ch}: \mathcal{C}\ell_{\mathbb{C}}(\mathfrak{S}) \to \Lambda_{\mathbb{C}}
\]
defined on \(f \in \mathcal{C}\ell(\mathfrak{S}_n)\) by
\[
\text{ch}(f) = \langle f, \psi \rangle_{\mathfrak{S}_n} = \frac{1}{n!} \sum_{w \in \mathfrak{S}_n} f(w) p_{\text{sh}(w)}.
\]

**Theorem 10.12.2.** The Frobenius characteristic \(\text{ch}\) has the following properties:

1. \(\text{ch}(1_\lambda) = p_\lambda/z_\lambda\).
2. \(\text{ch}\) is an isometry, i.e., it preserves inner products:
\[
\langle f, g \rangle_{\mathfrak{S}_n} = \langle \text{ch}(f), \text{ch}(g) \rangle_{\Lambda}.
\]
3. \(\text{ch}\) is a ring isomorphism.
4. \(\text{ch}(\text{Ind}_{\mathfrak{S}_\lambda}^{\mathfrak{S}_n} \chi_{\text{triv}}) = h_\lambda\).
5. \(\text{ch}(\text{Ind}_{\mathfrak{S}_\lambda}^{\mathfrak{S}_n} \chi_{\text{sign}}) = e_\lambda\).
6. Let \(\chi\) be any character of \(\mathfrak{S}_n\) and let \(\chi_{\text{sign}}\) be the sign character on \(\mathfrak{S}_n\). Then \(\text{ch}(\chi \otimes \chi_{\text{sign}}) = \omega(\text{ch}(\chi))\), where \(\omega\) is the involution of 10.5.4.
7. \(\text{ch}\) restricts to an isomorphism \(\mathcal{C}\ell_{\mathbb{Z}}(\mathfrak{S}) \to \Lambda_{\mathbb{Z}}\), where \(\mathcal{C}\ell_{\mathbb{Z}}(\mathfrak{S})\) 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 \(\{\text{ch}^{-1}(s_\lambda) \mid \lambda \vdash n\}\).

**Proof.** (1): Recall from (10.8.1) that \(|C_\lambda| = n!/z_\lambda\), where \(z_\lambda = 1^{r_1}2^{r_2}r_2! \ldots\), where \(r_i\) is the number of occurrences of \(i\) in \(\lambda\). Therefore
\[
\text{ch}(1_\lambda) = \frac{1}{n!} \sum_{w \in C_\lambda} p_\lambda = p_\lambda/z_\lambda.
\]

170
It follows that \( \text{ch} \) is (at least) a graded \( \mathbb{C} \)-vector space isomorphism, since \( \{1_\lambda\} \) and \( \{p_\lambda/z_\lambda\} \) are graded \( \mathbb{C} \)-bases for \( C\ell(\mathfrak{S}) \) and \( \Lambda \) respectively.

(2): We have
\[
\langle 1_\lambda, 1_\mu \rangle_{\mathfrak{S}_n} = \frac{1}{n!} \sum_{w \in \mathfrak{S}_n} 1_\lambda(w)1_\mu(w) = \frac{1}{n!}\langle C_\lambda|\delta_{\lambda\mu} = \delta_{\lambda\mu}/z_\lambda = \langle p_\lambda/z_\lambda, p_\mu/z_\mu \rangle_{\Lambda}
\]
(we previously proved the last equality by expanding the Cauchy kernel; see Section 10.8). (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{align*}
\text{ch}(fg) &= \langle \text{Ind}^{\mathfrak{S}_n}_{\mathfrak{S}_j \times \mathfrak{S}_k} (f \times g), \psi \rangle_{\mathfrak{S}_n} \quad \text{(where \( \psi \) is defined as in (9.10.4)}) \\
&= \langle f \times g, \text{Res}^{\mathfrak{S}_n}_{\mathfrak{S}_j \times \mathfrak{S}_k} \psi \rangle_{\mathfrak{S}_j \times \mathfrak{S}_k} \quad \text{(by Frobenius reciprocity)} \\
&= \frac{1}{j!k!} \sum_{(w,x) \in \mathfrak{S}_j \times \mathfrak{S}_k} f(w)g(x)p_{sh(wx)} \\
&= \left( \frac{1}{j!} \sum_{w \in \mathfrak{S}_j} f(w) p_{sh(w)} \right) \left( \frac{1}{k!} \sum_{x \in \mathfrak{S}_k} g(x) p_{sh(x)} \right) \quad \text{(because the power-sum basis is multiplicative)} \\
&= \text{ch}(f) \text{ch}(g).
\end{align*}
\]

(4), (5): Using the identities of Corollary 10.8.2, we get
\[
\begin{align*}
\text{ch}(\chi_{\text{triv}}^{\mathfrak{S}_n}) &= \langle \chi_{\text{triv}}, \psi \rangle_{\mathfrak{S}_n} = \frac{1}{n!} \sum_{w \in \mathfrak{S}_n} p_{sh(w)} = \sum_{\lambda \vdash n} \frac{p_\lambda}{z_\lambda} = h_n, \\
\text{ch}(\chi_{\text{sign}}^{\mathfrak{S}_n}) &= \langle \chi_{\text{sign}}, \psi \rangle_{\mathfrak{S}_n} = \frac{1}{n!} \sum_{w \in \mathfrak{S}_n} \varepsilon_{sh(w)} p_{sh(w)} = \sum_{\lambda \vdash n} \varepsilon_{\lambda} \frac{p_\lambda}{z_\lambda} = e_n,
\end{align*}
\]
and since \( \text{ch} \) is a ring homomorphism, we obtain
\[
\begin{align*}
h_\lambda &= \prod_{i=1}^\ell h_{\lambda_i} = \prod_{i=1}^\ell \text{ch}(\chi_{\text{triv}}^{\mathfrak{S}_n}) = \text{ch} \left( \prod_{i=1}^\ell \chi_{\text{triv}}^{\mathfrak{S}_n} \right) = \text{ch}(\text{Ind}^{\mathfrak{S}_n}_{\mathfrak{S}_\lambda} \chi_{\text{triv}}) \\
e_\lambda &= \text{ch}(\text{Ind}^{\mathfrak{S}_n}_{\mathfrak{S}_\lambda} \chi_{\text{sign}}).
\end{align*}
\]

(6): Left as an exercise.

(7), (8): Each of (4) and (5) says that \( \text{ch}^{-1}(\Lambda_\mathbb{Z}) \) 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 \( \text{ch} \) are genuine characters. On the other hand, \( \{s_\lambda\} \) is also a \( \mathbb{Z} \)-basis, and since \( \text{ch} \) is an isometry, the characters \( \sigma_\lambda := \text{ch}^{-1}(s_\lambda) \) satisfy
\[
\langle \sigma_\lambda, \sigma_{\mu} \rangle_{\mathfrak{S}_n} = \langle s_\lambda, s_\mu \rangle_\Lambda = \delta_{\lambda\mu}
\]
which must mean that \( \{\sigma_\lambda \mid \lambda + n\} \) is a \( \mathbb{Z} \)-basis for \( C\ell_\mathbb{V}(\mathfrak{S}_n) \), and that each \( \sigma_\lambda \) is either an irreducible character or its negative. We can determine which it is by checking the sign of \( \sigma_\lambda(\text{Id}_{\mathfrak{S}_n}) \).

To do this, expand the Schur functions in the basis \( \{p_\mu/z_\mu\} \). We don’t know what the coefficients are, so let’s call them \( \chi_{\lambda,\mu} \). That is,
\[
s_\lambda = \sum_{\mu \vdash n} \chi_{\lambda,\mu} \frac{p_\mu}{z_\mu} = \text{ch} \left( \sum_{\mu \vdash n} \chi_{\lambda,\mu} 1_\mu \right)
\]
(10.12.2)
so in fact $\chi_{\lambda,\mu}$ is the value of the irreducible character indexed by $\lambda$ on the conjugacy class $C_\mu$. So what we care about is $\chi_{\lambda,\nu}$, where $\nu = (1,1,\ldots,1) = sh(Id_{S_n})$.

The coefficient of $x_1x_2\ldots x_n$ in $s_\lambda$ is just $f^\lambda$ (the number of standard tableaux of shape $\lambda$), and every such monomial must come from the summand $p_\mu/z_\mu$ in (10.12.2). In particular $\sigma_\lambda(Id_{S_n}) = \chi_{\lambda,\nu} = f^\lambda > 0$, proving that $\sigma_\lambda$ (rather than $-\sigma_\lambda$) is a genuine irreducible character.

The Frobenius characteristic allows us to translate back and forth between representations (equivalently, characters) of symmetric groups, and symmetric functions. In particular, questions about representations of $S_n$ can now be translated into tableau combinatorics. Here are a few fundamental things we would like to know at this point.

1. Littlewood-Richardson numbers. Now that we know important the Schur functions are from a representation-theoretic standpoint, how do we multiply them? That is, suppose that $\mu,\nu$ 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:

$$s_\mu s_\nu = \sum_\lambda c_{\lambda,\mu,\nu} s_\lambda, \quad c_{\lambda,\mu,\nu} \in \mathbb{Z}. \quad (10.12.3)$$

The $c_{\lambda,\mu,\nu} \in \mathbb{Z}$ are called the Littlewood-Richardson numbers. They are the structure coefficients for $\Lambda$, regarded as an algebra generated as a vector space by the Schur functions. Note that they 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 (10.12.3) is equivalent to

$$c_{\mu,\nu}^{\lambda} = \langle s_\mu s_\nu, s_\lambda \rangle$$

and applying $\text{ch}^{-1}$ gives an interpretation of the $c_{\mu,\nu}^{\lambda}$ in terms of characters, namely

$$c_{\mu,\nu}^{\lambda} = \langle \text{Ind}_{S_q \times S_r}^{S_n}(\chi^\mu \otimes \chi^\nu), \chi^\lambda \rangle_{S_n} = \langle \chi^\mu \otimes \chi^\nu, \text{Res}_{S_q \times S_r}^{S_n}(\chi^\lambda) \rangle_{S_q \times S_r},$$

where the second equality comes from Frobenius reciprocity.

The Littlewood-Richardson rule gives a combinatorial interpretation for the numbers $c_{\mu,\nu}^{\lambda}$. There are many different Littlewood-Richardson rules!

2. Irreducible characters. What is the value $\chi_{\lambda,\mu}$ of the irreducible character $\sigma_\lambda$ on the conjugacy class $C_\mu$? In other words, what is the character table of $S_n$? 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 $\Lambda_n$. The precise combinatorial formulation is given by the Murnaghan-Nakayama Rule.

3. Dimensions of irreducible characters. A special case of the Murnaghan-Nakayama Rule is that the irreducible representation indexed by $\lambda \vdash n$ (i.e., with character $\sigma_\lambda = \text{ch}^{-1}(s_\lambda)$) has dimension $f^\lambda$, the number of standard tableaux of shape $\lambda$. What are the numbers $f^\lambda$? There is a beautiful interpretation called the hook-length formula of Frame, Robinson and Thrall, which again has many, many proofs in the literature.

4. Transition matrices. What are the coefficients of the transition matrices between different bases of $\Lambda_n$? 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). I also claimed that the Kostka numbers show up in representation theory as the multiplicities of Specht modules in tabloid representations of $S_n$ (see Section 9.11). Why is that the case?
10.13 Skew tableaux and the Littlewood-Richardson Rule

Let $\nu = \lambda/\mu$ be a skew shape, and let $\text{CST}(\lambda/\mu)$ denote the set of all column-strict skew tableaux of shape $\lambda/\mu$. It is natural to define the skew Schur function

$$s_{\lambda/\mu}(x_1, x_2, \ldots) = \sum_{T \in \text{CST}(\lambda/\mu)} x_T.$$

For example, suppose that $\lambda = (2, 2)$ and $\mu = (1)$, so that $\nu = (0)$. What are the possibilities for $T \in \text{CST}(\nu)$? Clearly the entries cannot all be equal. If $a < b < c$, then there are two ways to fill $\nu$ with $a, b, c$ (left, below). If $a < b$, then there is one way to fill $\nu$ with two $a$’s and one $b$ (center), and one way to fill $\nu$ with one $a$ and two $b$’s (right).

$$a \ b \ c \quad b \ a \ c \quad a \ b \ b \quad 1 \ b \ b$$

Therefore, $s_{\nu} = 2m_{111} + m_{21}$ (these are monomial symmetric functions). In fact, skew Schur functions are always symmetric. This is not obvious, but is not too hard to prove. (Like ordinary Schur functions, it is fairly easy to see that they are quasisymmetric.) Therefore, we can write

$$s_{\lambda/\mu} = \sum_{\nu} \tilde{c}_{\lambda/\mu, \nu}s_{\nu}$$

where $\tilde{c}_{\lambda/\mu, \nu} \in \mathbb{Z}$ for all $\lambda, \mu, \nu$. The punchline is that the tildes are unnecessary: these numbers are in fact the Littlewood-Richardson coefficients $c_{\mu, \nu}^\lambda$ of equation (10.12.3). Better yet, they are symmetric in $\mu$ and $\nu$.

**Proposition 10.13.1.** Let $x = \{x_1, x_2, \ldots\}$, $y = \{y_1, y_2, \ldots\}$ be two countably infinite sets of variables. Think of them as an alphabet with $1 < 2 < \cdots < 1' < 2' < \cdots$. Then

$$s_{\lambda}(x, y) = \sum_{\mu \subseteq \lambda} s_{\mu}(x)s_{\lambda/\mu}(y).$$

**Proof.** Every $T \in \text{CST}(\lambda)$ labeled with $1, 2, \ldots, 1', 2', \ldots$ consists of a CST of shape $\mu$ filled with $1, 2, \ldots$ (for some $\mu \subseteq \lambda$) together with a CST of shape $\lambda/\mu$ filled with $1', 2', \ldots$. \qed

**Theorem 10.13.2.** For all partitions $\lambda, \mu, \nu$, we have

$$\tilde{c}_{\lambda/\mu, \nu} = c_{\mu, \nu}^\lambda = c_{\nu, \mu}^\lambda.$$

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 $x, y, z$ for this. Consider the “double Cauchy kernel”

$$\Omega(x, z)\Omega(y, z) = \prod_{i,j}(1 - x_iz_j)^{-1}\prod_{i,j}(1 - y_iz_j)^{-1}.$$
On the one hand, expanding both factors in terms of Schur functions and then applying the definition of the Littlewood-Richardson coefficients to the $z$ terms gives
\[
\Omega(x, z) \Omega(y, z) = \left( \sum_{\mu} s_{\mu}(x) s_{\mu}(z) \right) \left( \sum_{\nu} s_{\nu}(y) s_{\nu}(z) \right) = \sum_{\mu, \nu} s_{\mu}(x) s_{\nu}(y) s_{\mu}(z) s_{\nu}(z) \]
\[
= \sum_{\mu, \nu} s_{\mu}(x) s_{\nu}(y) \sum_{\lambda \mu} c_{\mu, \nu}^\lambda s_{\lambda}(z). \tag{10.13.1}
\]
On the other hand, we also have (formally setting $s_{\lambda/\mu} = 0$ if $\mu \not\subseteq \lambda$)
\[
\Omega(x, z) \Omega(y, z) = \sum_{\lambda} s_{\lambda}(x, y) s_{\lambda}(z) = \sum_{\lambda} \sum_{\mu \subseteq \lambda} s_{\mu}(x) s_{\lambda/\mu}(y) s_{\lambda}(z) \]
\[
= \sum_{\lambda} \sum_{\mu} s_{\mu}(x) s_{\lambda}(z) \sum_{\nu} \tilde{c}_{\lambda/\mu, \nu} s_{\nu}(y) \]
\[
= \sum_{\mu, \nu} s_{\mu}(x) s_{\nu}(y) \sum_{\lambda} s_{\lambda}(z) \tilde{c}_{\lambda/\mu, \nu}. \tag{10.13.2}
\]
(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 $x \cup y$ and $z$. The second equality uses Proposition 10.13.1.) Now the theorem follows from the equality of (10.13.1) and (10.13.2).

There are a lot of combinatorial interpretations of the Littlewood-Richardson numbers. Here is one:

Theorem 10.13.3 (Littlewood-Richardson Rule). $c_{\lambda/\mu}^\nu$ equals the number of column-strict tableaux $T$ of shape $\lambda/\mu$, and content $\nu$ such that the reverse of row($T$) is a ballot sequence (or Yamanouchi word, or lattice permutation): that is, each initial sequence of it contains at least as many 1’s as 2’s, at least as many 2’s as 3’s, et cetera.

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 10.13.4 (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
\[
s_{\lambda}s_{(k)} = s_{\lambda}h_k = \sum_{\mu} s_{\mu}
\]
where $\mu$ ranges over all partitions obtained from $\lambda$ by adding $k$ boxes, no more than one in each column; and
\[
s_{\lambda}s_{(1^k)} = s_{\lambda}e_k = \sum_{\mu} s_{\mu}
\]
where $\mu$ ranges over all partitions obtained from $\lambda$ by adding $k$ boxes, no more than one in each row.

Another important, even more special case is
\[
s_{\lambda}s_1 = \sum_{\mu} s_{\mu}
\]
where $\mu$ ranges over all partitions obtained from $\lambda$ 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:
\[
\text{Res}_{\mathfrak{S}_{n-1}}^{\mathfrak{S}_n} (\chi^\lambda) = \bigoplus_{\mu} \chi^\mu
\]
where now $\mu$ ranges over all partitions obtained from $\lambda$ by deleting a single box.
10.14 The Murnaghan-Nakayama Rule

We know from Theorem 10.12.2 that the irreducible characters of $S_n$ are $\chi^\lambda = \text{ch}^{-1}(s_\lambda)$ for $\lambda \vdash n$. The Murnaghan-Nakayama Rule gives a formula for the value of the character $\chi^\lambda$ on the conjugacy class $C_\mu$ in terms of rim-hook tableaux. Here is an example of a rim-hook tableau of shape $\lambda = (5, 4, 3, 3, 1)$ and content $\mu = (6, 3, 3, 2, 1, 1)$:

Note that the columns and row are weakly increasing, and for each $i$, the set $H_i(T)$ of cells containing an $i$ is contiguous.

**Theorem 10.14.1** (Murnaghan-Nakayama Rule (1937)).

$$\chi^\lambda(C_\mu) = \sum_{\text{rim-hook tableaux } T \text{ of shape } \lambda \text{ and content } \mu} \prod_{i=1}^{n} (-1)^{1 + \text{ht}(H_i(T))}.$$

For example, the heights of $H_1, \ldots, H_6$ in the rim-hook tableau above are 4, 3, 2, 1, 1, 1. There are an even number of even heights, so this rim-hook tableau contributes 1 to $\chi^\lambda(C_\mu)$.

An important special case is when $\mu = (1, 1, \ldots, 1)$, i.e., since then $\chi^\lambda(C_\mu) = \chi^\lambda(\text{Id}_{S_n})$ i.e., the dimension of the irreducible representation $S^\lambda$ of $S_n$ indexed by $\lambda$. On the other hand, a rim-hook tableau of content $\mu$ is just a standard tableau. So the Murnaghan-Nakayama Rule implies the following:

**Corollary 10.14.2.** $\dim S^\lambda = f^\lambda$.

So, how do we calculate $f^\lambda$?

10.15 The Hook-Length Formula

As before, let $\lambda = (\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_\ell > 0) \vdash n$, and denote by $\text{SYT}(\lambda)$ the set of standard tableaux of shape $\lambda$, so $f^\lambda = |\text{SYT}(\lambda)|$. For each cell $c$ in row $i$ and column $j$ of the Ferrers diagram of $\lambda$, let $h(c) = h(i, j)$ denote its hook length\(^5\): the number of cells due east of, due south of, or equal to $c$. In the following example, $h(c) = h(2, 3) = 6$.

---

\(^5\)In this section, the letter $h$ always refers to hook lengths, never to the complete homogeneous symmetric function!
Explicitly,
\[ h(i, j) = \lambda_i - (i - 1) + \tilde{\lambda}_j - (j - 1) - 1 = \lambda_i + \tilde{\lambda}_j - i - j + 1 \] 

(10.15.1)

where \( \tilde{\lambda} \) is the conjugate partition to \( \lambda \).

The number of standard tableaux is given by the following amazing formula, first proved by Frame, Robinson, and Thrall [FRT54].

**Theorem 10.15.1 (Hook-Length Formula).** Let \( \lambda \vdash n \). Then

\[ f^\lambda = \frac{n!}{\prod_{c \in \lambda} h(c)}. \]

**Example 10.15.2.** For \( \lambda = (5, 4, 3, 3, 1) \vdash 16 \) as above, the tableau of hook lengths is

\[
\begin{array}{ccccc}
9 & 7 & 6 & 3 & 1 \\
7 & 5 & 4 & 1 \\
5 & 3 & 2 \\
4 & 2 & 1 \\
1 \\
\end{array}
\]

so \( f^\lambda = 14!(9 \cdot 7^2 \cdot 6 \cdot 5^2 \cdot 4^2 \cdot 3^2 \cdot 2^2 \cdot 1^4) = 2288 \). 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

\[ f^\lambda = \frac{(2n)!}{(n+1)! n!} = \frac{1}{n+1} \left( \begin{array}{c} 2n \\ n \end{array} \right), \]

the \( n^{th} \) Catalan number (as we already know).

There are many proofs of the hook-length formula in the literature. My personal favorite is the following one, due to Greene, Nijenhuis and Wilf [GNW79].

**Proof of the Hook-Length Formula [GNW79].** First, a little notation. If \( c = (i, j) \) is a cell of a tableau \( T \), then we write \( T(c) \) or \( T(i, j) \) for the entry in that cell. Also, for any partition \( \lambda \), define

\[ F(\lambda) = \frac{n!}{\prod_{c \in \lambda} h(c)}. \]

Let \( T \) be a standard tableau of shape \( \lambda \vdash n \). The cell \( c \in T \) containing the number \( n \) must occur at an outside corner (i.e., it is the rightmost element in its row and the bottom element 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 bijection

\[ \{ T \in \text{SYT}(\lambda) \mid T(c) = n \} \to \text{SYT}(\lambda - c). \]
The proof proceeds by induction on \( n \). The base case \( n = 1 \) is clear, and by the bijection just described, the inductive step is equivalent to showing that
\[
F(\lambda) = \sum_c F(\lambda - c),
\]
the sum over all corners. Abbreviating \( F = F(\lambda) \) and \( F_c = F(\lambda - c) \), we can rewrite this equation as
\[
\sum_{\text{corners } c} \frac{F_c}{F} = 1 \tag{10.15.2}
\]
and this is what we plan to prove.

Consider the following random process (called a hook walk). First choose a cell \((a_0, b_0)\) uniformly from \( \lambda \). Then for each \( t = 1, 2, \ldots \), choose a cell \((a_t, b_t)\) uniformly from all other cells in the hook (so that \((a_t, b_t)\) is either strictly to the right of or below \((a_{t-1}, b_{t-1})\)). The process stops when it reaches a corner; let \( p(c) \) be the probability of reaching a particular corner \( c \). Evidently \( \sum_c p_c = 1 \). The goal is now to show that
\[
p_c = \frac{F_c}{F} \tag{10.15.3}
\]
which will establish (10.15.2).

Let \( c = (x, y) \). Removing \( c \) decreases by 1 the sizes of the hooks with corners strictly north or west of \( c \), and leaves all other hook sizes unchanged. Therefore
\[
\frac{F_c}{F} = \frac{(n-1)!}{n!} \prod_{i=1}^{x-1} \frac{1}{h(i, y)} \prod_{j=1}^{y-1} \frac{1}{h(x, j)} \prod_{i=x+1}^{\infty} \frac{1}{h(i, y)} \prod_{j=y+1}^{\infty} \frac{1}{h(x, j)}
\]
\[
= \frac{1}{n} \prod_{i=1}^{x-1} \left( 1 + \frac{1}{h(i, y) - 1} \right) \prod_{j=1}^{y-1} \left( 1 + \frac{1}{h(x, j) - 1} \right). \tag{10.15.4}
\]

For a hook walk \( W = (a_1, b_1), \ldots, (a_m, b_m) = (x, y) \), let \( A = \{a_1, \ldots, a_m\} \) and \( B = \{b_1, \ldots, b_m\} \) be the sets of rows and columns encountered; these are called 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
\[
p(A, B \mid a, b) = \left( \prod_{i \in A \setminus x} \frac{1}{h(i, y) - 1} \right) \left( \prod_{j \in B \setminus y} \frac{1}{h(x, j) - 1} \right) \phi. \tag{10.15.5}
\]
We prove this by induction on \( m \). If \( m = 1 \) then the equation says that \( 1 = 1 \) since the RHS is the empty product. If \( m > 1 \), then
\[
p(A, B \mid a, b) = \frac{1}{h(a, b) - 1} \left[ p(A \setminus a_1, B \mid a_2, b_1) + p(A, B \setminus b_1 \mid a_1, b_2) \right]
\]
\[
= \frac{1}{h(a, b) - 1} \left[ (h(a, y) - 1)\phi + (h(x, b) - 1)\phi \right] \quad \text{(by induction)}
\]
\[
= \left( \frac{h(a, y) - 1 + h(x, b) - 1}{h(a, b) - 1} \right) \phi
\]
\[
= \phi.
\]
To see that the parenthesized expression is 1, consider the following diagram, with the hooks at \((a,y)\) and \((x,b)\) shaded in red and blue respectively, with the square \((x, y)\) omitted so that there are a total of \(h(a, y) - 1 + h(x, b) - 1\) shaded squares. Pushing some red squares north and some blue squares to the left produces the hook at \((a, b)\) with one square omitted, as on the right.

This proves (10.15.5). Now we compute \(p_c\), the probability that a walk ends at a particular corner \(c = (x, y)\). In this case we will have \(x = \max A\) and \(y = \max B\). Summing over all possible starting positions, we have

\[
p_c = \frac{1}{n} \sum_{(A, B, a, b): A \subseteq [x], B \subseteq [y], a = \min A, b = \min B, x = \max A, y = \max B} \left( \prod_{i \in A \setminus x} \frac{1}{h(i, y) - 1} \right) \left( \prod_{j \in B \setminus y} \frac{1}{h(x, j) - 1} \right)
\]

(by (10.15.5))

\[
= \frac{1}{n} \sum_{A \subseteq [x - 1], B \subseteq [y - 1]} \left( \prod_{i \in A} \frac{1}{h(i, y) - 1} \right) \left( \prod_{j \in B} \frac{1}{h(x, j) - 1} \right).
\]

But this is precisely the expansion of the right-hand side of (10.15.4)! This establishes that \(p_c = F_c / F\) (10.15.3) and completes the proof.

10.16 The Jacobi-Trudi determinant definition of the Schur functions

There is a formula for the Schur function \(s_\lambda\) as a determinant of a matrix whose entries are \(h_n\)'s or \(e_n\)'s, with an elegant proof due to the ideas of Lindström, Gessel, and Viennot. This exposition follows closely that of [Sag01, §4.5].

We adopt the convention that \(h_0 = e_0 = 1\), and that \(h_k = e_k = 0\) for \(k < 0\).

**Theorem 10.16.1.** For any \(\lambda = (\lambda_1, \ldots, \lambda_\ell)\) we have

\[
s_\lambda = \left| h_{\lambda_i - i + j} \right|_{i, j = 1, \ldots, \ell}
\]

and

\[
s_\lambda = \left| e_{\lambda_i - i + j} \right|_{i, j = 1, \ldots, \ell}.
\]
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 of (10.16.1).**  **Step 1:** For \( n \in \mathbb{N} \), express \( h_n \) and \( e_n \) as generating functions 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 \((x, \infty)\) for some \( x \in \mathbb{N} \). Label each eastward step \( e \) of \( P \) by the number \( L(e) \) that is its \( y \)-coordinate plus one, as in Figure 10.1.

![Figure 10.1: A lattice path \( P \) from (1,0) to (6,\infty) (blue) and the labeling of its east steps (red). Here \( x^P = x_1 x_2^2 x_4 x_6 \).](image)

Assign \( P \) the monomial weight
\[
x^P = \prod_{e} x_{L(e)} \in \mathbb{Q}[x_1, x_2, \ldots].
\]

Note that the path \( P \) can be recovered from the monomial \( x^P \) (up to horizontal shifting), and that \( x^P \) can be any monomial. Thus we have a bijection, and it follows that for any \((a,b) \in \mathbb{N}^2\)
\[
h_n = \sum_{\text{paths } P \text{ from } (a,b) \text{ to } (a+n,\infty)} x^P.
\] (10.16.3)

**Step 2:** For a partition \( \lambda = (\lambda_1, \ldots, \lambda_\ell) \), express \( h_\lambda \) and \( e_\lambda \) as generating functions for families of lattice paths.

Let \( U = (u_1, \ldots, u_\ell) \), \( V = (v_1, \ldots, v_\ell) \) be ordered \( \ell \)-tuples of lattice points in \( \mathbb{Z} \times (\mathbb{Z} \cup \{ \infty \}) \). A **\( U,V \)-lattice path family** is a tuple \( P = (\pi, P_1, \ldots, P_\ell) \), where \( \pi \in \mathcal{S}_\ell \) and each \( P_i \) is a path from \( u_{\pi(i)} \) to \( v_i \). Define
\[
x^P = \prod_{i=1}^{\ell} x^{P_i}, \quad (-1)^{\pi} = \varepsilon(\pi)
\]
where \( \varepsilon \) denotes the sign of the permutation \( \pi \).

For a partition \( \lambda \) of length \( \ell \), a \( \lambda \)-path family is a \((U,V)\)-path family, where \( U, V \) are defined by

\[
u_i = (\ell - i, 0), \quad v_i = (\lambda_i + \ell - i, \infty).
\]

for \( 1 \leq i \leq \ell \). For instance, if \( \lambda = (3, 3, 2, 1) \) then

\[
U = ((3, 0), (2, 0), (1, 0), (0, 0)), \quad V = ((6, \infty), (5, \infty), (3, \infty), (1, \infty))
\]

and Figure 10.2 shows a \( \lambda \)-path family (in which \( \pi = 3124 \) and so \((-1)^{P} = 1\)).

For each \( i \in [\ell] \), the path \( P_i \) from \( u(\pi(i)) \) to \( v_i \) has

\[
(\lambda_i + \ell - i) - (\ell - \pi(i)) = \lambda_i - i + \pi(i)
\]
estep steps. Therefore, expanding the determinant on the right-hand side of (10.16.1), we see that

\[
|h_{\lambda, -i+j}|_{i,j=1,\ldots,\ell} = \sum_{\pi \in \mathfrak{S}_\ell} \varepsilon(\pi) \prod_{i=1}^{\ell} h_{\lambda_i, -i+\pi(i)} \quad \text{(the usual determinant expansion)}
\]

\[
= \sum_{\pi \in \mathfrak{S}_\ell} \varepsilon(\pi) \prod_{i=1}^{\ell} \left( \sum_{\text{paths } P_i \text{ from } u(\pi(i)) \text{ to } v_i} x^{P_i} \right) \quad \text{(by (10.16.3))}
\]

\[
= \sum_{\pi \in \mathfrak{S}_\ell} \varepsilon(\pi) \sum_{\lambda\text{-path families } P = (P_1, \ldots, P_\ell)} x^{P_1} \cdots x^{P_\ell}
\]

\[
= \sum_{P} (-1)^{P} x^{P}.
\] (10.16.4)
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 $\pi$ must be the identity permutation, and in particular $(-1)^{\mathbf{P}} = 1$.

Step 3: Show that all the bad summands in (10.16.4) cancel out.

Define a sign-reversing, weight-preserving involution $\mathbf{P} \mapsto \mathbf{P}^\sharp$ on non-good $\lambda$-path families as follows.

1. Of all the lattice points contained in two or more paths in $\mathbf{P}$, choose the point $\alpha$ with the lex-greatest pair of coordinates.
2. Of all the half-paths from $\alpha$ 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 10.3.

Figure 10.3: The involution $\mathbf{P} \leftrightarrow \mathbf{P}^\sharp$ on bad path families.

Then

- 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 (10.16.4) we have

$$|h_{\lambda, -i+j}|_{i,j=1,\ldots,\ell} = \sum_{\mathbf{P}} (-1)^{\mathbf{P}} x^{\mathbf{P}} = \sum_{\mathbf{P} \text{ good}} x^{\mathbf{P}}. \quad (10.16.5)$$

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 each $j$ and $i < i'$, the $j^{th}$ step of the path $P_i$ is strictly southeast of the $j^{th}$ step of $P_{i'}$. Therefore, we can construct a column-strict tableau of shape $\lambda$ by reading off the labels of each path, and this gives a bijection between good $\lambda$-path families and column-strict tableaux of shape $\lambda$. An example is shown in Figure 10.4.
Consequently, (10.16.5) implies that \(|h_{\lambda_{i-j}}|_{i,j=1,...,\ell} = s_{\lambda}\), as desired.

The proof of (10.16.2) (the expression for \(s_{\lambda}\) as a determinant of \(e\)’s) proceeds along similar lines. The key difference is that instead of labeling each east step with its height, we number all the steps (north and east) consecutively, starting at 1, and define the weight of a path to be 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 (10.16.3). When we pass from paths to path families, we ignore the first \(i-1\) steps of \(P_i\) (which must all be northward). Bad path families cancel out by the same involution as before, and good path families now give rise to tableaux of shape \(\lambda\) in which rows strictly increase but columns weakly increase, so that transposing gives a column-strict tableau of shape \(\tilde{\lambda}\).

Corollary 10.16.2. For every partition \(\lambda\), the involution \(\omega\) interchanges \(s_{\lambda}\) and \(s_{\tilde{\lambda}}\).

Proof. We know that \(\omega\) interchanges \(h_{\lambda}\) and \(e_{\lambda}\), so it interchanges the RHS’s, hence the LHS’s, of (10.16.1) and (10.16.2).

10.17 Quasisymmetric functions

Definition 10.17.1. A quasisymmetric function is a formal power series \(F \in \mathbb{C}[[x_1, x_2, \ldots]]\) 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 \(\alpha_1, \ldots, \alpha_r \in \mathbb{N}\), 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 \(\mu\) 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 \( QSym \subseteq \mathbb{C}[[x]] \). We now describe a vector space basis for \( QSym \).

A composition \( \alpha \) is a sequence \((\alpha_1, \ldots, \alpha_r)\) 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 \( \alpha \vdash n \); the set of all compositions of \( n \) will be denoted \( \text{Comp}(n) \). 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

\[
S(\alpha) = \{\alpha_1, \alpha_1 + \alpha_2, \ldots, \alpha_1 + \cdots + \alpha_{r-1}\}.
\]

The map \( \alpha \mapsto S(\alpha) \) is a bijection from compositions of \( n \) to subsets of \([n-1]\); in particular, \( |\text{Comp}(n)| = 2^{n-1} \).

We can define a partial order on \( \text{Comp}(n) \) 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 \( \alpha = (\alpha_1, \ldots, \alpha_r) \vdash n \) is the power series

\[
M_{\alpha} = \sum_{i_1 < \cdots < i_r} x_{i_1}^{\alpha_1} \cdots x_{i_r}^{\alpha_r} \in \mathbb{Z}[[x_1, x_2, \ldots]]_n.
\]

(For example, the quasisymmetric function \( \sum_{i<j} x_i^2 x_j \); mentioned above is \( M_{21} \).) This is the sum of all the monomials whose coefficient is constrained by the definition of quasisymmetry to be the same as that of any one of them. Therefore, the set \( \{M_{\alpha}\} \) is a graded basis for \( QSym \).

**Example 10.17.2.** Let \( \mathcal{M} \) be a matroid on ground set \( E \) of size \( n \). Consider weight functions \( f : E \to \mathbb{F} \); 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 \( B \cup \{e_i\} \) is independent, then replace \( B \) with \( B \cup \{e_i\} \)). 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

\[
W(U_2(3)) = \sum_{i<j<k} 6x_ix_jx_k + \sum_{i<j} 3x_i^2x_j = 6M_{111} + 3M_{12}.
\]

Questions: How are \( W(\mathcal{M}) \) and \( W(\mathcal{M}^*) \) related?

### 10.18 Exercises

**Exercise 10.1.** Let \( \lambda \vdash n \). Verify that \( |C_\lambda| = n!/z_\lambda \), where \( z_\lambda \) is defined as in (10.8.1).
Exercise 10.2. Prove Proposition 10.6.1, namely that the coefficient of a monomial $m_{\mu}$ in the power-sum symmetric function $p_{\lambda}$ is given by the number $\chi_{\mu}(C_{\lambda})$ defined in Section 9.11. Include a nontrivial example to illustrate your argument.

Exercise 10.3. 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 (10.8.4), and compute the coefficient of $x^k$ for each $k$. Hint: Interpret the coefficients as counting permutations.

Exercise 10.4. Supply the proofs for the identities (10.8.3), i.e.,

$$\Omega^* = \sum_{\lambda} e_\lambda(x)m_\lambda(y) = \sum_{\lambda} e_\lambda p_\lambda(x)p_\lambda(y).$$

Exercise 10.5. Prove part (6) of Theorem 10.12.2.

Exercise 10.6. Fill in the proofs of the underlined assertions in Rule 2 and Rule X for the alternate RSK algorithm in Section 10.11.

Exercise 10.7. For this problem, you will probably want to use one of the alternate RSK algorithms from Sections 10.10 and 10.11.

(a) For $w \in S_n$, 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 $Q(w^*) = Q(w)^T$? Computation indicates that the number of such permutations is

$$\begin{cases} 
2^{(n-1)/2}(n-1)! & \text{if } n \text{ is odd,} \\
\frac{(n-1)/2)!^2}{2^{(n-1)/2}} & \text{if } n \text{ is even,}
\end{cases}$$

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 $q_1, q_2, \ldots = (1, 2, 2, 12, 24, 136, 344, 2872, 7108, \ldots)$, where $q_n = \#\{w \in S_n \mid Q(w^*) = Q(w)^T\}$, does not seem to appear in the Online Encyclopedia of Integer Sequences.
Chapter 11

More Topics
11.1 Oriented matroids

This section barely scratches the surface of an extremely important subject, on which the canonical work is [BLVS+99]. See also the lecture notes [Rei].

11.1.1 Big face lattices of hyperplane arrangements

Consider the following two hyperplane arrangements $A_1, A_2$ in $\mathbb{R}^2$, with lines $H_1, \ldots, H_5$ (abbreviated $1, \ldots, 5$ in the diagrams):

Their intersection posets are isomorphic:

$$L(A_1) \cong L(A_2) =$$

Therefore, by Zaslavsky’s theorems they have the same numbers of regions and bounded regions. However, there is good reason not to consider $A_1, A_2$ isomorphic as arrangements. For example, both two bounded regions in $A_1$ are triangles, while in $A_2$ there is a triangle and a trapezoid. Also, the point $H_3 \cap H_2 \cap H_4$ lies between the lines $H_3$ and $H_5$ in $A_1$, while it lies below both of them in $A_2$. The intersection poset lacks the power to model geometric data like “between,” “below,” and “trapezoid,” as does matroid theory. Accordingly, we need to define a stronger combinatorial invariant.

In general, let $A = \{H_1, \ldots, H_n\}$ be an essential hyperplane arrangement in $\mathbb{R}^d$, with normal vectors $n_1, \ldots, n_n$. For each $i$, let $\ell_i$ be an affine linear functional on $\mathbb{R}^n$ such that $H_i = \{x \in \mathbb{R}^d \mid \ell_i(x) = 0\}$.

The intersections of hyperplanes in $A$, 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 $e \leq d$. We can encode each cell by recording whether the linear functionals $\ell_1, \ldots, \ell_n$ are positive, negative or zero on it. Specifically, for
\( k = (k_1, \ldots, k_n) \in \{+, -, 0\}^n \), let

\[
F = F(k) = \left\{ x \in \mathbb{R}^d \mid \begin{array}{ll}
\ell_i(x) > 0 & \text{if } k_i = + \\
\ell_i(x) < 0 & \text{if } k_i = - \\
\ell_i(x) = 0 & \text{if } k_i = 0
\end{array} \right\}
\]

Some notation: we write \( k_+ = \{i \mid k_i = +\} \) and \( k_- = \{i \mid k_i = -\} \). A convenient shorthand is to represent \( k_i \) by a list of digits, using \( i \) and \( \bar{i} \) indicate respectively that \( k_i = + \) or \( k_i = - \), with \( i \) omitted if \( k_i = 0 \). For instance, \( k = 0 + -00 + -0 \) would be written \( 2 \bar{3} \bar{6}7 \); here \( k_+ = \{2, 7\} \) and \( k_- = \{3, 6\} \).

If \( F \neq \emptyset \) then it is called a **face** of \( \mathcal{A} \), and \( k = k(F) \) is the corresponding **covector**. The set of all faces is denoted \( \mathcal{F}(\mathcal{A}) \). The poset \( \mathcal{F}(\mathcal{A}) = \mathcal{F}(\mathcal{A}) \cup \{0, 1\} \), ordered by containment of closures (\( F \leq F' \) if \( \bar{F} \subseteq \bar{F}' \)), is a lattice, called the **big face lattice of \( \mathcal{A} \)** (that is, the big lattice of faces, not the lattice of big faces!). If \( \mathcal{A} \) is central, then \( \mathcal{F}(\mathcal{A}) \) already has a unique minimal element and we don’t bother adding an additional \( 0 \). For example, if \( \mathcal{A} = \mathcal{B}_2 = \{y = 0, x = 0\} \subseteq \mathbb{R}^2 \), then the big face lattice is as follows:

Combinatorially, the order relation in \( \mathcal{F}(\mathcal{A}) \) is given by \( k \leq k' \) if \( k_+ \subseteq k'_+ \) and \( k_- \subseteq k'_- \). This is very easy to read in digital notation. The big face lattice is able to capture 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’ll see one soon.)

The maximal covectors (or **topes**) are precisely those with no zeroes; they correspond to the regions of \( \mathcal{A} \).

### 11.1.2 Oriented matroids

Consider the linear forms \( \ell_i \) that were used in representing each face by a covector. Recall that specifying \( \ell_i \) is equivalent to specifying a normal vector \( n_i \) to the hyperplane \( H_i \) (with \( \ell_i(x) = n_i \cdot x \)). As we know, the vectors \( n_i \) represent a matroid whose lattice of flats is precisely \( L(\mathcal{A}) \). Scaling \( n_i \) (equivalently, \( \ell_i \)) by a nonzero constant \( \lambda \in \mathbb{R} \) has no effect on the matroid represented by the \( n_i \)'s, but what does it do to the covectors? If \( \lambda > 0 \), then nothing happens, but if \( \lambda < 0 \), then we have to switch + and − signs in the \( i^{th} \) position of every covector. So, in order to figure out the covectors, we need not just the normal vectors \( n_i \), 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.
**Example 11.1.1.** Consider our running example of $A_1$ and $A_2$. In the following figure, the red arrows point from each hyperplane to its positive halfspace, and the blue strings are the maximal covectors.

**Proposition 11.1.2.** The maximal covectors whose negatives are also covectors are precisely those that correspond to unbounded faces. In particular, $A$ is central if and only if every negative of a covector is a covector.

The proof, as you can see, is presently omitted.

In order to describe circuits of oriented matroids, let us cone $A_1$ and $A_2$ to obtain central, essential arrangements $C A_1, C A_2$ in $\mathbb{R}^3$, whose normal matroids can be represented respectively by the matrices

$$
\begin{bmatrix}
 1 & -1 & 0 & 0 & 0 \\
 1 & 1 & 1 & 1 & 1 \\
 0 & 0 & 1 & 0 & -1
\end{bmatrix} \quad \begin{bmatrix}
 1 & -1 & 0 & 0 & 0 \\
 1 & 1 & 1 & 1 & 1 \\
 0 & 0 & 1 & 0 & 2
\end{bmatrix}
$$

The two matroids are both isomorphic, with circuit system $\{124, 345, 1235\}$. However, let’s look a little more closely. The minimal linear dependencies realizing the circuits in each case are

$$
\begin{align*}
&n_1 + n_2 - 2n_4 = 0 \\
&n_3 - 2n_4 + n_5 = 0 \\
&n_1 + n_2 - n_3 - n_5 = 0
\end{align*}
\quad \begin{align*}
&n'_1 + n'_2 - 2n'_4 = 0 \\
&2n'_3 - n'_4 - n'_5 = 0 \\
&n'_1 + n'_2 - 4n'_5 + 2n'_5 = 0
\end{align*}
$$

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 $124$ is a circuit in both cases. However, in the first case $345$ is a circuit, while in the second it is $345$. Note that if $c$ is a circuit then so is $-c$, where, e.g., $-124 = \overline{124}$. In summary, the oriented circuit systems for $CA_1$ and $CA_2$ are respectively

$$
\mathcal{C}_1 = \{124, \overline{124}, 345, 345, 1235, \overline{1235}\},
\mathcal{C}_2 = \{124, \overline{124}, 345, 345, 1235, \overline{1235}\}.
$$

Oriented circuits are minimal obstructions to covector-ness. For example, $124$ is a circuit of $A_1$ because the linear functionals defining its hyperplanes satisfy $\ell_1 + \ell_2 - 2\ell_4 = 0$. But if a covector of $A_1$ contains $124$, then any point in the corresponding face of $A$ would have $\ell_1, \ell_2, -\ell_4$ all positive, which is impossible.

Oriented circuits can be axiomatized.
Definition 11.1.3. Let $n$ be a positive integer. A **oriented circuit system** is a collection $\vec{C}$ of $n$-tuples $c \in \{+,-,0\}$ satisfying the following properties:

1. $00 \cdots 0 \notin \vec{C}$.
2. If $c \in \vec{C}$, then $-c \in \vec{C}$.
3. If $c,c' \in \vec{C}$ and $c \neq c'$, then either $c_+ \nsubset c'_-$ or $c_- \nsubset c'_+$.
4. If $c,c' \in \vec{C}$ and $c \neq c'$, and there is some $i$ with $c_i = +$ and $c'_i = -$, then there exists $d \in \vec{C}$ with $d_i = 0$, and, for all $j \neq i$, $d_+ \subseteq c_+ \cup c'_+$ and $d_- \subseteq c_- \cup c'_-$.

Again, the idea is to record not just the linearly dependent subsets of a set $\{\ell_i, \ldots, \ell_n\}$ of linear forms, but also the sign patterns of the corresponding linear dependences ("syzygies").

Condition (1) says that the empty set is linearly independent.

Condition (2) says that multiplying any syzygy by $-1$ gives a syzygy.

Condition (3), as in the definition of the circuit system of an (unoriented) matroid, must hold if we want circuits to record syzygies with minimal support.

Condition (4) is the oriented version of circuit exchange. Suppose that we have two syzygies

$$
\sum_{j=1}^{n} \gamma_j \ell_j = \sum_{j=1}^{n} \gamma'_j \ell_j = 0,
$$

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

$$
\sum_{j=1}^{n} \delta_j \ell_j = 0,
$$

where $\delta_j = \gamma_j + \gamma'_j$. In particular, $\delta_i = 0$, and $\delta_j$ is positive (resp., negative) if and only if at least one of $\gamma_j, \gamma'_j$ is positive (resp., negative).

**Remark 11.1.4.** If $\vec{C}$ is an oriented circuit system, then $\mathcal{C} = \{c_+ \cup c_- \mid c \in \vec{C}\}$ 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 $\vec{C}$.

As in the unoriented setting, the circuits of an oriented matroid represent minimal obstructions to being a covector. That is, every real hyperplane arrangement $A$ gives rise to an oriented circuit system $\vec{C}$ such that if $k$ is a covector of $A$ and $c$ is a circuit, then it is not the case that $k_+ \supseteq c_+$ and $k_- \supseteq c_-$. 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 4.5.7). 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.

11.1.3 Oriented matroids from graphs

Recall (§4.3) 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 \( \vec{C} \) is the family of oriented cycles. This is best shown by an example.

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 \( \mathcal{C}_1 \) seen previously. More generally, for every oriented graph \( \vec{G} \), the signed set system \( \vec{\mathcal{C}} \) formed in this way satisfies the axioms of Definition 11.1.3. To understand axiom (4) of that definition, suppose \( e \) is an edge that occurs forward in \( c \) and backward in \( c' \). Then \( c - e \) and \( c' - e \) 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 \( \mathcal{C}_2 \).

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 $ij$ is an edge in $\vec{G}$ then the hyperplane $x_i = x_j$ is assigned the normal vector $e_i - e_j$. The maximal covectors are precisely the regions of the graphic arrangement.

11.2 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 11.2.1.** 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 $c$.](image_url)

**Definition 11.2.2.** A flow on $N$ is a function $f : E \to \mathbb{N}$ that satisfies the capacity constraints

$$0 \leq f(e) \leq c(e) \quad \forall e \in E$$

and the conservation constraints

$$f^-(v) = f^+(v) \quad \forall v \in V \setminus \{s,t\}$$

191
where
\[ f^-(v) = \sum_{e = \overrightarrow{uv}} f(e), \quad f^+(v) = \sum_{e = \overleftarrow{uv}} f(e). \]

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 (11.2.1). 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 11.2.3.** 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
\[ [S, T] = \{ \overrightarrow{xy} \in E \mid x \in S, y \in T \} \]
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 \( f^-(A) = \sum_{e \in [A, \overline{A}]} f(e) \), \( f^+(A) = \sum_{e \in [A, \overline{A}]} f(e) \).

**Proposition 11.2.4.** Let \( f \) be a flow, and let \( A \subseteq V \). Then:
\[ f^+(A) - f^-(A) = \sum_{v \in A} (f^+(v) - f^-(v)). \] (11.2.3a)

In particular, if \( [S, T] \) is a cut, then
\[ f^+(S) - f^-(S) = f^-(T) - f^+(T) = |f|, \] (11.2.3b)
\[ |f| \leq c(S, T). \] (11.2.3c)

The proof (which requires little more than careful bookkeeping) is left as an exercise.

The inequality (11.2.3c) 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.
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:

Fortunately, there is a more general way to increase the value of a flow. The key idea is that flow along an edge $x\rightarrow y$ 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 $e$.

1. If possible, find an augmenting path, i.e., a sequence of edges and vertices

   $P : x_0 = s, \ x_1, \ x_2, \ \ldots, \ x_{n-1}, \ e_n, \ x_n = t$

   such that the $x_i$ are distinct and for every $i$, $i = 0, \ldots, n-1$, either
   - $e_i = x_{i-1}x_i \in E$, and $f(e_i) < c(e_i)$ ("$e_i$ is a forward edge"); or
   - $e_i = x_ix_{i-1} \in E$, 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 $P$; 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 11.2.5. Suppose that $f$ is a flow that has no augmenting path. Let 

$$S = \{ v \in V \mid \text{there is an augmenting path from } s \text{ to } v \}, \quad T = V \setminus S.$$ 

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 (11.2.3b) gives 

$$|f| = f^+(S) - f^-(S) = \sum_{e \in [S,S]} f(e) - \sum_{e \in [S,S]} f(e) = \sum_{e \in [S,S]} f(e).$$

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. \qed

We have proven:

Theorem 11.2.6 (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.)

**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 $f_2$. 

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 $P$; 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 11.2.5. Suppose that $f$ is a flow that has no augmenting path. Let 

$$S = \{ v \in V \mid \text{there is an augmenting path from } s \text{ to } v \}, \quad T = V \setminus S.$$ 

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 (11.2.3b) gives 

$$|f| = f^+(S) - f^-(S) = \sum_{e \in [S,S]} f(e) - \sum_{e \in [S,S]} f(e) = \sum_{e \in [S,S]} f(e).$$

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. \qed

We have proven:

Theorem 11.2.6 (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 11.2.7.** 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 $e \in E$,

$$f(e) = \# \{i \mid e \in P_i \}.$$  

(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 11.2.8.** 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 $k = \min \{ f(e) \mid e \in C \}$. 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 $|f|$. \qed

**Remark 11.2.9.** 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.

### 11.3 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 11.3.1.** 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 11.3.2** (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 “≤” 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 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 $P^+ = \{x \in P \mid x \geq a \text{ for some } a \in A\}$, $P^- = \{x \in P \mid x \leq a \text{ for some } a \in A\}$.

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

$$\{C_a^+ \cap C_a^- \mid a \in A\}$$

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 11.3.3** (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 $A$ is an antichain cover, then no antichain in $A$ can contain more than one element of $C$, so $|A| \geq |C|$. On the other hand, let

$$A_i = \{x \in P \mid \text{the longest chain headed by } x \text{ has length } i\};$$

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 \cite{GK76, Gre76}. An excellent source on this topic, including multiple proofs, is the survey article \cite{BF01} by Thomas Britz and Sergey Fomin.

**Theorem 11.3.4** (Greene-Kleitman). Let $P$ be a finite poset. Define two sequences of positive integers

$$\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k), \quad \mu = (\mu_1, \mu_2, \ldots, \mu_m)$$

by

$$\lambda_1 + \cdots + \lambda_k = \max \{|C_1 \cup \cdots \cup C_k| : C_i \subseteq P \text{ chains }\}, \quad \mu_1 + \cdots + \mu_k = \max \{|A_1 \cup \cdots \cup A_k| : A_i \subseteq P \text{ disjoint antichains }\}.$$ 

Then:

\[\text{I.e., a chain of size } m(P) \text{ — not merely a chain that is maximal with respect to inclusion, which might have smaller cardinality.}\]
1. \( \lambda \) and \( \mu \) are both partitions of \(|P|\), i.e., weakly decreasing sequences whose sum is \(|P|\).
2. \( \lambda \) and \( \mu \) are conjugates (written \( \mu = \tilde{\lambda} \)): the row lengths of \( \lambda \) are the column lengths in \( \mu \), and vice versa.

Note that Dilworth’s Theorem is just the special case \( \mu_1 = \ell \). As an example, the poset with Hasse diagram

![Hasse Diagram](image)

has

\[
\lambda = (3, 2, 2, 2) = \begin{array}{cccc}
\text{ } & \text{ } & \text{ } & \\
\text{ } & \text{ } & \text{ } & \\
\text{ } & \text{ } & \text{ } & \\
\text{ } & \text{ } & \text{ } & \\
\text{ } & \text{ } & \text{ } & \\
\text{ } & \text{ } & \text{ } & \\
\text{ } & \text{ } & \text{ } & \\
\end{array}
\quad \text{and} \quad
\mu = (4, 4, 1) = \begin{array}{cccc}
\text{ } & \text{ } & \text{ } & \\
\text{ } & \text{ } & \text{ } & \\
\text{ } & \text{ } & \text{ } & \\
\text{ } & \text{ } & \text{ } & \\
\end{array} = \tilde{\lambda}.
\]

### 11.4 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.

![Necklace Diagram](image)

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

\[
\binom{7}{4, 2, 1} = \frac{7!}{4! \cdot 2! \cdot 1!} = 105.
\]

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 \( \alpha : G \rightarrow \mathfrak{S}_X \), the group of permutations of \( X \).

Equivalently, an action can also be regarded as a map \( G \times X \rightarrow X \), sending \((g, x)\) to \( gx \), such that

- \( \text{Id}_G x = x \) for every \( x \in X \) (where \( \text{Id}_G \) 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 \text{ is the set } O_x = \{gx \mid g \in G\} \subseteq X \text{ and its stabilizer is } S_x = \{g \in G \mid gx = x\} \subseteq G, \text{ 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 \text{ is the set of all 105 necklaces, and the group acting on them is the dihedral group } D_7 \text{ (the group of symmetries of a regular heptagon). The number we are looking for is the number of orbits of } D_7.

**Lemma 11.4.1.** Let } x \in X. \text{ Then } |O_x||S_x| = |G|.

**Proof.** The element } gx \text{ depends only on which coset of } S_x \text{ contains } g, \text{ so } |O_x| \text{ is the number of cosets, which is } |G|/|S_x|.

**Proposition 11.4.2.** [Burnside’s Theorem] The number of orbits of the action of } G \text{ on } X \text{ equals the average number of fixed points:

\[
\frac{1}{|G|} \sum_{g \in G} \#\{x \in X \mid gx = x\}
\]

**Proof.** For a sentence } P, \text{ let } \chi(P) = 1 \text{ if } P \text{ is true, or 0 if } P \text{ is false (the “Garsia chi function”). Then

\[
\text{Number of orbits } = \sum_{x \in X} \frac{1}{|O_x|} = \frac{1}{|G|} \sum_{x \in X} |S_x| = \frac{1}{|G|} \sum_{x \in X} \sum_{g \in G} \chi(gx = x) = \frac{1}{|G|} \sum_{g \in G} \sum_{x \in X} \chi(gx = x) = \frac{1}{|G|} \sum_{g \in G} \#\{x \in X \mid gx = x\}.
\]

Typically, it is easier to count fixed points than to count orbits directly.

**Example 11.4.3.** We can apply this technique to the necklace example above.

- The identity of } D_7 \text{ has 105 fixed points.
- Each of the seven reflections in } D_7 \text{ 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 11.4.4.** 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 \text{ has } 3^7 = 2187 \text{ 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

$$\frac{k^7 + 7k^4 + 6k}{14}. \quad (11.4.1)$$

As this example indicates, it is helpful to look at the cycle structure of the elements of $G$, or more precisely on their images $\alpha(g) \in \Sigma_X$.

**Proposition 11.4.5.** Let $X$ be a finite set, and let $\alpha : G \to \Sigma_X$ 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,

$$\text{# equivalence classes of colorings} = \frac{1}{|G|} \sum_{g \in G} k^{\ell(g)}. \quad (11.4.2)$$

Let’s rephrase Example 11.4.4 in this notation. The identity has cycle-shape 1111111 (so $\ell = 7$); each of the six reflections has cycle-shape 2221 (so $\ell = 4$); and each of the seven rotations has cycle-shape 7 (so $\ell = 1$). Thus (11.4.1) is an example of the general formula (11.4.2).

**Example 11.4.6.** 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 $S_4$ that contains the identity, of cycle-shape 1111; the eight permutations of cycle-shape 31; and the three permutations of cycle-shape 22. Therefore, the number of colorings is

$$\frac{k^4 + 11k^2}{12}.$$ 

**11.5 Grassmannians**

A standard reference for everything in this and the following section is Fulton [Ful97].

Part of the motivations for the combinatorics of partitions and tableaux comes from classical enumerative geometric questions like this:

**Problem 11.5.1.** 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 the matroid represented by four direction vectors 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 $k$ (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 (short for Grassmannian variety) and denoted by $\text{Gr}(k, V)$ (warning: this notation varies considerably from source to source). As we’ll see, $\text{Gr}(k, V)$ has a lot of nice properties:

- It is a smooth manifold of dimension $k(n-k)$ over $k$.
- It can be decomposed into pieces, called Schubert cells, each of which is naturally diffeomorphic to $k^j$, for some appropriate $j$.
- Here’s where combinatorics comes in: the Schubert cells correspond to the interval $Y_{n,k} := [\emptyset, k^{n-k}]$ 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 $\lambda$ there is a corresponding Schubert cell $\Omega_{\lambda}$ of dimension $|\lambda|$ (the number of boxes in the Ferrers diagram).
- How these cells fit together topologically is described by $Y_{n,k}$ in the following sense: the closure of $\Omega_{\lambda}$ 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).
- Consequently, the Poincaré polynomial of $\text{Gr}(k, \mathbb{C}^n)$ (i.e., the Hilbert series of its cohomology ring) is the rank-generating function for the graded poset $Y_{n,k}$ — namely, the $q$-binomial coefficient $\binom{n}{k}_q$.

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 $W \in \text{Gr}(k, \mathbb{C}^n)$; that is, $W$ is a $k$-dimensional subspace of $V = \mathbb{C}^n$. We can describe $W$ as the column space of a $n \times k$ matrix $M$ of full rank:

\[ M = \begin{bmatrix} m_{11} & \cdots & m_{1k} \\ \vdots & \ddots & \vdots \\ m_{nk} & \cdots & m_{nk} \end{bmatrix}. \]

Is the Grassmannian therefore just the space $\mathbb{C}^{n \times k}$ of all such matrices? No, 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, we can write

\[ \text{Gr}(k, \mathbb{C}^n) = \mathbb{C}^{n \times k}/\text{GL}_k(\mathbb{C}). \quad (11.5.1) \]

---

If these terms don’t make sense, here’s what you need to know. Some of you will recognize that I have omitted lots of technical details from the explanation that is about to follow — that’s exactly the point. 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:

\[ \text{Poin}(X, q) = \sum_{i=0}^{d} (\dim_{\mathbb{Q}} H^i(X))q^i. \]

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 $H^*(X) \cong \mathbb{Q}[z]/(z^{d+1})$. Each graded piece $H^i(X)$, for $0 \leq i \leq d$, is a 1-dimensional $\mathbb{Q}$-vector space (generated by the monomial $z^i$), and $\text{Poin}(X, q) = 1 + q + q^2 + \cdots + q^d = (1 - q^{d+1})/(1 - q)$. In general, if $X$ is a compact orientable manifold, then Poincaré duality implies (among other things) that $\text{Poin}(X, q)$ is a palindrome.
That is, the $k$-dimensional subspaces of $k^n$ can be identified with the orbits of $\mathcal{F}^{n \times k}$ under the action of the general linear group $\text{GL}_k(k)$.

(In fact, as one should expect from (11.5.1),

$$\dim \text{Gr}(k, k^n) = \dim k^{n \times k} - \dim \text{GL}_k(k) = nk - k^2 = k(n - k)$$

where “dim” means dimension as a manifold over $k$. Technically, this dimension calculation does not follow from (11.5.1) alone; you need to know that the action of $\text{GL}_k(k)$ on $k^{n \times k}$ is suitably well-behaved. Nevertheless, we will soon be able to calculate the dimension of $\text{Gr}(k, k^n)$ more directly.)

Is there a canonical representative for each $\text{GL}_k(k)$-orbit? In other words, given $W \in \text{Gr}(k, k^n)$, can we find some “nicest” matrix whose column space is $W$? Yes: 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:

- 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 $i^{th}$ column has its pivot. Then $1 \leq p_1 < p_2 < \cdots < p_k \leq n$.
- 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:

$$
\begin{pmatrix}
1 & 0 \\
0 & 1 \\
0 & 0 \\
0 & 0
\end{pmatrix}
\quad
\begin{pmatrix}
1 & 0 \\
0 & * \\
0 & 0 \\
0 & 0
\end{pmatrix}
\quad
\begin{pmatrix}
1 & 0 \\
0 & 1 \\
0 & 0 \\
0 & 0
\end{pmatrix}
\quad
\begin{pmatrix}
* & * \\
0 & 1 \\
0 & 0 \\
0 & 0
\end{pmatrix}
\quad
\begin{pmatrix}
* & * \\
* & 1 \\
0 & 0 \\
0 & 0
\end{pmatrix}
\quad
\begin{pmatrix}
* & * \\
* & * \\
1 & 0 \\
0 & 1
\end{pmatrix}
$$

(11.5.2)

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 $k$ is infinite and you choose the entries of $M$ randomly (for a suitable definition of “random” — for a precise formulation, consult your local probabilist), then you will almost always end up with a matrix $M^*$ of the last form.

Definition 11.5.2. Let $0 \leq k \leq n$ and let $p = \{p_1 < \cdots < p_k\} \in \binom{[n]}{k}$ (i.e., $p_1, \ldots, p_k$ are distinct elements of $[n]$, ordered least to greatest). The Schubert cell $\Omega_p$ is the set of all elements $W \in \text{Gr}(k, k^n)$ such that, for every $i$, the $i^{th}$ column of $M(W)$ has its pivot in row $p_i$.

Theorem 11.5.3. 1. Every $W \in \text{Gr}(k, k^n)$ belongs to exactly one Schubert cell; that is, $\text{Gr}(k, k^n)$ is the disjoint union of the subspaces $\Omega_p$.

2. For every $p \in \binom{[n]}{k}$, there is a diffeomorphism

$$\Omega_p \xrightarrow{\sim} k^{|p|}$$

where $|p| = (p_1 - 1) + (p_2 - 2) + \cdots + (p_k - k) = p_1 + p_2 + \cdots + p_k - \binom{k+1}{2}$.

3. Define a partial order on $\binom{[n]}{k}$ as follows: for $p = \{p_1 < \cdots < p_k\}$ and $q = \{q_1 < \cdots < q_k\}$, set $p \geq q$ if $p_i \geq q_i$ for every $i$. Then

$$p \geq q \implies \overline{\Omega_p} \supseteq \Omega_q.$$  \hspace{1cm} (11.5.3)

4. The poset $\binom{[n]}{k}$ is isomorphic to the interval $Y_{k,n}$ in Young’s lattice.

5. $\text{Gr}(k, k^n)$ is a compactification of the Schubert cell $\Omega_{(n-k+1,n-k+2,\ldots,n)}$, which is diffeomorphic to $k^{k(n-k)}$. In particular, $\dim k \text{Gr}(k, k^n) = k(n - k)$.

The cell closures $\overline{\Omega_p}$ are called Schubert varieties.
Proof. (1) is immediate from the definition.

For (2), the map $\Omega_p \to k^{\lvert p \rvert}$ is given by reading off the $*$s in the reduced column-echelon form of $M(W)$. (For instance, let $n = 4$ and $k = 2$. Then the matrix representations in (11.5.2) give explicit diffeomorphisms of the Schubert cells of $Gr(k, k^n)$ to $C^0$, $C^1$, $C^2$, $C^3$, $C^4$ respectively.) The number of $*$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 $*$s is $\lvert p \rvert$.

For (3): This is best illustrated by an example. Consider the second matrix in (11.5.2):

$$M = \begin{bmatrix} 1 & 0 \\ 0 & z \\ 0 & 1 \\ 0 & 0 \end{bmatrix},$$

where I have replaced the entry labeled $*$ 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.,

$$\text{colspace } M = \text{colspace } \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1/z \\ 0 & 0 \end{bmatrix}.$$

Therefore, it makes sense to say that

$$\lim_{|z| \to \infty} \text{colspace } M = \text{colspace } \lim_{|z| \to \infty} M = \text{colspace } \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix},$$

which is the first matrix in (11.5.2). 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 $p$ corresponds to taking a limit of column spans in this way, so the covering relations in the poset $([n]_k)$ give containment relations of the form (11.5.3).

For (4), 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 $p \mapsto (p_k - k, p_{k-1} - (k - 1), \ldots, p_1 - 1)$.

(5) now follows because $p = (n - k + 1, n - k + 2, \ldots, n)$ is the unique maximal element of $\binom{n}{k}$, and an easy calculation shows that $\lvert p \rvert = k(n - k)$.

This theorem amounts to a description of $Gr(k, k^n)$ 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 copies of vector spaces — such that the closure of every cell is itself a union of cells.) Furthermore, the poset isomorphism with $Y_{n,k}$ says that for every $i$, the number of cells of $Gr(k, k^n)$ of dimension $i$ is precisely the number of Ferrers diagrams with $i$ blocks that fit inside $k^{n-k}$ (recall that this means a $k \times (n-k)$ rectangle). Combinatorially, the best way to express this equality is this:

$$\sum_i (\text{number of Schubert cells of dimension } i) q^i = \sum_i \# \{ \lambda \subseteq k^{n-k} \} q^i = \binom{n}{k}_q.$$

(For those of you who know some algebraic topology: Suppose that $k = \mathbb{C}$. Then $Gr(k, k^n)$ is a cell complex with no odd-dimensional cells (because, topologically, the dimension of cells is measured over $\mathbb{R}$). Therefore,
in cellular homology, all the boundary maps are zero — because for each one, either the domain or the range is trivial — and so the homology groups are exactly the chain groups. So the Poincaré series of $\text{Gr}(k, k^n)$ is exactly the generating function for the dimensions of the cells. If $k = \mathbb{R}$, then things are not nearly this easy — the boundary maps aren’t necessarily all zero, and the homology can be more complicated.)

Example 11.5.4. If $k = 1$, then $\text{Gr}(1,k^n)$ is the space of lines through the origin in $k^n$; that is, projective space $kP^{n-1}$. As a cell complex, this has one cell of every dimension; for instance, the projective plane consists of a 2-cell, the 1-cell and an 0-cell, 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 $Y_{n,k}$ is a chain of rank $n-1$. Its rank-generating function is $1 + q + q^2 + \cdots + q^{n-1}$, which is the Poincaré polynomial of $\mathbb{R}P^{n-1}$.

(For $k = \mathbb{C}$, double the dimensions of all the cells, and substitute $q^2$ for $q$ in the Poincaré polynomial.)

Example 11.5.5. If $n = 4$ and $k = 2$, then the interval in Young’s lattice looks like this:

![Diagram of Young's lattice]

These six partitions correspond to the six matrix-types in (11.5.2). 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 11.5.6. What does all this have to do with enumerative geometry questions such as Problem 11.5.1? The answer (modulo technical details) is that the cohomology ring $H^*(X)$ encodes intersections of subvarieties of $X$: for every subvariety $Z \subseteq \text{Gr}(k,k^n)$ 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 (whatever that means), but the consequence is that Problem 11.5.1 reduces to a computation in $H^*(\text{Gr}(k,k^n))$: find the cohomology class $[Z]$ of the subvariety

$$Z = \{ W \in \text{Gr}(2, \mathbb{C}^4) \mid W \text{ meets some plane in } \mathbb{C}^4 \text{ nontrivially} \}$$

and compare $[Z]^4$ to the cohomology class $[\bullet]$ of a point. In fact, $[Z]^4 = 2[\bullet]$; this says that the answer to Problem 11.5.1 is (drum roll, please) two, which is hardly obvious! To carry out this calculation, one needs to calculate an explicit presentation of the ring $H^*(\text{Gr}(k,k^n))$ 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.

\footnote{If you are more comfortable with differential geometry than algebraic geometry, feel free to think “submanifold” instead of “subvariety.”}
11.6 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 $L_n(k)$. 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 $w \in S_n$ 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 $F_k = k\langle F_1, \ldots, F_k \rangle$ 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

$$v'_1 = b_{11}v_1, \quad v'_2 = b_{12}v_1 + b_{22}v_2, \quad \ldots, \quad v'_n = b_{1n}v_1 + b_{2n}v_2 + \cdots + b_{nn}v_n,$$

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

$$F\ell(n) = \bigcoprod_{w \in S_n} X_w.$$ 

Here the dimension of a Schubert cell $X_w$ is the number of inversions of $w$, i.e.,

$$\dim X_w = \text{inv}(w) = \#\{(i,j) : 1 \leq i < j \leq n \text{ and } w(i) > w(j)\}.$$ 

Recall that this is the rank function of the Bruhat and weak Bruhat orders on $S_n$. In fact, the (strong) Bruhat order is the cell-closure partial order (analogous to (11.5.3)). 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

$$X_w = \bigcup_{v \in S_n, v \leq w} X_v,$$

where $\leq$ means (strong) Bruhat order (see Ex. 1.4.12). These are much-studied objects in combinatorics; for example, determining which Schubert varieties is singular turns out to be a combinatorial question involving the theory of pattern avoidance. Even more generally, instead of $S_n$, 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. The reader who wants the full story should see a source such as [BT82].

204
Definition 11.6.1. 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 11.6.2. 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 11.6.3. Another important example is the tautological bundle on projective space $\mathbb{P}^{d-1}_k = \text{Gr}(1, k^d)$. Recall that this is the space of lines $\ell$ through the origin in $k^d$. The tautological bundle $\mathcal{T}$ 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.

Let $k$ be either $\mathbb{R}$ or $\mathbb{C}$, and let us work in the category of closed compact manifolds over $k$. A vector bundle of rank $d$ is a bundle whose fiber is $k^d$. (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 $\mathbb{P}^{d-1}_k$ 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 $\ell$ through the origin in $\mathcal{E}_m \cong k^d$. In turn, $\mathbb{P}(\mathcal{E})$ has a tautological line bundle $\mathcal{L} = \mathcal{L}(\mathcal{E})$ whose fiber over $(\ell, m)$ is $\ell$.

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 $i > \text{rank} \mathcal{E}$.
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.

$$H^*(\mathbb{P}(\mathcal{E}); \mathbb{Z}) = H^*(M; \mathbb{Z})[x]/(x^d + c_1(\mathcal{E})x^{d-1} + \cdots + c_{d-1}(\mathcal{E})x + c_d(\mathcal{E}))$$ (11.6.1)

The standard symbol for the tautological bundle is actually $\mathcal{O}(-1)$; let’s not get into why.
where $x = c_1(\mathcal{L})$.

**Example 11.6.4 (Projective space).** $\mathbb{P}^{d-1}$ is the projectivization of the trivial rank-$d$ bundle over $M = \{\bullet\}$. Of course $H^*(M; \mathbb{Z}) = \mathbb{Z}$, so $H^*(\mathbb{P}^{d-1}; \mathbb{Z}) = \mathbb{Z}[x]/(x^d)$.

**Example 11.6.5 (The flag variety $F_{\ell}(3)$).** Let $M = \mathbb{P}^2 = \text{Gr}(1, \mathbb{C}^3)$. 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 $\ell$. Let $\mathcal{L} = \mathcal{L}(\mathcal{E}^2)$. 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 tell us that

$$(1 + x)(1 + c_1(\mathcal{E}^2) + c_2(\mathcal{E}^2)) = 1$$

and extracting the graded pieces we get

$$x + c_1(\mathcal{E}^2) = 0, \quad xc_1(\mathcal{E}^2) + c_2(\mathcal{E}^2) = 0$$

so $c_1(\mathcal{E}^2) = -x$ and $c_2(\mathcal{E}^2) = -xc_1(\mathcal{E}^2) = x^2$. Now (11.6.1) tells us that

$$H^*(F_{\ell}(3)) = H^*(\mathbb{P}^2)[y]/(y^2 - xy + x^2) = \mathbb{Q}[x, y]/(x^3, y^2 - xy + x^2).$$

In fact this ring is isomorphic to

$$\mathbb{Q}[a, b, c]/(a + b + c, ab + ac + bc, abc).$$

(For the isomorphism, set $a = x$, $b = -y$, $c = -x + y$.)

**Example 11.6.6 (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,

$$H^*(F_{\ell}(n)) \cong \mathbb{Q}[x_1, \ldots, x_n]/(e_1, e_2, \ldots, e_n)$$

where $e_k$ is the $k^{th}$ elementary symmetric function, i.e.,

$$e_k = \sum_{1 \leq i_1 < \cdots < i_k \leq n} x_{i_1} \cdots x_{i_k}.$$
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,

\[
Poin(F\ell(n), q) = \sum_k q^k \dim QH^2(F\ell(n)) = (1)(1 + q)(1 + q + q^2) \cdots (1 + q + \cdots + q^{n-1}) = [q]_n!.
\]

This expression has a lovely combinatorial interpretation:

\[
[q]_n! = \sum_{w \in S_n} q^{\text{inv}(w)}
\]

where \( S_n \) is the symmetric group on \( n \) letters and \( \text{inv}(w) \) is the number of inversions:

\[
\text{inv}(w) = \# \{(i, j) : 1 \leq i < j \leq n, w(i) > w(j)\}.
\]

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 \( E_k \), 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 \( w \in S_n \). 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 \( X_{\mu}^{4132} \) 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

\[ F\ell(n) = \bigcup_{w \in S_n} X_w^{\circ} \]

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).
11.7 Combinatorial Hopf algebras

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.
- 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 ranked posets $P,Q$ can be combined into one by taking the ordinal sum $P \oplus Q$ (see Definition 1.2.1) by identifying $\hat{1}_P$ with $\hat{0}_Q$.
- Let $\alpha = (\alpha_1,\ldots,\alpha_\ell) |\!| n$. For $0 \leq k \leq \ell$, we can break $\alpha$ 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:

\[
\begin{align*}
|V(G|_W)| + |V(W|_{V \setminus W}| = |V(G)| & \quad |V(G + H)| = |V(G)| + |V(H)| \\
|E(M|_A)| + |E(M/A)| = |E(M) & \quad |E(M \oplus M')| = |E(M)| + E(M')| \\
_r([\hat{0},x]) + _r([x,\hat{1}]) = _r(P)_r(P \oplus Q) = _r(P) + _r(Q) & \quad |\alpha^{(k)}| + |\alpha^{(k)}| = |\alpha|_r(\alpha \beta) = _r(\alpha) + _r(\beta)
\end{align*}
\]

We seek an algebraic structure to record all this.

First, here is an algebraic description.

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 $\mathbb{C}$-linear 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

\[
\begin{array}{c}
A \otimes A \otimes A \xrightarrow{\mu \otimes I} A \otimes A \\
\downarrow I \otimes \mu \quad \mu \downarrow \quad a \otimes b \otimes c \xrightarrow{\text{id}} ab \otimes c \\
A \otimes A \xrightarrow{\mu} A \\
\downarrow \quad a \otimes bc \xrightarrow{\mu} abc
\end{array}
\]

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 $\mu$ to the first two tensor factors and tensor what you get with $[I \text{ applied to}]$ 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 \textit{coassociative}, a condition defined by reversing the arrows in the previous diagram:

```
\[
\begin{array}{c}
Z & \xrightarrow{\Delta} & Z \otimes Z \\
\downarrow & & \downarrow \iota \otimes \Delta \\
Z \otimes Z & \xleftarrow{\Delta \otimes I} & Z \otimes Z \otimes Z
\end{array}
\]
```

Just as an algebra has a unit, a coalgebra has a \textit{counit}. To say what this is, let us diagrammatize the defining property of the multiplicative unit $1_A$ in an algebra $A$: it is the image of $1_C$ under a map $u : C \to A$ such that the diagram

```
\[
\begin{array}{c}
C \otimes A & \xrightarrow{\mu} & A \otimes C \\
\downarrow u \otimes I & & \downarrow I \otimes u \\
A \otimes A & \xleftarrow{I \otimes \varepsilon} & A \otimes A
\end{array}
\]
```

commutes. (Here $I$ is the identity map, and the top diagonal maps take $a \in A$ to $1 \otimes A$ and $a \otimes 1$ respectively.) Thus a counit of a coalgebra is a map $\varepsilon : Z \to C$ such that the diagram

```
\[
\begin{array}{c}
C \otimes A & \xrightarrow{\Delta} & A \otimes C \\
\downarrow \varepsilon \otimes I & & \downarrow I \otimes \varepsilon \\
A \otimes A & \xleftarrow{I \otimes \varepsilon} & A \otimes A
\end{array}
\]
```

A \textit{bialgebra} is a vector space 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

```
\[
\begin{array}{c}
A \otimes A & \xrightarrow{\Delta \otimes \Delta} & A \otimes A \otimes A \otimes A \\
\downarrow \mu & & \downarrow \mu \otimes \mu \\
A & \xrightarrow{\Delta} & A \otimes A
\end{array}
\]
```

Comultiplication takes some getting used to. 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.

\textbf{Example 11.7.1} (The graph algebra). For $n \geq 0$, let $G_n$ 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 $G$), and let $G = \bigoplus_{n \geq 0} G_n$. Thus $G$ is a graded vector space, which we make into a $\mathbb{C}$-algebra by defining $\mu([G] \otimes [H]) = [G \sqcup H]$, where $\sqcup$ denotes disjoint union. The unit is the unique graph $K_0$ with no vertices (or, technically, the map $u : \mathbb{C} \to G_0$ sending $c \in \mathbb{C}$ to $c[K_0]$). Comultiplication in $G$ can be defined by

$$\Delta[G] = \sum_{A, B} [G|_A] \otimes [G|_B]$$
which can be checked to be a coassociative algebra morphism, making $G$ into a bialgebra. This comultiplication is in fact cocommutative. Let $f$ 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 by the diagrams

$$
\begin{array}{c}
\begin{array}{c}
B \otimes B \\
\mu
\end{array}
\end{array}
\xrightarrow{f}
\begin{array}{c}
\begin{array}{c}
B \otimes B \\
\mu
\end{array}
\end{array}
$$

and

$$
\begin{array}{c}
\begin{array}{c}
B \otimes B \\
\Delta
\end{array}
\end{array}
\xrightarrow{f}
\begin{array}{c}
\begin{array}{c}
B \otimes B \\
\Delta
\end{array}
\end{array}
$$

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 11.7.2** (Rota’s Hopf algebra of posets). For $n \geq 0$, let $P_n$ be the vector space of formal $\mathbb{C}$-linear combinations of isomorphism classes $[P]$ of finite graded posets $P$ of rank $n$. Thus $P_0$ and $P_1$ are one-dimensional (generated by the chains of lengths 0 and 1), but dim $P_n = \infty$ for $n \geq 2$. We make $P = \bigoplus_n P_n$ 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

$$
\Delta[P] = \sum_{x \in P} [\hat{0}, x] \otimes [x, \hat{1}].
$$

Coassociativity is checked by the following calculation:

$$
\Delta \otimes I(\Delta(P)) = \Delta \otimes I \left( \sum_{x \in P} [\hat{0}, x] \otimes [x, \hat{1}] \right)
$$

$$
= \sum_{x \in P} \Delta([\hat{0}, x]) \otimes [x, \hat{1}]
$$

$$
= \sum_{x \in P} \left( \sum_{y \in [\hat{0}, x]} [\hat{0}, y] \otimes [y, x] \right) \otimes [x, \hat{1}]
$$

$$
= \sum_{x \leq y \in P} [\hat{0}, y] \otimes [y, x] \otimes [x, \hat{1}]
$$

$$
= \sum_{y \in P} [\hat{0}, y] \otimes \left( \sum_{x \in [y, \hat{1}]} [y, x] \otimes [x, \hat{1}] \right)
$$

$$
= \sum_{y \in P} [\hat{0}, y] \otimes \Delta([y, \hat{1}]) = I \otimes \Delta(\Delta(P)).
$$

This Hopf algebra is commutative, but not cocommutative; there’s no reason for the switching map to fix $\Delta(P)$ unless $P$ is self-dual.

The ring $\Lambda$ of symmetric functions is a coalgebra in the following way. We regard $\Lambda$ as a subring of the ring of formal power series $\mathbb{C}[[x]] = \mathbb{C}[[x_1, x_2, \ldots]]$ First, the counit is just the map that takes a formal power series to its constant term. To figure out the coproduct, we then make a “Hilbert Hotel substitution”:

$$
x_1, x_2, x_3, x_4, \ldots \mapsto x_1, y_1, x_2, y_2
$$

to obtain a power series in $\mathbb{C}[[x, y]] = \mathbb{C}[[x]] \otimes \mathbb{C}[[y]]$. This is symmetric in each of the variable sets $x$ and $y$, i.e.,

$$
\Lambda(x, y) \subseteq \Lambda(x) \otimes \Lambda(y).
$$

---

5 There are those who call this “immutative”.
So every symmetric function $F(x, y)$ can be written in the form $\sum F_1(x)F_2(y)$; we set $\Delta(F) = \sum F_1 \otimes F_2$.

For example, clearly $\Delta(c) = c = c \otimes 1 = 1 \otimes c$ for any scalar $c$.

$$h_1(x, y) = \sum_{i \geq 1} x_i + \sum_{i \geq 1} y_i = \left( \sum_{i \geq 1} x_i \right) \cdot 1 + 1 \cdot \left( \sum_{i \geq 1} y_i \right) = h_1(x) \cdot 1 + 1 \cdot h_1(y)$$

$\therefore \Delta(h_1) = h_1 \otimes 1 + 1 \otimes h_1$

$$h_2(x, y) = h_2(x) + h_1(x)h_1(y) + h_2(y) \quad \Delta(h_2) = h_2 \otimes 1 + h_1 \otimes h_1 + 1 \otimes h_2$$

and more generally

$$\Delta(h_k) = \sum_{j=0}^{k} h_j \otimes h_{k-j}, \quad \Delta(e_k) = \sum_{j=0}^{k} e_j \otimes e_{k-j}.$$

We can finally define a Hopf algebra!

**Definition 11.7.3.** A Hopf algebra is a bialgebra $\mathcal{H}$ with a antipode $S : \mathcal{H} \to \mathcal{H}$, which satisfies the commutative diagram

$$\begin{array}{ccc}
\mathcal{H} & \xrightarrow{\Delta} & \mathcal{H} \otimes \mathcal{H} \\
\epsilon & \downarrow & \mathcal{H} \otimes \mathcal{H} \xrightarrow{\mu} \mathcal{H} \\
\varepsilon & \downarrow & \mathcal{H} \otimes \mathcal{H} \xrightarrow{\mu} \mathcal{H} \\
\mathbb{C} & \xrightarrow{S \otimes I} & \mathcal{H} \\
\end{array} \quad \text{(11.7.1)}$$

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 $\sum S(g_1) \cdot g_2$. 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:

$$\begin{align*}
\mu(S \otimes I(\Delta 1)) &= \mu(S \otimes I(1 \otimes 1)) = \mu(S(1) \otimes 1) = S(1) \\
u(\varepsilon(1)) &= 1 \\
S(1) &= 1 \\
\mu((S \otimes I)(\Delta h_1)) &= \mu((S \otimes I)(h_1 \otimes 1 + 1 \otimes h_1)) = \mu(S h_1 \otimes 1 + S(1) \otimes h_1) = Sh_1 + h_1 \\
u(\varepsilon(h_1)) &= 0 \\
Sh_1 &= -h_1 \\
\end{align*}$$

**Proposition 11.7.4.** Let $B$ be a bialgebra that is graded and connected, i.e., the 0th graded piece has dimension 1 as a vector space. Then the commutative diagram (11.7.1) defines a unique antipode $S : B \to B$, 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".

In general the antipode is not very nice, but for symmetric functions it is. Our calculation of $\Delta(h_k)$ says that

$$\sum_{j=0}^k S(h_j)h_{k-j} = \begin{cases} 1 & \text{if } k = 0, \\ 0 & \text{if } k > 0 \end{cases}$$

and comparing with the Jacobi-Trudi relations says that $S(h_k) = (-1)^k e_k$, i.e., $S = (-1)^k \omega$.

**Characters.** A character on a Hopf algebra $H$ is a $C$-linear map $\zeta : H \to C$ that is multiplicative, i.e., $\zeta(1_H) = 1_C$ and $\zeta(h \cdot h') = \zeta(h)\zeta(h')$. For example, if $H$ is the graph Hopf algebra, then we can define a character by

$$\zeta(G) = \begin{cases} 1 & \text{if } G \text{ has no edges}, \\ 0 & \text{if } G \text{ has one or more edges}, \end{cases}$$

for a graph $G$, and then extending by linearity to all of $G$. This map is multiplicative (because $G \cdot H$ has an edge iff either $G$ or $H$ does); it also looks completely silly to define such a thing.

The reason this is interesting is that characters can be multiplied by the convolution product defined as follows: if $h \in H$ 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 $\varepsilon$ is a two-sided identity for convolution, i.e., $\zeta * \varepsilon = \varepsilon * \zeta = \zeta$ for all characters $\zeta$. Moreover, the definition (11.7.1) of the antipode implies that

$$\zeta * (\zeta \circ S) = \varepsilon$$

(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 $\zeta$, and let $k \in \mathbb{N}$. The $k^{th}$ convolution power of $\zeta$ is given by

$$\zeta^k(G) = \zeta * \cdots * \zeta(G) = \sum_{\substack{V(G) = V_1 \cup \cdots \cup V_k \\text{for which } V_1, \ldots, V_k \text{ are all cocliques}}} \zeta(G|V_1) \cdots \zeta(G|V_k)$$

$$= \sum_{\substack{V(G) = V_1 \cup \cdots \cup V_k \\text{for which } V_1, \ldots, V_k \text{ are all cocliques,}}} \begin{cases} 1 & \text{if } V_1, \ldots, V_k \text{ are all cocliques,} \\ 0 & \text{otherwise.} \end{cases}$$

(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 5.4). 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].

212
11.8 Exercises

Oriented matroids

Max-flow/min-cut and min-max theorems on posets

Exercise 11.1. Prove Proposition 11.2.4.

Exercise 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

Exercise 11.3. Let \( n \geq 2 \) and for \( \sigma \in S_n \), let \( f(\sigma) \) denote the number of fixed points. Prove that for every \( k \geq 1 \), the number \( \frac{1}{n!} \sum_{\sigma \in S_n} f(\sigma)^k \) is an integer.

Grassmannians and flag varieties

Combinatorial Hopf algebras

Exercise 11.4. 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 11.7.4.

Exercise 11.5. Prove that the Billera-Jia-Reiner invariant defines a Hopf algebra morphism \( \mathcal{M} \to QSym \).

(First I’d need to tell you how to comultiply in QSym....)
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, Exercise 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’ll 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.](image)

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 11.8.1.** 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); the number of such things is evidently \(\binom{2n + 1}{n}\).

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 \( z \).)

\[
\begin{align*}
\frac{1}{2n+1} \binom{2n+1}{n} &= \frac{(2n+1)!}{(2n+1) \ (n+1)! \ n!} = \frac{(2n)!}{(n+1) \ n! \ n!} = \frac{(2n)!}{n!} = \frac{1}{n+1} \binom{2n}{n}.
\end{align*}
\]

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:

\[
C_n = C_{n-1} + \sum_{k=1}^{n-1} C_{k-1} C_{n-k} \quad \text{for } n \geq 1.
\] (11.8.1)

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.

\[
\begin{align*}
\text{primitive; } P' \text{ highlighted} & \quad \text{not primitive; } k = 2
\end{align*}
\]

Figure 11.7: Primitive and non-primitive Dyck paths.
Notational Index

Basics

\[ [n] \quad \{1, \ldots, n\}\]
\[ \mathbb{N} \quad \text{nonnegative integers } 0, 1, 2, \ldots \]
\[ \mathbb{P} \quad \text{positive integers } 1, 2, \ldots \]
\[ 2^S \quad \text{power set of a set } S \text{ (or the associated poset)} \]
\[ \cup \quad \text{disjoint union} \]
\[ \mathfrak{S}_n \quad \text{symmetric group on } n \text{ letters} \]
\[ \binom{S}{k} \quad \text{set of } k\text{-element subsets of a set } S \]
\[ C_n \quad \text{Catalan numbers} \]
\[ S(n, k) \quad \text{Stirling number of the second kind} \]
\[ k\langle v_1, \ldots, v_n \rangle \quad k\text{-vector space with basis } \{v_1, \ldots, v_n\} \]

Posets

\[ \geq, \leq \quad \text{“covers”, “is covered by”} \]
\[ \mathbf{0}, \mathbf{1} \quad \text{unique minimum and maximum elements of a poset} \]
\[ [x, y] \quad \text{interval in a poset} \]
\[ P^* \quad \text{dual poset to } P \]
\[ \Pi_n \quad \text{partition lattice (the lattice of all set partitions of } [n] \text{)} \]
\[ K(G) \quad \text{connectivity lattice of a graph } G \]
\[ Y \quad \text{Young’s lattice of integer partitions} \]
\[ \lambda \vdash n \quad \lambda \text{ is a partition of } n \]
\[ \tilde{\lambda} \quad \text{conjugate of a partition } \lambda \]
\[ \text{NC}(n) \quad \text{Set of noncrossing partitions of } [n] \]

Lattices

\[ \land, \lor \quad \text{meet, join} \]
\[ \mathbb{B}_n \quad \text{Boolean lattice of rank } n \]
\[ D_n \quad \text{Lattice of divisors of an integer } n \]
\[ \mathbb{F}_q \quad \text{finite field of order } q \]
\[ L_n(q) \quad \text{lattice of vector subspaces of } \mathbb{F}_q^n \]
\[ J(P) \quad \text{lattice of order ideals of a poset } P \]
\[ \text{Irr}(L) \quad \text{poset of join-irreducible elements in a lattice } L \]
\[ N_5 \quad \text{nonmodular, nonranked 5-element lattice} \]
\[ M_5 \quad \text{modular, ranked, nondistributive 5-element lattice} \]
\[ L(E) \quad \text{geometric lattice represented by a set of vectors } E \]
\[ L^\text{aff}(E) \quad \text{geometric lattice represented by a set of affine points } E \]
Poset Algebra

\text{Int}(P) \quad \text{set of intervals of poset } P
\text{I}(P) \quad \text{incidence algebra of } P
f \ast g \quad \text{convolution product in } \text{I}(P)
\delta \quad \text{Kronecker delta function (identity of } \text{I}(P))
\zeta \quad \text{zeta function in } \text{I}(P)
\mu \quad \text{Möbius function in } \text{I}(P)
\chi_P(x) \quad \text{characteristic polynomial of poset } P
A(L) \quad \text{Möbius algebra of a lattice } L

Matroids and the Tutte Polynomial

\bar{A} \quad \text{closure operator applied to } A
M(G) \quad \text{graphic matroid of a graph } G
\mathcal{I} \quad \text{matroid independence system}
\mathcal{B} \quad \text{matroid basis system}
\mathcal{C} \quad \text{matroid circuit system}
M^* \quad \text{dual of a matroid } M
M \oplus M' \quad \text{direct sum of matroids}
M - e \quad \text{matroid deletion}
M/e \quad \text{matroid contraction}
U_k(n) \quad \text{uniform matroid of rank } k \text{ on set of size } n
T_M, T_M(x, y) \quad \text{Tutte polynomial of } M
p_G(k) \quad \text{chromatic polynomial of a graph } G
C(e, B) \quad \text{fundamental circuit of } e \text{ w/r/t basis } B
C^*(e, B) \quad \text{fundamental cocircuit of } e \text{ w/r/t basis } B

Hyperplane Arrangements

\mathcal{B}_n \quad \text{Boolean arrangement}
\mathcal{B}_G \quad \text{braid arrangement}
L(A) \quad \text{intersection poset of arrangement } A
\text{ess}(A) \quad \text{essentialization of } A
r(A) \quad \text{number of regions of a real arrangement } A
b(A) \quad \text{number of bounded regions of } A
A_x, A^x \quad \text{See Eqn. 6.3.1}
\text{Shi} \quad \text{Shi arrangement}
\mathbb{P}^d_k \quad d\text{-dimensional projective space over field } k
\text{proj}(A) \quad \text{projectivization of a central arrangement } A
cA \quad \text{cone over } A
\mathcal{A}_G \quad \text{arrangement associated with a graph } G

Simplicial Complexes and Polytopes

\langle \cdots \rangle \quad \text{simplicial complex generated by a set of faces}
|\Delta| \quad \text{(standard) geometric realization of } \Delta
\Delta(P) \quad \text{order complex of a poset } P
\mathbb{k}[\Delta] \quad \text{Stanley-Reisner (face) ring of } \Delta \text{ over a field } \mathbb{k}
C_k(\Delta, R) \quad \text{simplicial chain groups over a ring } R
\check{H}_k(\Delta, R) \quad \text{reduced simplicial homology groups}
\mathbb{S}^d \quad d\text{-dimensional sphere}
P^* \quad \text{dual of a polytope } P
Representation Theory

Id identity element of a group
$D_n$ dihedral group of symmetries of a regular $n$-gon
$\rho_{\text{triv}}, \rho_{\text{reg}}$ trivial and regular representations
$\rho_{\text{sign}}, \rho_{\text{def}}$ character of representation $\rho$
$C\ell(G)$ space of class functions of $G$
$\langle \rho, \chi \rangle_G$ inner product on $C\ell(G)$ (see Thm. 9.7.1)
$\text{Hom}_C(V, W)$ $C$-linear maps $V \to W$
$\text{Hom}_G(V, W)$ $G$-equivariant $C$-linear maps $V \to W$
$V^G$ space of invariants of a $G$-action on $V$
$[a, b]$ commutator: $aba^{-1}b^{-1}$
$[G, G]$ commutator subgroup of $G$
$\mathfrak{A}_n$ alternating group on $n$ letters
$\text{Par}(n)$ partitions of $n$
$C_\lambda$ conjugacy class of permutations with cycle shape $\lambda$
$\lambda < \mu$ lexicographic (total) order on partitions
$\lambda \lessdot \mu, \lambda \lessdotdot \mu$ dominance (partial) order on partitions
$\text{sh}(T)$ shape of a tabloid $T$
$\langle \rho_\mu, V_\mu \rangle$ tabloid representation of shape $\mu$
$\chi_\mu$ character of tabloid representation
$\text{Sp}_\lambda$ Specht module
$K_{\lambda, \mu}$ Kostka numbers
$\text{Res}_H^G \rho, \text{Res}_H^G \chi$ restricted representation/character
$\text{Ind}_H^G \rho, \text{Ind}_H^G \chi$ induced representation/character

Symmetric Functions

$x^\alpha$ monomial in variables $x$ with exponent vector $\alpha$
$R[[x]]$ formal power series in $x$ with coefficients in $R$
$m_\lambda$ monomial symmetric function
$e_\lambda$ elementary symmetric function
$h_\lambda$ (complete) homogeneous symmetric function
$p_\lambda$ power-sum symmetric function
$\Lambda_\lambda, \Lambda_{R,d}(x)$ $R$-module of degree-$d$ symmetric functions in $x$
$\Lambda, \Lambda_R(x)$ $R$-algebra of symmetric functions in $x$
$\omega$ involutory automorphism $\Lambda \to \Lambda$ swapping $e$’s and $h$’s
(not to be confused with $w$!)
$\text{CST}(\lambda)$ set of column-strict tableaux of shape $\lambda$
$s_\lambda$ Schur function
$\Omega, \Omega^*$ Cauchy kernel and dual Cauchy kernel
$z_\lambda$ size of centralizer of a partition of shape $\lambda$ (see (10.8.1))
$\varepsilon_\lambda$ sign of a partition of shape $\lambda$ (see (10.8.1))
$\text{SYT}(\lambda)$ set of standard tableaux of shape $\lambda$
$f^\lambda$ number of standard tableaux of shape $\lambda$
$T \leftarrow x$ row-insertion ($\S$10.9)
$\text{ch}$ Frobenius characteristic ($\S$10.12)
$\epsilon_{\mu, \nu}^\lambda$ Littlewood-Richardson coefficients
$h(x)$ Length of the hook with corner $x$

218
Combinatorial Algebraic Varieties

\( \text{Gr}(k, V) \) Grassmannian of \( k \)-dimensional subspaces of \( V \)

\( \Omega_\lambda \) Schubert cell in a Grassmannian

\( F\ell(n) \) (complete) flag variety in dimension \( n \)

\( X_w \) Schubert cell in a flag variety

**Hopf Algebras**

\( \mu \) product

\( \Delta \) coproduct

\( u \) unit

\( \varepsilon \) counit

\( S \) antipode
Bibliography


