Honors Abstract Algebra

Course Notes
Math 55a, Harvard University

Contents

1 Introduction

This course will provide a rigorous introduction to abstract algebra, including group theory and linear algebra.

Topics include:

1. Set theory. Formalization of $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$.

2. Linear algebra. Vector spaces and transformations over $\mathbb{R}$ and $\mathbb{C}$. Other ground fields. Eigenvectors. Jordan form.

3. Multilinear algebra. Inner products, quadratic forms, alternating forms, tensor products, determinants.

4. Abstract groups.

5. Groups, symmetry and representations.

2 Set Theory

Halmos reading. Read Halmos, *Naive Set Theory*, sections 1–15 to learn the foundations of mathematics from the point of view of set theory, and its use in formalizing the integers. Most of this should be review, although the systematic use of a small set of axioms to rigorously establish set theory may be new to you. We will also formalize the rational, real and complex numbers.

Then read 22-23 to learn about cardinality and countable sets.

Finally, read 16–21 and 24–25 to learn about other versions of the axiom of choice, ordinals and cardinals.

Axiom of choice. *For any set A, there is a function $c : \mathcal{P}(A) - \{\emptyset\} \to A$ such that $c(B) \in B$ for all $B \subset A$.*

Theorem 2.1 *The Axiom of Choice is equivalent to the assertion that every set can be well-ordered.*

Proof. If $(A, <)$ is well-ordered, we can define $c(B)$ to be the least element of B.

For the converse, let c be a choice function for A. Let us say a well-ordering $(B, <)$ of a subset of A is *guided by c* if for all $x \in B$, we have $x = c(A - \{y \in B : y < x\})$. It is easy to see that if orderings on B and B' are both guided by c, then $B \subset B'$ or vice-versa, and the orderings agree on $B \cap B'$. Taking the union of order relations compatible with c, we obtain a well-ordering of A. ∎

Here is a sample application that conveys the power of this axiom.

Theorem 2.2 *Any vector space V has a basis. For example, $\mathbb{R}$ has a basis as a vector space over $\mathbb{Q}$.*

Proof. Choose a well-ordering $<$ for V (using the Axiom of choice). Let S be the set of $v \in V$ such that v is not in the linear span of $\{w \in V : w < v\}$. It is easy to see that the elements of S are linearly independent. Suppose the span S' of S is not all of V. Then, by well-ordering, there is a *least* element $v \in V - S'$. But then $v = \sum_1^n a_i v_i$ with $v_i < v$, else we would have $v \in S$. And each v_i lies in S', since $v_i < v$. But then $v \in S'$. ∎

Theorem 2.3 *There is a map $f : \mathbb{R} \to \mathbb{R}$ satisfying $f(x+y) = f(x) + f(y)$, $f(x) = 0$ if x is rational, and $f(\sqrt{2}) = 1$.*

Proof. Let $B_0 = \{1, \sqrt{2}\}$. Using a small variation of the proof above, we can extend B_0 to a basis B for $\mathbb{R}$ over $\mathbb{Q}$. Then any $x \in \mathbb{R}$ can be written uniquely as $x = \sum_B x_b \cdot b$ with $x_b \in \mathbb{Q}$ and $x_b = 0$ for all but finitely many $b \in B$. This implies $(x+y)_b = x_b + y_b$. Now define $f(x) = x_{\sqrt{2}}$. ∎

3 Vector spaces

Axler reading. We will discuss groups, rings, fields, and vector spaces over arbitrary fields. Our main text Axler [Ax] discusses only the fields $\mathbb{R}$ and $\mathbb{C}$. For more general definitions, see [Ar, Ch. 2,3].

Note also that Axler discusses the direction sum $S \oplus T$ of two *subspaces* of a given vector space V. In general one also uses the same notation, $S \oplus T$, to construct a *new* vector space from two given ones, whose elements are ordered pairs $(s, t) \in S \times T$ with the obvious coordinate-wise operations.

Finite fields. When p is a prime, the ring $\mathbb{Z}/p$ is actually a field (usually denoted $\mathbb{F}_p$). It is the unique field (up to isomorphism) containing p elements.

To see $\mathbb{Z}/p$ is a field, just note that if $xy = 0 \bmod p$ then $p | xy$, which (by unique factorization) means $p | x$ or $p | y$, and hence $x = 0$ or $y = 0 \bmod p$. So if $x \neq 0$, the map $\mathbb{Z}/p \to \mathbb{Z}/p$ given by $y \mapsto xy$ is $1 - 1$. By the pigeonhole principle, it is onto, so there is a y such that $xy = 1 \bmod p$.

Polynomials. Let K be a field. The polynomials $K[x]$ form a vector space over K. The elements of $K[x]$ are formal sums $\sum_0^n a_i x^i$ where $a_i \in K$. Thus the polynomials of degree d or less form a vector space of dimension $d + 1$.

Axler defines polynomials (p.10) as certain *functions* $f : K \to K$, namely those of the form $f(x) = \sum_0^n a_i x^i$. This is fine for fields like $\mathbb{Q}$ and $\mathbb{R}$, but it is not the right definition in general. For example, if $K = \mathbb{F}_p$ is the field with p elements, there are infinitely many polynomials in $\mathbb{F}_p[x]$, but only finitely many maps $f : \mathbb{F}_p \to \mathbb{F}_p$. An important special case is the polynomial $f(x) = x^p - x$, which vanishes for every $x \in \mathbb{F}_p$ but is not the zero polynomial.

$\mathbb{R}^4$. It is sometimes useful to appreciate the geometry and topology of higher-dimensional vector spaces. Here is an example. In $\mathbb{R}^2$, a circle can 'enclose' a point. The two objects are linked. But in $\mathbb{R}^3$ you can move the point out

of the plane and then transport it to the outside of the circle, without ever crossing it.

As a test of visualization: show that two circles which are linked in $\mathbb{R}^3$ can be separated in $\mathbb{R}^4$.

Linear interindependence. A basic feature of linear dependence is the following. Suppose

$$0 = \sum_1^n a_i x_i,$$

and all $a_i \neq 0$. (One might then say the x_i are *interdependent*). Then the span of $(x_1, \ldots, \widehat{x}_i, \ldots, x_n)$ is the same as the span of $(x_1, \ldots, x_n)$, for all $i i$. In other words, *any one* of the x_i can be eliminated, without changing their span.

From this we get the main fact regarding bases.

Theorem 3.1 *Let A be a linear independent set and B a finite spanning set for a vector space V. Then $|A| \leq |B|$.*

Proof. Write $A_i = (a_1, \ldots, a_i)$ (so $A_0 = \emptyset$). We inductively construct a sequence of spanning sets of form form $A_i \cup B_i$, as follows. Let $B_0 = B$; then $A_0 \cup B_0$ spans. Assuming $A_i \cup B_i$ spans, we can express a_{i+1} as a linear combination of elements in $A_i \cup B_i$. These interdepedent vectors must include at least one from B_i, since A is an independent set. Thus we can remove one element of B_i, to obtain a set B_{i+1} such that $A_{i+1} \cup B_{i+1}$ still spans. Note that $|B_i| = |B| - i$.

The induction can proceed until i reaches the minimum n of $|A|$ and $|B|$. If $n = |B| < |A|$ then $B_n = \emptyset$, while A_n is a proper subset of A that spans V. This contradicts the linear independence of A. Thus $|B| \geq |A|$. ∎

Assuming V is finite-dimensional, we obtain:

Corollary 3.2 *Any lineary independent set be extended to a basis.*

Corollary 3.3 *All bases have the same number of elements.*

When applied to the case where A and B are both bases, the proof gives more: it gives a sequence of bases $A_i \cup B_i$ that interpolates between A and B. This can be expressed as a factorization theorem for general automorphisms of V.

4 Polynomials

A ratio of polynomials p/q, $q \neq 0$, can always be written as a 'proper fraction', $p/q = s + r/q$, where $\deg(r) < \deg(q)$. Equivalently, we have:

Theorem 4.1 *Given $p, q \in K[x]$, $q \neq 0$, there exist unique polynomials $s, r \in K[x]$ such that $p = sq + r$ and $\deg(r) < \deg(q)$.*

Using this fact one can show that polynomials have gcd's. In particular we have:

Theorem 4.2 *If $p, q \in \mathbb{C}[x]$ have no common zeros, then there exist $r, s \in \mathbb{C}[x]$ such that $sp + rq = 1$.*

See [Ar, Ch. 11] for more details on polynomials.

5 Linear Operators

Theorem 5.1 (Conservation of dimension) *For any linear map $T : V \to W$, we have $\dim V = \dim \operatorname{Ker} T + \dim \operatorname{Im} T$.*

Proof. By lifting a basis for $\operatorname{Im} T$ we get a subspace $S \subset V$ mapping bijectively to the image, and with $V = \operatorname{Ker} T \oplus S$. ∎

Corollary 5.2 *There exists a basis for V and W such that T has the form of a projection followed by an inclusion: $T : \mathbb{R}^n \to \mathbb{R}^i \subset \mathbb{R}^j$.*

This result shows that not only is the theory of finite-dimensional vector spaces trivial (they are classified by their dimension), but the theory of maps between *different* vector spaces V and W is also trivial. It is for this reason that we will concentrate on the theory of *operators*, that is the (dynamical) theory of maps $T : V \to V$ from a vector space to itself.

Rank. The dimension of $\operatorname{Im} T$ is also known as the *rank* of T. When T is given by a matrix, the columns of the matrix span the image. Thus the rank of T is the maximal number of linearly independent columns.

Clearly the rank of ATB is the same as the rank of T, if A and B are automorphisms.

The row rank and the column rank of a matrix are equal. This is clear when the matrix of T is a projection of $\mathbb{R}^n$ onto $\mathbb{R}^i$; it then follows from the Corollary above, by invariance of both quantities under composition with automorphisms.

We will later see a more functorial explanation for this, in terms of vector spaces and their duals.

The rank can be found by repeated row and/or column reduction. Row reduction can be made into an algorithm to compute the inverse of T, as well as a basis for the kernel when T is not invertible.

Example. Row operations on $T = \left(\begin{smallmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 4 & 9 \end{smallmatrix}\right)$ lead to $\left(\begin{smallmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 3 & 8 \end{smallmatrix}\right)$ and then $\left(\begin{smallmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 2 \end{smallmatrix}\right)$, showing this matrix has rank 3.

Quotient spaces. If $U \subset V$ is a subspace, the *quotient space* V/U is defined by $v_1 \sim v_2$ if $v_1 - v_2 \in U$. It is a vector space in its own right, whose elements can be regarded as the parallel translates $v + U$ of the subspace U. There is a natural surjective linear map $V \to V/U$. We have

$$\dim(V/U) = \dim V - \dim U.$$

For more background see [Hal].

If $T \in \operatorname{Hom}(V,V)$ and $T(U) \subset U$ then T induces a *quotient map* on V/U by $v + U \mapsto T(v + U) + U = T(v) + U$.

Exact sequences. A pair of linear maps

$$U \xrightarrow{S} V \xrightarrow{T} W$$

is said to be *exact* at V if $\operatorname{Im} S = \operatorname{Ker} T$. A sequence of linear maps $V_1 \to V_2 \to V_3 \to \cdots$ is said to be *exact* if it is exact at every V_i. Any linear map $T : V \to W$ gives rise to a *short exact sequence* of the form

$$0 \to \operatorname{Ker}(T) \to V \xrightarrow{T} \operatorname{Im}(T) \to 0.$$

Block diagonal and block triangular matrices. Suppose $V = A \oplus B$ with a basis (e_i) that runs first through a basis of A, then through a basis of B.

If $T(A) \subset A$, then the matrix of T is *block triangular*. It has the form $T = \left(\begin{smallmatrix} T_{AA} & T_{BA} \\ 0 & T_{BB} \end{smallmatrix}\right)$. The T_{AA} block gives the matrix for $T|A$. The T_{BB} block gives the matrix for $T|(V/A)$. So quotient spaces occur naturally when we

6

consider block triangular matrices. Note that if $S(A) = A$ as well, then the diagonal blocks for TS can be computed from those for S and those for T.

Finally if $T(B) \subset B$ then $T_{BA} = 0$ and the matrix is block diagonal. This means $T = (T|A) \oplus (T|B)$.

Flags. A *flag* is an ascending sequence of subspaces, $0 \subset V_1 \subset \cdots V_n = V$. A flag is *maximal* if $\dim V_i = i$.

Assume V is a vector space *over* $\mathbb{C}$ (or any algebraically closed field). The theorem on upper triangular form can be re-phrased and proved as follows.

Theorem 5.3 *Any linear operator $T : V \to V$ leaves invariant a maximal flag.*

Proof. Since $\mathbb{C}$ is algebraically closed, T has an eigenvector, and hence an invariant 1-dimensional subspace V_1. Now proceed by induction. Consider the quotient map $T : V/V_i \to V/V_i$. This map also has an eigenvector, and hence there is an invariant 1-dimensional subspace $W_i \subset V/V_i$. Let V_{i+1} be the preimage of this subspace. Then $\dim V_{i+1} = \dim V_i + \dim W_i = i + 1$. Continue in this way until $\dim V/V_i = 0$, i.e. until $i = n$. ∎

Generalized kernels. If $T^i v = 0$ for some $i > 0$, we say v is in the *generalized kernel* of T. Since $\mathrm{Ker}(T^i)$ can only increase with i, it must stabilize after at most $n = \dim V$ steps. Thus the generalized kernel of T coincides with $\mathrm{Ker}(T^n)$, $n = \dim V$.

Similarly we say v is a *generalized eigenvector* of T if v is in the generalized kernel of $T - \lambda$, i.e. $(T - \lambda)^i v = 0$ for some i. The set of all such v forms the *generalized eigenspace* V^λ of T, and is given by $V_\lambda = \mathrm{Ker}(T - \lambda)^n$. The number $m(\lambda) = \dim V^\lambda$ is the *multiplicity* of V^λ as an eigenvalue of T.

Proposition 5.4 *We have $T(V_\lambda) = V_\lambda$.*

Example. Let $T(x, y) = (x + y, y)$. The matrix for this map, $T = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, is already upper triangular. Clearly $v_1 = (1, 0)$ is an eigenvector with eigenvalue 1. But there is no second eigenvector to form a basis for V. This is clearly seen geometrically in $\mathbb{R}^2$, where the map is a shear along horizontal lines (fixing the x-axis, and moving the line at height y distance y to the right.)

However if we try $v_2 = (0, 1)$ we find that $T(v_2) - v_2 = v_1$. In other words, taking $\lambda = 1$, $v_1 (T - \lambda) v_2 \neq 0$ *is* an eigenvector of T. And indeed

$(T - \lambda)^2 = 0$ so $V^\lambda = V$. This is the phenomenon that is captured by generalized eigenspaces.

Here is another way to think about this: if we conjugate T by $S(x, y) = (ax, y)$, then it becomes $T(x + (1/a)y, y)$. So the conjugates of T converge to the identity. The eigenvalues don't change under conjugacy, but in the limit the map is diagonal.

The main theorem about a general linear map over $\mathbb{C}$ is the following:

Theorem 5.5 *For any $T \in \mathrm{Hom}_{\mathbb{C}}(V, V)$, we have $V = \oplus V_\lambda$, where the sum is over the eigenvalues of V.*

This means T can be put into block diagonal form, where there is one $m(\lambda) \times m(\lambda)$ block for each eigenvalue λ, which is upper triangular and has λ's on the diagonal.

The characteristic polynomial. This is given by $p(x) = \prod(x - \lambda)^{m(\lambda)}$. Since $T(V_\lambda) = V_\lambda$, we have $(T - \lambda)^{m(\lambda)}|V_\lambda = 0$. This shows:

Corollary 5.6 (Cayley-Hamilton) *A linear transformation satisfies its characteristic polynomial: we have $p(T) = 0$.*

Determinants. Although we have not official introduced determinants, we would be remiss if we did not mention that $p(x) = \det(xI - T)$.

Lemma 5.7 *If T_{ij} is upper-triangular, then the dimension of the generalized kernel of T ($\mathrm{Ker}\, T^n$, $n = \dim V$) is the same as the number of zero on the diagonal.*

Proof. The proof is by induction on $\dim V = n$. Let $V_1 \subset \cdots \subset V_n = V$ be an invariant flag, and let $(\lambda_1, \ldots, \lambda_n)$ be the eigenvalues of T on V_i/V_{i-1} (equivalently, the diagonal entries of T_{ij}). We have a natural sequence of linear maps

$$0 \to \mathrm{Ker}\, T^n|V_{n-1} \to \mathrm{Ker}\, T^n|V \to \mathrm{Ker}\, T^n|V/V_{n-1} \to 0.$$

To complete the proof it suffices to show this sequence is exact. Indeed, the first term coincides with $\mathrm{Ker}\, T^{n-1}|V_{n-1}$, which by induction is the number of zeros among $(\lambda_1, \ldots, \lambda_{n-1})$; and the last term is 0 or 1, depending on whether or not $\lambda_n = 0$.

The only stage where exactness is not obvious is surjectivity to $\operatorname{Ker} T^n V/V_{n-1}$. To prove this we use the fact that $T^{n-1}(V_{n-1}) = T^n(V_{n-1})$, i.e. the image of $T^i|V_{n-1}$ stabilizes after $n-1$ iterations.

Suppose $[v] \in \operatorname{Ker} T^n|V/V_{n-1}$. Since $\dim V/V_{n-1} = 1$, this is the same as saying $T(v) \in V_{n-1}$. To show surjectivity, we must find a $w \in V_{n-1}$ such that $T^n(v+w) = 0$. But $T^n(v) \in T^{n-1}(V_{n-1}) = T^n(V_{n-1})$, so this is possible and we are done. ∎

Corollary 5.8 *The dimension of V_λ coincides with the number of times λ occurs on the diagonal of T_{ij} in upper triangular form.*

Corollary 5.9 *We have $\sum \dim V_\lambda = \dim V$.*

Corollary 5.10 *We have $V = \oplus V_\lambda$.*

Proof. Let S be the span of $\bigcup V_\lambda$. It remains only to show that $S = V$. But since $T(S) \subset S$, if we decompose S into generalized eigenspaces, then we have $S_\lambda = V_\lambda$. Thus $\dim S = \sum \dim S_\lambda = \sum \dim V_\lambda = \dim V$ and so $S = V$. ∎

Note that $T|V_\lambda = \lambda I + N$ where N is nilpotent. Since $N^m = 0$ with $m = m(\lambda)$ we have

$$T^k|V_\lambda = \sum_{i=0}^{\min(k,m)} \binom{k}{k-i} \lambda^{k-i} N^i. \tag{5.1}$$

Note that this is the sum of at most m terms, each of which is a polynomial in k times λ^k.

Spectral radius. The following result is often useful in application with a dynamical flavor. Let $\|T\|$ denote any reasonable measurement of the size of T, for example $\sup |T_{ij}|$, such that $\|\lambda T\| = |\lambda| \cdot \|T\|$.

Theorem 5.11 *We have $\lim \|T^n\|^{1/n} = \sup |\lambda|$, where λ ranges over the eigenvalues of T.*

The eigenvalues of T are also known as its *spectrum*, and $\sup |\lambda| = \rho(T)$ as the *spectral radius* of T. This follows easily from (5.1).

9

Jordan blocks and similarity. A *Jordan block* is an upper triangular matrix such as

$$
T = \begin{pmatrix}
\lambda & 1 & & & \\
& \lambda & 1 & & \\
& & \lambda & 1 & \\
& & & \lambda & 1 \\
& & & & \lambda
\end{pmatrix}
$$

(where the missing entries are zero). It has the form $T = \lambda I + N$, where N is an especially simple nilpotent operator: it satisfies $N(e_i) = e_{i-1}$, and $N(e_1) = 0$.

Two operators on V are *similar* if $S T_1 S^{-1} = T_2$ for some $S \in \mathrm{GL}(V)$. This means T_1 and T_2 have the same matrix, for suitable choice of bases.

Theorem 5.12 *Every $T \in \mathrm{M}_n(\mathbb{C})$ is similar to a direct sum of Jordan blocks. Two operators are similar iff their Jordan blocks are the same.*

Nilpotence. The *exponent* of a nilpotent operator is the least $q \geq 0$ such that $T^q = 0$.

By considering $(T - \lambda)|V_\lambda$, the proof of the existence of Jordan form follows quickly from the nilpotent case, i.e. the case $\lambda = 0$. In this case the Jordan block is a nilpotent matrix, representing the transitions a graph such as the one shown for $q = 5$ in Figure 1.

Theorem 5.13 *Let T be nilpotent with exponent q, and suppose $v \notin \mathrm{Im}(T)$. Then V admits a T-invariant decomposition of the form $V = A \oplus B$, where A is spanned by $(v, Tv, \ldots, T^{q-1}v)$.*

Proof. Let $A_0 = A$ and let $A_1 = T(A_0)$. Note that $T|\,\mathrm{Im}(T)$ has exponent $q - 1$; thus by induction we can have a T-invariant decomposition $\mathrm{Im}(T) = A_1 \oplus B_1$. Let $B_0 = T^{-1}(B_1)$.

We claim that (i) $A_0 + B_0 = V$ and (ii) $A_0 \cap B_1 = (0)$. To see (i) suppose $v \in V$; then $T(v) = a_1 + b_1 \in A_1 \oplus B_1$ and $a_1 = T(a_0)$ for some $a_0 \in A_0$; and since $T(v - a_0) = b_1 \in B_1$, we have $v - a_0 \in B_0$.

To see (ii) just note that $B_1 \subset \mathrm{Im}(T)$, so $A_0 \cap B_1 = A_1 \cap B_1 = (0)$.

Because of (i) and (ii) we can choose B such that $B_1 \subset B \subset B_0$ and $V = A_0 \oplus B$. Then $T(C) \subset T(B_0) \subset B_1 \subset C$ and we are done. ∎

Graphs. Suppose G is a directed graph with vertices $1, 2, \ldots, n$. Let $T_{ij} = 1$ if i is connected to j, and 0 otherwise. Then $(T^k)_{ij}$ is just the *number of directed paths* of length k that run from i to j.

Theorem 5.14 *The matrix T is nilpotent if and only if T has no cycle.*

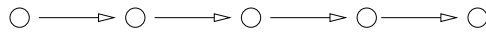
Figure 1. A nilpotent graph.

The minimal polynomial. This is the unique monic polynomial $q(x)$ of minimal degree such that $q(T) = 0$. For example, the characteristic and minimal polynomials of the identity operator are given by $p(x) = (x - 1)^n$ and $q(x) = (x - 1)$. It is straightforward to verify that

$$q(x) = \prod (x - \lambda)^{M(\lambda)},$$

where $M(\lambda)$ is the exponent of $(T - \lambda)|V_\lambda$. This is the same as the maximal dimension of a Jordan block of T with eigenvalue λ.

Classification over $\mathbb{R}$: an example.

Theorem 5.15 *Every $T \in \mathrm{SL}_2(\mathbb{R})$ is either elliptic, parabolic, or hyperbolic.*

6 Inner product spaces

In practice there are often special elements in play that make it possible to diagonalize a linear transformation. Here is one of the most basic.

Theorem 6.1 (Spectral theorem) *Let $T \in \mathrm{M}_n(\mathbb{R})$ be a diagonal matrix. Then $\mathbb{R}^n$ has a basis of orthogonal eigenvectors for T.*

In particular, T is diagonalizable, all its eigenvalues are real, and $V_\lambda = \mathrm{Ker}(T - \lambda I)$ (there are no nilpotent phenomena).

But what does it mean for a matrix to be symmetric? In fact, a matrix which is symmetric in one basis need not be symmetric in another. For example, if T has distinct eigenvalues (the typical case) then there is a basis where it is symmetric (even diagonal), even if it did not start that way.

Inner products. Let V be a vector space over $\mathbb{R}$. With just this structure, there is no notion of lengths or angles. This is added by specifying an *inner product* $V \times V \to \mathbb{R}$, written $\langle x, y \rangle$ or $x \cdot y$.

An inner product is an example of a *bilinear form* $B(x, y)$. This means a map $B : V \times V \to \mathbb{R}$ such that B is linear in each variable individually. An inner product satisfies the additional requirements that it is *symmetric* and *positive definite*. This means $\langle x, y \rangle = \langle y, x \rangle$, a$\langle x, x \rangle \geq 0$ and $\langle x, x \rangle = 0$ iff $x = 0$.

We will *define* the length of a vector in V by $|x|^2 = \langle x, x \rangle$. (Geometrically, it turns out that $\langle x, y \rangle = |x||y| \cos \theta$, where θ is the angle between x and y.)

The main example is of course $\mathbb{R}^n$ with $\langle x, y \rangle = \sum x_i y_i$.

Basic facts:

1. If $\langle x, y \rangle = 0$ then $|x + y|^2 = |x|^2 + |y|^2$ (Pythagorean rule).

2. We have $|\langle x, y \rangle| \leq |x||y|$ (Cauchy-Schwarz inequality).

3. We have $|x + y| \leq |x| + |y|$ (Triangle inequality).

The Pythagorean rule follows from from

$$\langle x + y, x + y \rangle = |x|^2 + |y|^2 + 2\langle x, y \rangle,$$

and the same computation shows the triangle inequality follows from Cauchy-Schwarz.

Projections. To prove Cauchy-Schwarz it is useful, more generally, to introduce the notion of a orthogonal projection onto a subspace.

Given a subspace S, we let $S^\perp = \{v : \langle v, s \rangle = 0 \forall s \in S\}$. This is clearly again a subspace, and clearly $S \cap S^\perp = (0)$. A linear map $P : V \to S$ gives *orthogonal projection onto* S if $\langle Pv, s \rangle = \langle v, s \rangle$ for all $s \in S$. Clearly $\operatorname{Ker} P = S^\perp$, so projection to S exists iff $V = S \oplus S^\perp$, in which case it is unique. It satisfies $P^2 = P$ and $P|S = \mathrm{id}$.

Let us say $(e_i)_1^m$ is an *orthonormal basis* for S if $\langle e_i, e_j \rangle = \delta_{ij}$ and the span of (e_i) is S. If S *has* an orthonormal basis, then we get an explicit formula for P:

$$P(v) = \sum \langle v, e_i \rangle e_i.$$

Theorem 6.2 *Any finite-dimensional inner product space has an orthonormal basis.*

Proof. By induction on $\dim V$. Clear if $\dim V \leq 1$, so assume $\dim V > 1$. Pick any $e_1 \in V$ with $|e_1| = 1$. Then (e_1) is an orthonormal basis for $S = \mathbb{R}e_1$, so the projection to S is defined and $V = S \oplus S^\perp$. By induction, $S^\perp$ has an orthonormal basis $(e_2, \ldots, e_n)$, which together with e_1 gives one for V. ∎

Corollary 6.3 *For any $S \subset V$ we have $V = S \oplus S^\perp$.*

Corollary 6.4 *Any finite-dimensional inner product space is isomorphic to $\mathbb{R}^n$ with the usual inner product.*

Proof of Cauchy-Schwarz: We can assume $|x| = |y| = 1$. Let P be projection to $S = \mathbb{R}x$. Then $Py = \langle x, y \rangle x$ so $|Py| = |\langle x, y \rangle|$. Moreover $Py - y \in S^\perp$, so by the Pythagorean rule we have

$$1 = |y|^2 \geq |Py|^2 + |Py - y|^2 \geq |\langle x, y \rangle|^2$$

and hence $|x||y| = 1 \geq |\langle x, y \rangle|^2$. ∎

Note that equality holds iff $y \in \mathbb{R}x$ or $x = 0$.

Function spaces. Another typical inner-product space is the space of polynomials $\mathbb{R}[x]$ with

$$\langle f, g \rangle = \int_a^b f(x)g(x) \, dx.$$

Note that the monomials $(1, x, x^2, \ldots)$ form a basis for this space but not an *orthonormal* basis. The *Gram-Schmidt* process gives an inductive procedure for generating an orthonormal basis from a basis.

Hermitian inner products. We mention that similar definitions and results hold for vector spaces over $\mathbb{C}$. In this setting the standard example is $V = \mathbb{C}^n$ with $\langle x, y \rangle = \sum x_i \overline{y}_i$. Note that $\langle y, x \rangle = \overline{\langle x, y \rangle}$ and we only have (complex) linearity in the first coordinate. A positive-definite bilinear form with these properties is called a *Hermitian* inner product.

Example: the space $V = \mathbb{C}[z, z^{-1}]$ of function on the circle $S^1 \subset \mathbb{C}$ that are polynomials in z and $1/z$ forms a Hermitian vector space with respect to the inner product

$$\langle f, g \rangle = \frac{1}{2\pi} \int_{S^1} f(z)\overline{g(z)} \, |dz|.$$

More on projections. Let $P : V \to S$ be orthogonal projection to S. We remark that Pv is the unique point in S closest to v. To see this, recall that $t = v - Pv \in S^{\perp}$. So if $s' \in S$ we have

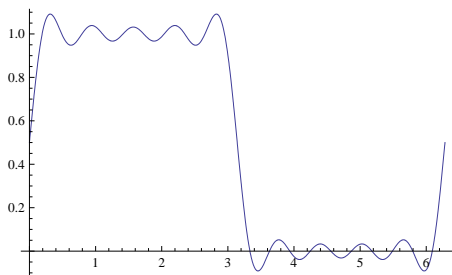$$|v - s'|^2 = |s - s' + t|^2 + |s - s'|^2 + |t|^2 \geq |t|^2 = |Pv - v|^2.$$

Figure 2. Square wave approximated by terms up to $\sin(9x)$.

Signal processing. For example, let $V = C[0, 2\pi]$ be the vector space of continuous functions $f : [0, 2\pi] \to \mathbb{R}$ with the inner product $\langle f, g \rangle = \pi^{-1} \int_0^{2\pi} f(x)g(x)\,dx$. Let $e_0(x) = 2^{-1/2}$, $c_n = \cos(nx)$ and $s_n(x) = \sin(nx)$, and let V_n the span of e_0 and c_i, s_i, $i = 1, \ldots, n$. One can show that this is an orthogonal basis, so it can be used to compute the projection g_n of f to V_n. We find

$$g_n(x) = \frac{1}{2} + \sideset{}{'}\sum_{1}^{k=n} \frac{2}{\pi k} \sin kx,$$

where only *odd* values of k are included in the sum.

The dual space. For a more functorial discussion, we introduce the *dual space* $V^* = \mathrm{Hom}(V, \mathbb{R})$. This is a new vector space built naturally from V; its elements consists of linear maps or *functionals* $\phi : V \to \mathbb{R}$.

Theorem 6.5 *For all $\phi \in V^*$ there exists a unique $y \in V$ such that $\phi(x) = \langle x, y \rangle$ for all $x \in V$.*

Proof. Choose an orthonormal basis (e_i) and let $y = \sum \phi(e_i)e_i$. ∎

14

Corollary 6.6 *We have* $\dim V = \dim V^*$.

Adjoints. Let T be a linear operator on V. Then for any $y \in V$, the map $x \mapsto \langle Tx, y \rangle$ is a linear functional on V. Thus there is a unique vector z such that $\langle Tx, y \rangle = \langle x, z \rangle$ for all $x \in V$. We write $z = T^*(y)$ and use it to define a new linear operator, the *adjoint* of T. It satisfies

$$\langle Tx, y \rangle = \langle x, T^*y \rangle$$

for all $x, y \in V$.

Theorem 6.7 *Let* e_i *be an orthonormal basis for* V. *Then the matrix for* T^* *with respect to* (e_i) *is the transpose of the matrix for* T.

Corollary 6.8 *An operator* $T \in M_n(\mathbb{R})$ *is self-adjoint iff* $T_{ij} = T_{ji}$.

Note that $(ST)^* = T^*S^*$, $T^{**} = T$, and if $T(S) \subset S$ then $T^{(}S^\perp) \subset S^\perp$.

The spectral theorem. We can now prove that a self-adjoint operator is diagonal with respect to a suitable orthonormal basis. In other words, if $T = T^*$ then V has an orthonormal basis of eigenvectors for T. Note that we are working over $\mathbb{R}$, not $\mathbb{C}$, and we are getting honest eigenvectors, not generalized eigenspaces.

The proof is based on some preliminary good properties for T. Suppose T is self-adjoint.

Lemma 6.9 *If* $T(S) \subset S$ *then* $T(S^\perp) \subset S^\perp$.

Lemma 6.10 *We have* $\langle v, T^2v \rangle \geq 0$ *for all* v.

Corollary 6.11 *If* $a > 0$ *then* $T^2 + a$ *is invertible. More generally,* $T^2 + aT + b$ *is invertible if* $p(x) = x^2 + ax + b$ *has no real root.*

Corollary 6.12 *The operator* T *has a (real) eigenvector.*

Proof. As usual, by considering $v, Tv, \ldots, T^n v$ as find a nontrivial monic polynomial $p \in \mathbb{R}[x]$ such that $\dim \operatorname{Ker} p(T) > 0$. Now over $\mathbb{R}$, $p(x)$ factors into linear terms $(x - r_i)$ (with real roots) and quadratic terms $q_i(x)$ (with no real roots). By the preceding Corollary, $q_i(T)$ is invertible for each i, so $(T - r_i I)$ must be noninvertible for some i. ∎

Proof of the spectral theorem. By induction on $\dim V$. Let $S = \mathbb{R}e_1$ be the span of a unit eigenvector. Then $T(S^\perp) \subset S^\perp$, and $T|S^\perp$ has a basis of eigenvectors by induction. ∎

(Note: one should verify that $T^*(S) \subset S$.)

The orthogonal group. The symmetries of V which preserve its inner product form the *orthogonal group* $O(V)$. These are exactly the linear transformations T such that

$$\langle Tx, Ty \rangle = \langle x, y \rangle$$

for all x, y. Equivalently, we have

$$\langle x, T^*Ty \rangle = \langle x, y \rangle.$$

But this happens iff $T^*T = I$. Thus:

$$O(V) = \{T \in \operatorname{Hom}(V, V) \ : \ T^*T = I\}.$$

Equivalently, we have $T^{-1} = T^*$. So these transformations are sort of 'anti-self adjoint'.

For example, the rotation of $\mathbb{R}^2$ given by $T = \left(\begin{smallmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{smallmatrix} \right)$ are orthogonal. The terminology is explained by the following observation: *a matrix T_{ij} belongs to $O(\mathbb{R}^n)$ iff its columns form an orthonormal basis for $\mathbb{R}^n$.*

Theorem 6.13 *For any orthogonal transformation T, there exists an orthogonal splitting $V = \oplus V_i$ where each V_i has dimension 1 or 2. When V_i is two-dimensional, $T|V_i$ is a rotation; when V_i is one-dimensional, $T|V_i = \pm I$.*

Proof. Recall that every real operator has an invariant subspace V_1 of dimension one or two. This is the span of v and Tv where v is in the kernel of $T - a$ or $T^2 - bT - c$. Since T is an isometry, $T(V_1) = V_1$ and $T(V_1^\perp) = V_1^\perp$. The proof then proceeds by induction. ∎

Corollary 6.14 *Every isometry of the 2-sphere that preserves its orientation is a rotation about an axis.*

Polar decomposition. A self-adjoint operator T is *positive* if $T = T^*$ and $\langle Tv, v \rangle \geq 0$ for all v. This is equivalent to its eigenvalues satisfying $\lambda_i \geq 0$. (Q. Is the condition that $T = T^*$ redundant?)

16

Proposition 6.15 *A positive operator has a unique positive square-root.*

Cleary A^*A is positive for any operator A.

Theorem 6.16 (Polar decomposition) *Any invertible operator A on V can be unique expressed as $A = TR$, where T is positive and $R \in \mathrm{O}_n(\mathbb{R})$.*

Proof. Let $T = \sqrt{A^*A}$; then $R = AT^{-1}$ satisfies $RR^* = AT^{-2}A^* = A(A^*A)^{-1}A = I$, so $R \in \mathrm{O}_n(\mathbb{R})$. ∎

Normal operators. Here is a useful extension to a vector space V over $\mathbb{C}$ equipped with a Hermitian form. We say an operator is *normal* if $TT^* = T^*T$.

Theorem 6.17 *A transformation $T \in \mathrm{GL}_n(V)$ has an orthonormal basis of eigenvectors iff T is normal.*

Proof. Clearly diagonal matrices are normal. For the converse, let $V_1 = \mathrm{Ker}(T - \lambda_1 I) \subset V$ be a nontrivial eigenspace. We claim T and T^* both preserve the splitting $V = V_1 \oplus V_1^{\perp}$.

First, we always have $T^*(V_1^{\perp}) \subset V_1^{\perp}$ (whether or not T is normal). But we also have $T^*(V_1) \subset V_1$, since for $v \in V_1$ we have

$$TT^*(v_1) = T^*T(v_1) = T^*(\lambda_1 v_1) = \lambda_1 T^*(v_1).$$

This implies $T(V_1^{\perp}) \subset V_1^{\perp}$. Then $T|V_1^{\perp}$ is also normal and the proof proceeds by induction. ∎

7 Bilinear forms

Let $B : V \times V \to \mathbb{R}$ be a bilinear form on a finite-dimensional real vector space.

We say B is *degenerate* if there is an $x \neq 0$ such that $B(x, y) = 0$ for all y; otherwise it is *nondegenerate*. Any bilinear form is the sum of a symmetric and an antisymmetric one. Our aim is to classify these.

It is often useful to give V an inner product, which we can then compare to B.

Proposition 7.1 *If V has an inner product, then there is a unique operator A such that*

$$B(x, y) = \langle x, Ay \rangle.$$

The form B is symmetric iff A is self-adjoint; antisymmetric iff $A^* = -A$; and B is nondegenerate iff A is invertible.

Quadratic forms. Suppose B is symmetric. Associated to $B(x, y)$ is the *quadratic form*

$$Q(x) = B(x, x).$$

The form Q determines B, since

$$2B(x, y) = Q(x + y) - Q(x) - Q(y).$$

(Note that for an antisymmetric form, Q would be zero.)

(Aside: this formula is related to the famous equation

$$(x^2 + y^2) - (x^2 - y^2)^2 = (2xy)^2,$$

which allows one to generate all Pythagorean triples. The geometric meaning of this equation is that the rational solutions to $a^2 + b^2 = 1$, projected from $(a, b) = (0, 1)$, are in bijection with the rational numbers on the a-axis.)

A quadratic form on $\mathbb{R}^n$ is nothing more than a homogeneous quadratic polynomial in the coordinates on $\mathbb{R}^n$:

$$Q(x) = \sum a_{ij} x_i x_j,$$

and the coefficients a_{ij} are unique if we require $a_{ij} = a_{ji}$. If we say $A(x) = \sum a_{ij} x_j$, then

$$B(x, y) = \langle x, Ay \rangle.$$

Examples with mixed signatures. For the usual inner product, $Q(x) = |x|^2$. But there are other important examples! The standard quadratic form of *signature* (p, q) on $\mathbb{R}^n$, $n \geq p + q$, is given by

$$Q(x) = \sum_{1}^{p} x_i^2 - \sum_{p+1}^{q} x_i^2.$$

This form is nondegenerate iff $p + q = n$.

Note that even when Q is nondegenerate, the 'lengths' of vectors can have both signs, or even be zero.

Another important example is provided by $Q(x, y) = xy$ on $\mathbb{R}^2$. This is in fact equivalent to the standard form of signature $(1, 1)$. In fact, if we let $e_1 = (1, 1)/\sqrt{2}$ and $e_2 = (1, -1)/\sqrt{2}$, then $B(e_i, e_j) = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix} \right)$.

Theorem 7.2 *For any real quadratic form on a finite-dimensional vector space V, there exists a basis such that that $Q(x)$ is the standard quadratic form of signature (p, q) with $p + q \leq n$.*

Proof. *Choose* an inner product on V and write the associated bilinear form as $B(x, y) = \langle x, Ay \rangle$ where A is self-adjoint. Then choose an orthonormal basis e_i of eigenvectors for A with eigenvalues λ_i. Replacing e_i with $|\lambda_i|^{-1/2} e_i$, we can assume $B(e_i, e_i) = \pm 1$ or 0. ∎

Conics and quadrics. A *real quadric* $X \subset \mathbb{R}^n$ is the locus defined by an equation of the form

$$f(x) = Q(x) + L(x) + C = 0,$$

where $Q(x)$ is a quadratic form, $L(x)$ is linear and C is a constant. We say two quadrics are equivalent if there is an (invertible) affine transformation $T(x) = Ax + b$ such that $T(X_1) = X_2$. A quadric is *degenerate* if Q is zero, or it is equivalent to one where f does not depend on all the coordinates, or where the linear and constant term both vanish.

To classify nondegenerate quadrics, we first arrange that Q has signature (p, q), $1 \leq p + q \leq n$. Then we can complete the square to eliminate all but $r = n - p - q$ coordinates from $L(x)$. If $r > 0$ then we can translate in these coordinates to absorb the constant. If $r > 1$ then we can get rid of one of the linear coordinates completely, so the quadric is degenerate. And if there is no linear term, we can scale Q so the constant becomes 1. So in the end we find:

Theorem 7.3 *Any nondegenerate quadric is equivalent ton one given by a purely quadratic equation of the form*

$$Q_{pq}(x) = 1,$$

with $p + q = n$, $p \geq 1$; or by a parabolic equation

$$x_n = Q_{pq}(x),$$

where $p \geq q$, $p + q = n - 1$.

Corollary 7.4 *Any nondegenerate conic is equivalent to the circle $x^2 + y^2 = 1$, the hyperbola $x^2 - y^2 = 1$, or the parabola $y = x^2$.*

Theorem 7.5 *Any nondegenerate quadric in $\mathbb{R}^3$ is equivalent to either:*

1. *The sphere $x^2 + y^2 + z^2 = 1$; or*

2. *The 1-sheeted hyperboloid $x^2 + y^2 - z^2 = 1$; or*

3. *The 2-sheeted hyperboloid $x^2 - y^2 - z^2 = 1$; or*

4. *The elliptic paraboloid $z = x^2 + y^2$; or*

5. *The hyperbolic paraboloid $z = x^2 - y^2$.*

Alternating forms. We now handle the alternating case.

Suppose $B(x, y) = -B(y, x)$. The main example of such a form is

$$B(x, y) = \det \begin{pmatrix} x_1 & x_2 \\ y_1 & y_2 \end{pmatrix} = x_1 y_2 - x_2 y_1.$$

This form measures the signed area of the parallelogram with sides x and y. By taking the sum of n such forms, we obtain a nondegenerate alternating form on $\mathbb{R}^{2n}$. It is given by $B(x, y) = \langle x, J_{2n} y \rangle$ where J_{2n} has n blocks of the form $\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix} \right)$ on the diagonal.

A nondegenerate alternating form is also called a *symplectic form*

Theorem 7.6 *Let $B(x, y)$ be a symplectic form on finite-dimensional real vector space V. Then $\dim V = 2n$ is even, and there exists a basis such that $B(e_i, e_j) = (J_{2n})_{ij}$.*

Note. The prototypical example of such a form arises on the homology of a closed orientable surface of genus g.

Proof. Choose any $e_1 \neq 0$. Then there exists an e_2 such that $B(e_1, e_2) = 1$ (else B is degenerate). Let V_1 be the span of e_1 and e_2; this is 2-dimensional, else we would have $B(e_1, e_2) = 0$. The matrix of $B|V_1$ is now J_2.

Let $V_1^\perp = \{v \in V : B(v, e_1) = B(v, e_2) = 0\}$. It is easy to see that $V = V_1 \oplus V_1^\perp$. First, these two subspaces meet only at 0 since $B|V_1$ is non-degenerate. And second, $V_1^\perp$ is the kernel of a map to $\mathbb{R}^2$, so its dimension is (at least) $n - 2$.

It follows that $B|V_1^\perp$ is non-degenerate, and we proceed by induction. ∎

Change of basis. The matrix of $T : V \to V$ relative to a given basis e_i is determined by the condition:

$$T(e_j) = \sum_i T_{ij} e_i.$$

What is the matrix of T in another basis (f_i)? To find this we form the matrix S satisfying $S(e_j) = f_j$, so the columns of S give the new basis.

Proposition 7.7 *The matrix of T relative to the new basis (f_i) is the same as the matrix of $U = S^{-1}TS$ relative to the old basis.*

Proof. We have $U(e_j) = \sum U_{ij} e_i$ and $SU = TS$, so

$$T(f_j) = TS(e_j) = SU(e_j) = \sum U_{ij} S(e_i) = \sum U_{ij} f_j.$$

∎

On the other hand, the matrix of a bilinear for B relative to (e_i) is given by

$$B_{ij} = B(e_i, e_j).$$

It satisfies $B(v, w) = \sum v_i B_{ij} v_j$.

Proposition 7.8 *The matrix of B relative to the new basis (f_i) is given by $S^* B S$.*

Proof. We have

$$B(f_i, f_j) = B(S(e_i), S(e_j)) = \sum_{k,l} S_{ki} B_{kl} S_{lj} = \sum S^*_{ik} B_{kl} S_{lj} = (S^* B S)_{ij}.$$

∎

Note that these operations are very different: if $S = \lambda I$, $\lambda \neq 0$, then $S^{-1}TS = T$ but $S^* B S = |\lambda|^2 B$. They agree if S is orthogonal (or unitary), i.e. if $SS^* = I$.

8 Trace and determinant

The trace and determinant can be approach in many different ways: using bases, using eigenvalues, and using alternating forms. It is useful to understand all approaches and their interrelations.

Trace. Let V be a finite-dimensional space over K, and let $T : V \to V$ be a linear map. The trace of T is defined by choosing a basis and then setting $\mathrm{tr}(T) = \sum T_{ii}$. We then find:

$$\mathrm{tr}(ST) = \sum_{i,j} S_{ij} T_{ji} = \mathrm{tr}(TS).$$

This shows $\mathrm{tr}(S^{-1}TS) = \mathrm{tr}(T)$ so in fact the trace doesn't depend on the choice of basis.

Now suppose $K = \mathbb{R}$ or $\mathbb{C}$. By putting T into upper-triangular form, we see:

$$\mathrm{tr}(T) = \sum m(\lambda)\lambda.$$

This gives another, basis-free definition of the trace. And finally we see that characteristic polynomial satisfies

$$p(t) = \prod(t - \lambda)^{m(\lambda)} = t^n - \mathrm{Tr}(T)t^{n-1} + \cdots$$

This shows that any convenient definition of $p(t)$ also gives a definition of the trace.

Determinant. We begin with the case $V = \mathbb{R}^n$. We can associate to a matrix T its column vectors $(v_1, \ldots, v_n) = (T(e_1), \ldots, T(e_n))$. We then wish to define a quantity which measures the volume of the parallelopiped with edges $v_1, \ldots, v_n$. We will denote this quantity by $\det(v_1, \ldots, v_n)$ and *define* it by induction on the dimension using expansion by minors in the first row. That is, writting $v_i = (a_i, v'_i)$, we set

$$\det(v_1, \ldots, v_n) = a_1 \det(v_2, \ldots, v_n) - a_2 \det(v_1, v_3, \ldots, v_n) + \cdots \pm v_n \det(v_1, v_2, \ldots, v_{n-1}).$$

By induction on n, we find this function has three key properties:

1. It is linear in each v_i.

2. If $v_i = v_{i+1}$ then the determinant is zero.

3. It is normalized so that $\det(e_1, \ldots, e_n) = 1$. (Volume of a cube).

From these properties one easily deduces the behavior of the determinant under *elementary operations*:

1. The determinant is unchanged if a multiple of one vector is added to another,

2. The determinant changes sign when 2 adjacent entries are reversed;

3. The determinant scales by a if one of the v_i is scaled by a.

Theorem 8.1 *There is a unique function* $\det(v_1, \ldots, v_n)$ *with these properties.*

Proof. Observe that if the (v_i) form a basis, we can repeatedly apply the operations above to convert (v_i) to (e_i), and hence compute det. Similarly if there is a linear relation, we can apply these operations to get $v_i = 0$ for some i, and then by linearity $\det(v_i) = 0$. Thus the properties above above uniquely determe $\det(v_i)$. ∎

Elementary matrices. Let E_{ij} denote the matrix with 1 in position (i, j) and 0 elsewhere. The *elementary matrices* are those of the form

1. $E = I + aE_{ij}$, $i \neq j$; which satisfies $E(e_j) = e_j + ae_i$;

2. $I + E_{ij} + E_{ji} - E_{ii} - E_{jj}$; which satisfies $E(e_i) = e_j$ and vice-versa; and

3. $I + (a - 1)E_{ii}$; which satisfies $E(e_i) = ae_i$.

In each case the vectors $(w_i) = TE(e_i)$ are obtained from $(v_i) = T(e_i)$ by performing an elementary operation. The argument just given shows:

Theorem 8.2 *The elementary matrices generate* $\mathrm{GL}_n(\mathbb{R})$.

We have also seen that the elementary operation changes the determinant in a predictable way: $\det(TE) = \det(T), -\det(T)$ or $a\det(T)$. But $\det(E)$ is also easily computed, allowing us to verify:

Theorem 8.3 *We have* $\det(TE) = \det(T)\det(E)$ *for any elementary matrix.*

Corollary 8.4 *We have* $\det(ST) = \det(S)\det(T)$ *for all* $S, T \in M_n(\mathbb{R})$.

(Here we use the fact that $\det(T) = 0$ if T is not invertible.)

Eigenvalues and determinants. Since $\det(S^{-1}TS) = \det(S)$, we also find

$$\det(T) = \prod \lambda^{m(\lambda)}$$

and hence $p(T) = T^n + \cdots + (-1)^n \det(T)$.

Alternating forms. Here is a more functorial approach to the determinant. Recall that V^* denotes the space of linear forms on V, i.e. linear maps $\phi : V \to K$.

Let $V^* \otimes V^*$ denote the space of *bilinear* forms. The notation is meant to suggest that $B(v_1, v_2) = \phi(v_1)\psi(v_2)$ is always such a form, so there is a natural map $V^* \oplus V^* \to V^* \otimes V^*$. A bilinear form is determined by its values $B_{ij} = B(e_i, e_j)$ on pairs of basis elements.

Similarly let $\otimes^k V^*$ denotes the space of k-linear forms. Such a form is determined by a 'matrix' with k indices, and hence $\dim \otimes^k V^* = n^k$.

A k-form is *alternating* if it vanishes whenever two entries are equal. In characteristic zero, this is equivalent to saying that ϕ changes sign whenever any two adjacent entries are interchanged. We let

$$\wedge^k V^* \subset \otimes^k V^*$$

denote the vector space of alternating k-linear forms $\phi(v_1, \ldots, v_k)$ on V.

For any linear map $T : V \to W$ we get a natural map $T^* : W^* \to V^*$. (How is this related to the adjoint in an inner product space?) Similarly, by 'pullback', we get a natural map

$$\wedge^k T^* : \wedge^k W^* \to \wedge^k V^*,$$

defined by

$$\wedge^k T^*(\phi) = \phi(Tv_1, \ldots, Tv_k).$$

Sign of a permutation. A *permutation* $\sigma \in S_n$ is a bijection from $\{1, \ldots, n\}$ to itself. Every permutation is a product of transpositions, indeed, adjacent transpositions would suffice.

The *sign* function sign $: S_n \to (\pm 1)$ is the function that is 1 if σ is a product of an even number of transpositions, and -1 if it is a product of an

odd number of transpositions. The fact that this function is well-defined can be seen as follows. Consider the polynomial

$$P(t_1, \ldots, t_n) = \prod_{i<j}(t_i - t_j).$$

Given σ we can form a new polynomial

$$Q = Q(t_1, \ldots, t_n) = P(t_{\sigma 1}, \ldots, t_{\sigma n}) = \pm P,$$

and we set $\text{sign}(\sigma) = P/Q$. Evidentally $\text{sign}(\sigma\sigma') = \text{sign}(\sigma)\,\text{sign}(\sigma')$ and $\text{sign}(\tau) = -1$ for a transposition, so sign gives the parity. In particular this definition of π shows that no permutation is both even and odd.

Theorem 8.5 *For any $\phi \in \wedge^k V^*$, and any $\sigma \in S_k$, we have*

$$\phi(v_{\sigma 1}, \ldots, v_{\sigma k}) = \text{sign}(\sigma)\phi(v_1, \ldots, v_k).$$

Theorem 8.6 *If $n = \dim V$, the space $\wedge^n V^*$ is 1-dimensional.*

Proof. Any ϕ is determined by its value on $(e_1, \ldots, e_n)$, and there is a unique ϕ satisfying $\phi(e_{\sigma 1}, \ldots, e_{\sigma n}) = \text{sign}(\sigma)$. ∎

To compute the action of T on $\wedge^n V^*$ it then suffices to compute the single number

$$\phi(Te_1, \ldots, Te_n)/\phi(e_1, \ldots, e_n).$$

Using the multilinear and alternating properties of ϕ it follows that this number obeys the same axioms as the determinant, which shows:

Theorem 8.7 *We have $\wedge^n T^*(\phi) = \det(T)\phi$ for all $\phi \in \wedge^n V^*$.*

Corollary 8.8 *We have $\det T = \sum \text{sign}(\sigma)T_{i,\sigma i}$.*

Theorem 8.9 *The characteristic polynomial is given by $p(t) = \det(tI - T)$.*

These result can also be read backwards, as alternative, basis-free/eigenvector-free definitions of $\det(T)$ and of the characteristic polynomial. For example we immediately see that the constant term in $p(t)$ is given by $p(0) = \det(-T) = (-1)^n \det(T)$.

Theorem 8.10 *We have $\det(T) = \det(T^*)$.*

Proof. This is clear in the one-dimensional case; now use the fact that $\wedge^n T^* = (\wedge^n T)^*$. ∎

Theorem 8.11 *The coefficients of the characteristic polynomial*

$$p(t) = \det(tI - t) = t^n + a_1 t^{n-1} + \cdots + a_n$$

are given by

$$a_i = (-1)^i \operatorname{Tr} \wedge^i T^*.$$

Cramer's rule. Let M'_{ij} be the $(n-1) \times (n-1)$ (minor) matrix obtained by deleting the ith row and jth column from T_{ij}, and let

$$T'_{ij} = (-1)^{i+j} \det M_{ij}.$$

Then $\sum T_{i1} T'_{i1} = \det(T)$. Similarly $\sum T_{i1} T'_{ij} = 0$ for $j \neq 1$, since this gives an expansion for the determinant of a matrix with two equal columns. The same reasoning shows

$$\sum_i T_{ij} T'_{ik} = (\det T) \delta_{ik},$$

or equivalently $T^{-1} = (T')^t$.

Corollary 8.12 *A matrix in $\mathrm{M}_n(K)$ is invertible iff its determinant is nonzero, in which case the inverse is also in $\mathrm{M}_n(K)$.*

If $A \subset K$ is a subring (e.g. $\mathbb{Z} \subset \mathbb{R}$), $T \in \mathrm{M}_n(A)$ and $\det(T)^{-1} \in A$, then $T^{-1} \in \mathrm{M}_n(A)$.

Intrinsically, T' gives the matrix of $\wedge^{n-1} T^*$ acting on $\wedge^{n-1} V^*$.

Tensor products. In general, the *tensor product* $V \otimes W$ of two vector spaces is a new vector space with a bilinear map $V \times W \to V \otimes W$, written $(v, w) \mapsto v \otimes w$, such that for any other bilinear map $\phi : V \times W \to Z$ there is a *unique* compatible linear map $\Phi : V \otimes W \to Z$.

If (v_i) and (w_j) are bases for V and W then $(v_i \otimes w_j)$ is a basis for $V \otimes W$, and $\Phi(\sum a_{ij} v_i \otimes w_j) = \sum a_{ij} \phi(v_i, w_j)$. In particular, we have

$$\dim V \otimes W = (\dim V)(\dim W).$$

Theorem 8.13 *There is a natural isomorphism between $\operatorname{Hom}(V, W)$ and $V^* \otimes W$.*

Proof. Define a bilinear map $V^* \times W \to \operatorname{Hom}(V, W)$ by sending (ϕ, w) to the map $T(v) = \phi(v)w$, then take its canonical extension to $V^* \otimes W$. ∎

Functoriality. If $T : V \to W$ then we have a natural map $T^* : V^* \to W^*$. If $S : V \to V'$ and $T : W \to W'$, then we get a map $S \otimes T : V \otimes W \to V' \otimes W'$ satisfying

$$(S \otimes T)(v \otimes w) = S(v) \otimes T(w).$$

Reciprocal polynomials. Let us say the *reciprocal* of a monic, degree n polynomial $p(t) = t^n + a_1 t^{n-1} + \cdots + a_n$ is the monic polynomial $p^*(t)$ whose roots are the reciprocals of the roots of $p(t)$. It is given explicitly by $p^*(t) = t^n p(1/t)/a_n$. We say p is a *reciprocal polynomial* if $p(t) = p^*(t)$.

Example: if $p(t) = \det(tI - T)$ is the characteristic polynomial of T, and T is invertible, then $p^*(t)$ is the characteristic polynomial of T^{-1}.

Theorem 8.14 *If $T \in \mathrm{Hom}(V, V)$ preserves a nondegenerate quadratic form B — that is, if $T \in \mathrm{O}(V, B)$ — then its characteristic polynomial satisfies $p^*(t) = p(t)$.*

Proof. To say V preserves B is to say that the isomorphism $B : V \to V^*$ conjugates T to T^*, reversing the direction of the arrow; i.e. T^* is conjugate to T^{-1}. Since T and T^* have the same characteristic polynomial, the result follows. In terms of equations, we have $T^* B T = B$ and so $BTB^{-1} = (T^*)^{-1}$. ∎

Corollary 8.15 *A rotation of $\mathbb{R}^{2n+1}$ preserves an axis.*

Proof. It must have an eigenvalue which satisfies $\lambda = 1/\lambda$, so $\lambda = \pm 1$. ∎

Diagonalization. One might hope that any $T \in \mathrm{O}(V, B)$ is diagonalizable, at least over $\mathbb{C}$. This is in fact true for symmetric forms, using the fact that if S is an eigenspace then $V = S \oplus S^\perp$. But it is false for antisymmetric forms; for example $T = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix} \right)$ preserves area on $\mathbb{R}^2$, so it lies in $\mathrm{Sp}_2(\mathbb{R})$, but it is not diagonalizable.

9 Introduction to Group Theory

A group is a G set with an associative operation $G \times G \to G$ that has a (two-sided) identity and inverses.

A *subgroup* $H \subset G$ is a subset that is closed under multiplication (i.e. $HH = H$) and, with respect to this operation, is a group in its own right.

The *intersection* $\bigcap H_i$ of any collection of subgroups is a subgroup. (Compare subspaces.) For example, if $S \subset G$ is a finite set, the subgroup $\langle S \rangle$ generated by S is the intersection of all subgroups containing S. (It is also the space of words in S and S^{-1}.)

Maps. A homomorphism of groups is just a map $f : G \to H$ such that $f(xy) = f(x)f(y)$. Its kernel and image are subgroups of G and H respectively. If f is a bijection, then its inverse is a homomorphism, and we say G and H are *isomorphic*.

Examples. The integers $\mathbb{Z}$ and the integers mod n, $\mathbb{Z}/n$. For any field (e.g. $\mathbb{R}$, $\mathbb{Q}$, $\mathbb{F}_p$, $\mathbb{C}$) we have (abelian) subgroups K and K^*. The group $(\mathbb{Z}/n)^*$.

Linear groups: $\mathrm{GL}_n(K)$; $\mathrm{O}_n(\mathbb{R})$; $\mathrm{SO}_n(\mathbb{R})$; $\mathrm{U}_n(\mathbb{C})$; $\mathrm{SU}_n(\mathbb{C})$; $\mathrm{Sp}_{2g}(\mathbb{R})$; $\mathrm{SO}(p,q)$; $\mathrm{GL}_n(\mathbb{Z})$; $\mathrm{O}_n(\mathbb{Z})$.
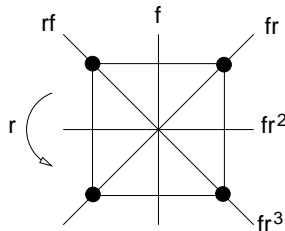

Figure 3. The dihedral group D_4.

The dihedral group. What is $\mathrm{O}_2(\mathbb{Z})$? It is the same as the symmetries of a diamond or square, allowing flips. We can record its 8 elements as a subgroup of S_4. It is generated by $r = (1234)$ and $f = (24)$. We have $rf = f^3 r$. More generally, $D_{2n} = \langle r, f \; : \; r^n = f^2 = e, rf = fr^{-1} \rangle$.

The symmetries of a cube. What is $G = \mathrm{SO}_3(\mathbb{Z})$? It is the orientation-preserving symmetric group of a cube or octahedron. It order is 24. In fact G is isomorphic to S_4. (Consider the pairs of antipodal vertices).

The quaternion group. This is a nonabelian group of order 8: $\langle i, j \; : \; i^2 = j^2 = (ij)^2 = e \rangle$.

Finite fields. What is $\mathrm{GL}_2(\mathbb{F}_2)$? It permutes the nonzero vectors in $\mathbb{F}_2$ in any way; so G is isomorphic to S_3.

The symmetric group. The symmetric group S_n. We write $(a_1, \ldots, a_r)$ for the cyclic permutation with $\sigma(a_i) = a_{i+1}$. We have $|S_n| = n!$

The (adjacent) transpositions generate S_n. Any permutation $\sigma \in S_n$ can be written uniquely as a product of disjoint cycles. The parity of a k-cycle is $k+1 \bmod 2$. Thus it is easy to calculate the parity of an given element:

$$\text{sign}(\sigma) = (\text{number of cycles}) + \sum (\text{lengths of each}) \bmod 2.$$

Proof that this is correct: if we multiply a product of cycles by a transposition (ij), we either meld or break apart the cycle(s) containing i and j.

Examples of subgroups. $\text{GL}_2(\mathbb{R})$ contains $\text{SL}_2(\mathbb{R})$ which in turn contains AN, A and N (isomorphic to $\mathbb{R}$ and $\mathbb{R}^*$). $\text{GL}_2(\mathbb{R})$ also contains the $f(x) = ax + b$ group as the matrices $\left(\begin{smallmatrix} a & b \\ 0 & 1 \end{smallmatrix}\right)$ acting on the line $y = 1$.

Subgroups of $\mathbb{Z}$. Every subgroup of $\mathbb{Z}$ has the form $H = a\mathbb{Z}$, $a \geq 0$. We have $a\mathbb{Z} \subset b\mathbb{Z}$ iff $b|a$. Thus the lattice of subgroups is the same as the lattice $\mathbb{N}$ under divisibility.

It follows that

$$a\mathbb{Z} + b\mathbb{Z} = \gcd(a, b)\mathbb{Z}.$$

In particular this shows $\gcd(a, b) = ar + bs$. Similarly $a\mathbb{Z} \cap b\mathbb{Z} = \text{lcm}(a, b)\mathbb{Z}$. Thus the additive theory of $\mathbb{Z}$ knows the multiplicative theory.

Cyclic subgroups. Given an $a \in G$ there is a map $\phi : \mathbb{Z} \to G$ given by $\phi(n) = a^n$. Its kernel is $b\mathbb{Z}$ where b is the order of a (or zero).

Isomorphisms. Example: $\phi : \mathbb{R} \to \text{SO}_2(\mathbb{R})$ given by $\phi(\theta) = \left(\begin{smallmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{smallmatrix}\right)$ is an isomorphism (verification includes the addition laws for sine and cosine).

For any group G, $\text{Aut}(G)$ is also a group. Does $\mathbb{Z}$ know the difference between ± 1? Between 1 and 2? What about $\mathbb{Z}/7$? To answer these questions, one can first show $\text{Aut}(\mathbb{Z}) \cong \mathbb{Z}/2$; and $\text{Aut}(\mathbb{Z}/n) \cong (\mathbb{Z}/n)^*$.

We also have $\text{Aut}(\mathbb{Z}^n) \cong \text{GL}_n(\mathbb{Z})$; $\text{Aut}(S_n) = ???$

Inner automorphisms. There is always a map $G \to \text{Aut}(G)$ sending g to $\phi_g(x) = gxg^{-1}$. The placement of g^{-1} is important; check: $\phi_{gh}(x) = g(hxh^{-1})g^{-1} = \phi_g(\phi_h(x))$. These are called *inner automorphisms* of G.

The center. The natural map $G \to \text{Aut}(G)$ has kernel $Z(G)$, the *center* of G. Example: $Z(\text{GL}_n(\mathbb{R})) \cong \mathbb{R}^* I$; $Z(S_n)$ is trivial for $n \geq 3$. (Proof: use the fact that $\text{Fix}(ghg^{-1}) = g(Fix(h))$ and the fact that there are permutations with single fixed-points.)

Two elements of G are *conjugate* if $gx_1g^{-1} = x_2$ for some $g \in G$. Compare similarity: two matrices in $\text{GL}_n(\mathbb{R})$ are similar iff they are conjugate. Thus

conjugate elements are the same up to a 'change of basis'. For example, x_1 and x_2 have the same order.

Example. In D_8 (or D_{4n}), the diagonal flips are not conjugate to the vertical and horizontal flips. That is, f and rf are not conjugate.

Homomorphisms $\phi : G \to H$. Image and kernel are subgroups; $\phi(e) = e$, $\phi(x)^{-1} = \phi(x^{-1})$.

Examples:

1. $\det : \mathrm{GL}_n(\mathbb{R}) \to \mathbb{R}^*$, kernel is $\mathrm{SL}_n(\mathbb{R})$.

2. $\mathrm{sign} : S_n \to \mathbb{Z}/2$, kernel is A_n.

3. $\mathbb{Z} \to \mathbb{Z}/n$, kernel is $n\mathbb{Z}$.

Normal subgroups. Problem: is there a homomorphism $\phi : \mathrm{SL}_2(\mathbb{R}) \to H$ whose kernel is the group A of diagonal matrices? No. The kernel is always a *normal subgroup*. It is a *union of conjugacy classes. A group homomorphism can only kill properties that are conjugacy invariant.*

Example: the center $Z(G)$ is always normal.

Cosets. Let H be a subgroup of G. The sets $aH \subset G$ are called the (left) *cosets* of G. Any two cosets have the same cardinality, since we can multiply by a^{-1} to get back to H. If aH and bH both contain $x = ah_1 = bh_2$ then $a = bh_2 h_1^{-1}$ so $a \in bH$ and $b \in aH$; but $aHH = aH$ so $aH = bH$. Thus any 2 cosets are equal or disjoint. In fact the cosets are the equivalence classes of the relation $a \equiv b$ iff $a^{-1}b \in H$. In particular we have:

Theorem 9.1 (Lagrange) *The order $|H|$ of any subgroup of G divides $|G|$.*

Corollary 9.2 *The only group of prime order p is $\mathbb{Z}/p$.*

Corollary 9.3 *For any $\phi : G \to H$ we have $|G| = |\operatorname{Ker}\phi| \cdot |\operatorname{Im}\phi|$.*

Compare this to the dimension addition formula for linear maps. Over finite fields, they are the same theorem.

We can similarly form the right cosets $\{Ha : a \in G\} = H\backslash G$.

Theorem 9.4 *A subgroup H is normal iff $aH = Ha$ for all $a \in G$.*

Internal and external products. For any pair of groups we can form the (external) product $A \times B$. In this product, elements of A commute with elements of B. In particular, A and B are normal subgroups of their product.

We can also make an internal version.

Theorem 9.5 *If A and B are normal subgroups of G meeting only at the identity, and $AB = G$, then $G \cong A \times B$.*

Proof. We claim elements of A and B commute. To see this we consider the commutator $[a, b] = aba^{-1}b^{-1}$. Parenthesizing one way, we see this lies in A; the other way, we see it lies in B; hence it is the identity.

Define a map $\phi : A \times B \to G$ by forming the product of two elements. By commutativity this map is a homomorphism, and our assumptions imply it is injective and surjective. ∎

Compare this to the decomposition of a vector space as $V = A \oplus B$. Note that it is enough to require that A and B generate G, since the proof shows AB is a subgroup.

Quotient groups.

Theorem 9.6 *If $N \subset G$ is a normal subgroup, then G/N is itself a group with the group law $(aN) * (bN) = (aN)(bN) = abN$, and the map $\phi(a) = aN$ is a homorphism with kernel N.*

This leads to the 'first isomorphism theorem':

Theorem 9.7 *If $\phi : G \to H$ is a surjective homomorphism, then H is naturally isomorphic to $G/\operatorname{Ker}\phi$.*

10 Symmetry

Group actions. Group theory emerged as an abstraction of the notion of symmetry. This notation is made systematic by the idea of a group acting on a set. Such an action is given by a map $G \times S \to S$, written $g(s)$ or gs or $g \cdot s$, satisfying $(gh)(s) = g(h(s))$ and $e(s) = s$. Equivalently, a group action is given by a homomorphism $G \to \operatorname{Sym}(S)$.

The *orbit* of $s \in S$ is the set Gs.

The *stabilizer* $\operatorname{Stab}(s) = G^s = \{g \in G : gs = s\}$ is a subgroup of G.

If s and t are in the same orbit, then G^s is conjugate to G^t. The following result is useful for counting orbits.

Theorem 10.1 *We have* $|G| = |Gs| \cdot |\operatorname{Stab}(s)|$.

We also define, for any set $A \subset S$, $\operatorname{Stab}(A) = \{g \in G : g(A) = A\}$.

Euclidean motions. As our first example we let $G = \operatorname{Isom}(\mathbb{R}^n)$. An isometry fixing the origin is a linear map (check this!) and hence an orthogonal transformation. Thus any $g \in G$ has the form

$$g(x) = T(x) + a$$

where $T \in O_n(\mathbb{R})$ and $a \in \mathbb{R}^n$. The group G can be represented by matrices of the form $\left(\begin{smallmatrix} T & a \\ 0 & 1 \end{smallmatrix} \right) \in \operatorname{GL}_{n+1}(\mathbb{R})$. Alternatively, we can let (T, a) represent g. Then we have the group law

$$(S, b)(T, a)x = S(Tx + a) + b = STx + Sa + b = (ST, Sa + b).$$

This presents G as the *semidirect product* of $O_n(\mathbb{R})$ and $\mathbb{R}^n$.

Motions of the plane. Now let us focus on the case of $\mathbb{R}^2$. In addition to rotations and translations, we have reflections in a line L and *glide reflections*: reflection in L followed by translation along L. It is crucial to notice that these classes are *conjugacy invariant*.

Theorem 10.2 *Every isometry of* $\mathbb{R}^2$ *is of one of these 4 types.*

Proof. Let $g(x) = Tx + a$. If $T = I$ we have a translation. If $T \neq I$ but $\det(T) = 1$, then T is a rotation and $\det(T - I) \neq 0$. It follows that g has a fixed-point p, and hence T is a rotation.

Now assume $\det(T) = -1$. Then $x \mapsto Tx$ is reflection, say through a line L; and g is the composition of a reflection and a translation. It is useful to write $a = b + c$ where b is orthogonal to L and c is parallel to L. Then $Tx + b$ is a reflection through $L' = L + b/2$, and thus g is the composition of reflection through L' with translation by c along L'. This is an ordinary reflection if $c = 0$ and a glide reflection if $c \neq 0$. ■

Theorem 10.3 *Every finite subgroup of* $O_2(\mathbb{R})$ *is cyclic or dihedral.*
The same is true for $\operatorname{Isom}(\mathbb{R}^2)$.

Proof. For the first statement, look at the smallest rotation in the group. For the second, observe that the barycenter $\beta(S) = |S|^{-1} \sum_S s$ satisfies $\beta(gS) = \beta(S)$ for any $g \in \operatorname{Isom}(\mathbb{R}^n)$; so for any x, $p = \beta(Gx)$ is a fixed-point of G. ■

Discrete groups. Next we analyze discrete subroups $G \subset \text{Isom}(\mathbb{R}^2)$.

Associated to any group G of isometries of $\mathbb{R}^2$, we have a translational subgroup $L \subset G$ and a rotational quotient group $DG \subset O_2(\mathbb{R})$.

Theorem 10.4 *If $L \subset \mathbb{R}^n$ is discrete, then it is generated by a set of linearly indepndent vectors, and L is isomorphic to $\mathbb{Z}^k$ for $1 \leq k \leq n$. Conversely, any set of linearly independent vectors generates a discrete subgroup of $\mathbb{R}^n$.*

Proof for $n = 2$. Consider the shortest vector, then the next shortest linearly independent vector, and use discreteness. ∎

Theorem 10.5 *There are 2, 4 or 6 shortest vectors in a lattice $L \subset \mathbb{R}^2$.*

Corollary 10.6 *The symmetry group of a lattice is typically $\mathbb{Z}/2$, but it can also be D_4 (rectangular or checkerboard case), D_6 (hexagonal case) or D_8 (square case).*

Theorem 10.7 *We have $g(L) = L$ for all $g \in DG$. Thus the rotational part DG of an infinite discrete subgroup of $\text{Isom}(\mathbb{R}^2)$ has order $1, 2, 3, 4$ or 6, and DG is cyclic or dihedral.*

Proof. If $Dg = A$ then $g = SA$ where S is a translation; consequently for any $T_a(x) = x + a \in L$ we have

$$gkg^{-1} = SAT_aA^{-1}S^{-1} = ST_{A(a)}S^{-1} = T_{A(a)}.$$

For the second part, let P the the convex hull of the shortest vectors in L. Then P is either an interval, a rectangle/square or a regular hexagon. Then DG must be contained in the symmetry group $\text{Isom}(P)$, which is D_4, D_8 or D_12. ∎

Example. You cannot order wallpaper with order 5 symmetry. (See however Penrose tilings!)

Constraints on lattices. Every lattice L has a rotational symmetry of order 2, but only the square and hexagonal lattices have symmetries of order 3, 4 or 6.

A lattice with a reflection symmetry must be rectangular *or* rhombic (the center of each rectangle is added).

The 17 types of wallpaper groups. A discrete subgroup $G \subset \mathrm{Isom}(\mathbb{R}^2)$ whose translational part L is a lattice is called a *wallpaper group.* There are 17 types of these. They can be classified according to their 'point groups'.

One might think that the point group DG almost determines G, but this is not quite right because the lift of a reflection from DG to G is not uniquely determined. We can always assume a given rotational element of DG fixes the origin.

1. (1) $DG = \langle 1 \rangle$. Then $G = L$.

2. (1) $DG = \mathbb{Z}/2$ acting as a rotation. Any L has such an extension. The points of $L/2$ are the poles of rotation. There are 4 types of poles.

3. (4) $DG = \mathbb{Z}/2$ acting as a reflection. There are 4 possiblities: (a) one type of reflection, (b) two types of parallel reflections, (c) a glide reflection and (d) a reflection and a glide reflection.

4. (3) $DG = \mathbb{Z}/3$, $\mathbb{Z}/4$, $\mathbb{Z}/6$. There are unique lattices that fit with these (rotational) symmetries.

5. (3) $DG = D_4 \cong \mathbb{Z}/2 \times \mathbb{Z}/2$. The points of $L/2$ are poles of rotation. In the simplest case, the lattice is rectangular and we add reflection through a line passing through the poles. But we can also reflect in a line passing midway between 2 poles. The trickiest case requires a 'checkerboard' lattice with a reflection through a pole.

6. (2) $DG = D_6$. The reflection group of the (60,60,60) triangle; and this example with face-center rotations adjoined.

7. (2) $DG = D_8$. The reflection group of the (45,45,90) triangle and this example with long-edge center rotations adjoined.

8. (1) $DG = D_{12}$. This is the reflection group of the $(30, 60, 90)$ triangle.

The wallpaper groups can also be classified topologically by forming the orbifold $X = \mathbb{R}^2/G$. The underlying space can be:

1. (4) The sphere; we get the $(2, 2, 2, 2)$, $(2, 4, 4)$, $(3, 3, 3)$ and $(2, 3, 6)$ orbifolds.

2. (8) The disk with mirror boundary. Then its double is a sphere as above. We write $(a_i|b_i)$ for a_i-type points in the interior and b_i on the boundary. The possibile signatures are:

$(|2, 2, 2, 2)$; $(2|2, 2)$; $(2, 2|)$
$(|2, 4, 4)$; $(4|2)$
$(|3, 3, 3)$; $(3|3)$
$(|2, 3, 6)$.

3. (1) The projective plane with signature $(2, 2)$.

4. (4) The torus, Klein bottle, cylinder and Möbius band.

Platonic solids. The symmetry group of a polyhedron $S \subset \mathbb{R}^3$ is its stabilizer in $\mathrm{Isom}^+(\mathbb{R}^3)$. (Note that we have required determinant one, so the symmetry can be seen by a rigid motion of the polyhedron.)

We say S is a *Platonic solid* if its symmetry group acts transitively on its vertices, edges and faces. As a consequence, the number of vertices, faces and edges must divide the order of the symmetry group.

There are exactly 5 Platonic solids: the tetrahedron, the cube, the octahedron, the dodecahedron and the icosahedron.

1. *The tetrahedron.* The symmetry group of a tetrahedron is A_4; it can be described as the orientation-preserving permutations of the vertices.

2. *The cube.* The symmetry group of a cube has 24 elements, since there are 6 faces each with stabilizer of order 4.

   In fact G is isomorphic to S_4, acting on the long diagonals! To see this, note that a rotation fixing a face gives the permutation $\sigma = (1234)$, and a rotation fixing an edge gives the permutation (12). These two elements together generate S_4.

3. The cube is dual to the octahedron.

4. *The dodecahedron.* How large is the symmetry group of a dodecahedron? A face has stabilizer of order 5, and there are 12 faces, so $|G| = t \times 12 = 60$. Similarly there are 30 edges (since each has stabilizer 2) and 20 vertices (since 5 faces come together at each).

   It turns out we have $G \cong A_5$. To see this, one can find *5 cubes* whose vertices lie on the vertices of a dodecahedron. There are 20 vertices all

together, and each belongs to two cubes — which works out, since 5 cubes have $5 \cdot 8 = 40$ vertices all together.

It is important to note that not every symmetry of an inscribed cube extends to a symmetry of the dodecahedron. In fact we have $S_4 \cap A_5 = A_4$ under the embedding. A given cube has 'clockwise' and 'counterclockwise' vertices: see Figure 4.

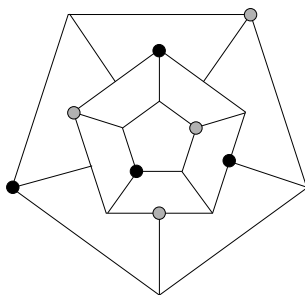5. The dodecahedron is dual to the icosahedron.

Figure 4. A cube inscribed in a dodecahedron has 2 types of vertices.

A non-Platonic solid. The rhombic dodecahedron is *not* a Platonic solid. All its 12 faces are equivalent, and their stabilizer is of order 2, so $|G| = 24$. There are 14 vertices, but they are *not* all equivalent! In fact they fall into two classes of sizes $6 + 8 = 14$, and each of those divides 24.

	G	$|G|$	V	E	F	$V - E + F$
Tetrahedron	A_4	12	4	6	4	2
Cube	S_4	24	8	12	6	2
Octahedron	S_4	24	6	12	8	2
Dodecahedron	A_5	60	20	30	12	2
Icosahedron	A_5	60	12	30	20	2
Rhombic Dodecahedron	S_4	24	6+8=14	12	12	2

Kepler's Cosmology. Kepler believed that the orbits of the planets were determined by the Platonic solids. Each eccentric orbit determines a thickened sphere or orb, centered at the Sun, that just encloses it. The 5 Platonic

solids thus correspond exactly to the gaps between the 6 planets known at that time. Between the orbits of Saturn and Jupiter you can just fit a cube; between Jupiter and Mars, a tetrahedron; between Mars and Earth, a dodecahedron; between Earth and Venus an icosahedron, and between Venus and Mercury, an octahedron.

This theory is discussed in the *Mysterium Cosmographicum*, 1596. Using the astronomical data of Copernicus, Kepler found:

	Predicted	Actual
Jupiter/Saturn	577	635
Mars/Jupiter	333	333
Earth/Mars	795	757
Venus/Earth	795	794
Mercury/Venus	577	723

See 'Kepler's Geometrical Cosmology', J. V. Field, University of Chicago Press, 1988.

Finite subgroups of $\mathrm{SO}_3(\mathbb{R})$. A modern, and more abstract version of the classification of Platonic solids is the classification of finite groups $G \subset \mathrm{SO}_3(\mathbb{R})$.

Theorem 10.8 *These finite subgroups of* $\mathrm{SO}_3(\mathbb{R})$ *are abstractly isomorphic to* $\mathbb{Z}/n$, D_{2n}, A_4, S_4 *or* A_5. *These subgroups are rigid: any 2 finite subgroups of* $\mathrm{SO}_3(\mathbb{R})$ *which are isomorphic abstractly are conjugate in* $\mathrm{SO}_3(\mathbb{R})$.

Compare the case of lattices $\mathbb{Z}^2 \subset \mathrm{Isom}(\mathbb{R}^2)$: these are *not* rigid, but they become rigid when more rotational symmetries are added.

Here is part of the proof.

Theorem 10.9 *If* $G \subset \mathrm{SO}_3(\mathbb{R})$ *is a finite group, then* G *is* $\mathbb{Z}/n$, D_{2n} *or a group of order 12, 24 or 60.*

Proof. Let $P \subset S^2$ be the set of poles of G, i.e. the endpoints of the axes of nontrivial rotations $G' \subset G$. Clearly G acts on P, say with orbits $P_1, \ldots, P_m$. Let N_i be the order of the stabilizer of a point in P_i. Now every $p \in P_i$ is the pole of $N_i - 1$ elements in G', and every element of G' has 2 poles. Thus if

we count up the number S of pairs (g, p) in $G' \times P$ such that p is a pole of g, we get

$$S = 2|G'| = 2|G|(1 - 1/|G|) = \sum |P_i|(N_i - 1) = |G| \sum (1 - 1/N_i).$$

In other words, we have

$$\sum (1 - 1/N_i) = 2 - 2/N < 2,$$

where $N = |G|$.

Now each term on the left is at least $1/2$, so the number of terms is 1, 2 or 3. And in fact (unless G is trivial) there must be at least 2 terms, because $1 - 1/N_1 < 1$ and $2 - 2/N \geq 1$.

If there are exactly 2 terms we get $2/N = 1/N_1 + 1/N_2$, and N_i divide N. The only possibility is $N_1 = N_2 = N$ and then G is cyclic.

If there are exactly 3 terms then we get $1 + 2/N = 1/N_1 + 1/N_2 + 1/N_3$. The possible solutions are $(N_1, N_2, N_3) = (2, 2, N/2)$, $(2, 3, 3)$, $(2, 3, 4)$ and $(2, 3, 5)$. In the first case G is dihedral, and in the last 3 cases we get $N = 12, 24$ or 60. ∎

Euler characteristic. From a topological point of view, what's going is that $\chi(S^2/G) = 2/|G| = 2 - \sum (1 - 1/N_i)$.

Abstract group actions. Any transitive action of G on a set S is isomorphic to the action by left translation of G on G/H, where $H = \text{Stab}(x)$.

This identification *depends* on the choice of x. The stabilizer of $y = gx$ is the conjugate gHg^{-1} of H; this observation applies in particular to the stabilizer of gH.

The stabilizers of the points of S are the conjugates of H.

Example: D_3 acts on the 3 vertices of a triangle with $H = \langle f \rangle$. The other vertices correspond to $\langle rf \rangle$ and $\langle r^2 f \rangle$, the conjugates of H.

Note however that H may fix many points in S, so there is not a bijection between the points of S and the conjugates of H in general.

The class equation and its consequences. The fact that G *acts* on G by translation (on the right or the left) is not very useful, since there is only one orbit.

But the fact that G also acts by *conjugation* is very useful. In this case the *orbit* of $x \in G$ is its *conjugacy class* $C(x)$, and its *stabilizer* is its *centralizer*. We have $|C(x)| \cdot |Z(x)| = |G|$.

Now recall that the partition of G into orbits of G yields the *class equation*

$$|G| = \sum C_i,$$

where C_i are the cardinalities of the conjugacy classes in G. Each C_i divides $|G|$.

We also know that we can take $C_1 = 1$, i.e. the conjugacy class of the identity is trival. More generally, the set of elements with $|C(x)| = 1$ forms the center $Z(G)$.

Examples: the symmetries G of the cube are comprised of the identity (1), elements of order 4 (6), elements of order 3 (8), and elements of order 2 flipping edges (6), and elements of order 2 spinning faces (3). We have

$$|G| = 24 = 1 + 6 + 8 + 6 + 3.$$

Note that an element of order 4 is determined by the choice of a face, which is then rotated 1/4 of a turn. The opposite face is rotated 3/4.

For the dodecahedron, the conjugacy class is almost determined by the order; we have elements of orders $1, 2, 3$ and 5, but the last fall into 2 types. The class equation is

$$|G| = 60 = 1 + 15 + 20 + 12 + 12.$$

Note that an element of order 5 is determined by a face plus a choice of whether to rotate 1/5 or 2/5 of a turn.

Note that $|C_i|$ must divide the order of G. Indeed, $|C_i| = |Z(g)|$ for any $g \in C_i$.

Simple groups. The class equation also helps us to recognize *simple groups*, i.e. groups with no normal subgroups. A general finite group is built up out of finite simple groups, so their classification is particularly important. By considering the conjugacy classes of G contained in N, we find:

Theorem 10.10 *Let $|G| = \sum C_i$ be the class equation. If G has a proper normal subgroup, then there is a proper sub-sum containing $C_1 = 1$ such that $\sum C_i'$ divides $|G|$.*

Corollary 10.11 *The icosahedral group A_5 has no normal subgroup.*

Proof. Recall its class equation is $60 = 1 + 15 + 20 + 12 + 12$. There is no sub-sum of the required type. ∎

Riddle. What is the next term in the sequence

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59?$$

(Ans: 60.)

Burnside counting theorem. We now return to the setting of G acting on a set S. If the action of G on S is not necessarily transitive, then S breaks up into orbits. We write the space of orbits as S/G, although $G \backslash S$ would be more accurate. Let $S^g = \{x \in S : gx = x\}$.

We then have the following result of Burnside:

Theorem 10.12 *The number of orbits $|S/G|$ is given by $|G|^{-1} \sum_G |S^g|$.*

Equivalently, the number of orbits is the same as the average number of fixed points.

Proof. We consider the function on $G \times S$ which is 1 if $gx = x$ and otherwises 0. Summing one way gives $\sum_G |S^g|$; the other way gives $\sum_S |\operatorname{Stab}(x)|$. Now $|\operatorname{Stab}(x)|$ is constant on each orbit, so the second sum can be rewritten as

$$\sum_S |\operatorname{Stab}(x)| = \sum_{S/G} |\operatorname{Stab}(x)||Gx| = \sum_{S/G} |G| = |S/G||G|.$$

$\blacksquare$

What makes this formula useful is that $S^g = S^h$ whenever g and h are conjugate. So we can organize the sum over G as a sum over conjugacy classes.

Example: coloring the cube. (The Heinz problem.) How many rotationally distinct ways are there to color the faces of a cube with 3 colors? Let S be the set of all colorings and let $G \cong S_4$ be the symmetries of the cube. Then $|S^g| = 3^n$ where n is the number of orbits of faces under g.

There are 5 cases to consider. The unique element of order 1 contributes 3^6. The 6 elements of order 4 each contribute 3^3. The 8 elements of order 3 each contribute 3^2. The 6 edge-type elements of order 2 each contribute 3^3. And the 3 face-type elements of order 2 each contribute 3^4 as well. So we get:

$$|S/G| = (1/24)(1 \cdot 3^6 + 6 \cdot 3^3 + 8 \cdot 3^2 + 6 \cdot 3^3 + 3 \cdot 3^4) = 1368/24 = 57.$$

Semidirect products (This is a supplement to Artin, which does not treat this topic.)

Here is a very important generalization of the notion of a product of groups. It uses the idea that both G and $\mathrm{Aut}(G)$ are subgroups of the permutation group $\mathrm{Sym}(G)$.

Let G and H be groups, and let $\phi : H \to \mathrm{Aut}(G)$ be a homomorphism. With ϕ understood, it is convenient to use the notation

$$g^h = \phi(h)(g);$$

the fact that $\phi(h)$ is an automorphism of G implies $(g_1 g_2)^h = g_1^h g_2^h$.

We can now form a new group $P = G \rtimes_\phi H$, the *semidirect product* of G and H. (Often the ϕ is understood and suppressed.)

As a set, P consists of all pairs (g, h). We interpret such a pair as the product gh; thus any element of P can be uniquely expressed as such a product.

What about products in the other order? The crucial relation is that we define:
$$hg = g^h h.$$

Using this relation, any product of elements of P can be rewritten uniquely in the form gh again. For example:

$$g_1 h_1 g_2 h_2 = g_1 g_2^{h_1} h_1 h_2.$$

An equivalent and more formal defintion is that P consists of the ordered pairs (g, h) with the group law $(g_1, h_1)(g_2, h_2) = (g_1 g_2^{h_1}, h_1 h_2)$.

Note that G is always a *normal* subgroup of $P = G \rtimes H$, and $P/G \cong H$. As a universal example, one can always take $G \rtimes \mathrm{Aut}(G)$.

Recognizing semidirect products. The basic result is:

Theorem 10.13 *Let P be a group containing subgroups G and H such that:*

1. G is normal,

2. $GH = \{gh : g \in G, h \in H\} = P$, and

3. $G \cap H = \{e\}$.

Then P is isomorphic to $G \rtimes_\phi H$, where $\phi(h)(g) = hgh^{-1} \in G$.

Remark. Since G is normal, GH is always a *subgroup* of S. This often helps in verifying that $GH = S$.

Splitting exact sequences. We say a surjective homomorphism $\pi : A \to B$ *splits* if it has a 'section' $\xi : B \to A$, i.e. a homomorphism satisfying $\pi \xi(b) = b$ for all $b \in B$.

Theorem 10.14 *Let N be a normal subgroup of G. Then G is the semidirect product of N and G/N iff the exact sequence*

$$0 \to N \to G \to G/N \to 0$$

splits.

Examples.

1. The group $\mathbb{Z}/n$ admits an automorphism of order 2 given by $\alpha(k) = -k$. Thus there is a map $\phi : \mathbb{Z}/2 \to \mathrm{Aut}(\mathbb{Z}/n)$. The resulting semidirect product is the usual dihedral group:

$$D_n = \langle r, \rho : r^2 = \rho^n = 1, r\rho r^{-1} = \rho^{-1} \rangle.$$

2. If ϕ is trivial, so $g^h = g$, then the semidirect product becomes the ordinary product of groups $G \times H$.

3. The group of motions of the plane is given by $M = \mathbb{R}^2 \rtimes O_2(\mathbb{R})$.

4. The orthogonal group itself is the semidirect product $O_2 = SO_2(\mathbb{R}) \rtimes \mathbb{Z}/2$, since we have $r\rho_\theta r^{-1} = \rho_{-\theta}$.

5. The group $AN \subset \mathrm{SL}_2(\mathbb{R})$ is the semidirect product of A and N.

6. The group $\mathbb{Z}/7$ admits an automorphism of the form $\alpha(k) = 2k$. We have $\alpha(\alpha(\alpha(k))) = 8k = k \bmod 7$, so α has order 3 in $\mathrm{Aut}(\mathbb{Z}/7)$. Thus there is a *nonabelian* group of order 21 given by $S = \mathbb{Z}/7 \rtimes_\phi \mathbb{Z}/3$.

7. The sequence $0 \to \mathbb{Z} \xrightarrow{2} \mathbb{Z} \to \mathbb{Z}/2$ does not split.

We recall from Theorem 9.5 that as a special case: if G is generated by a pair of normal subgroups A and B meeting only at the identity, then G is isomorphic to $A \times B$.

11 Finite group theory

Classification of finitely-generated abelian groups. We begin with the classification of finite abelian groups. The finitely-generated case is no more difficult. (Note that infinitely-generated abelian groups like $\mathbb{Q}$ can have a complicated structure.)

Theorem 11.1 *Any finitely-generated abelian group is isomorphic to a product of cyclic groups.*

Proof. Consider a presentation matrix $T : \mathbb{Z}^r \to \mathbb{Z}^g$ such that $G \cong \mathbb{Z}^g/T(\mathbb{Z}^r)$, and then apply row and column operations to diagonalize it. ∎

Note that $\mathbb{Z}/a \times \mathbb{Z}/b \cong \mathbb{Z}/ab$ whenever $\gcd(a, b) = 1$. Because of this, the factorization of G into cyclic groups is not quite unique.

There are 2 standard forms for the torsion part $\mathbb{Z}/a_1 \times \ldots \times \mathbb{Z}/a_n$ of the product. In one form, $a_1|a_2|\cdots a_n$. These integers are called the *elementary divisors* of G. In the other form, we have $a_i = p_i^{e_i}$ where p_i are prime. The same prime can appear multiple times.

In both cases, the numbers (a_i) are invariants of G. For example, in the elementary divisor case, a_n is the largest order of an element in G.

See Artin, Chapter 12.6 for a treatment of this classification in the context of modules.

Classification of small groups. We will next explore some methods for classifying groups of 'small order'. We already understand the groups whose order is prime. The first 'open' case is $|G| = 4$.

Lemma 11.2 *Any group G in which every element has order 1 or 2 is abelian. In fact, G is a vector space over $\mathbb{F}_2$.*

For $|G| = 4$ we either have a cyclic group $\mathbb{Z}/4$ or, by the preceding Lemma, the field $\mathbb{Z}/2 \times \mathbb{Z}/2$.

A similar argument shows for $|G| = 6$ there are 2 just groups, $\mathbb{Z}/6$ and S_3. E.g. we know G must have an element of order 3 or 6, else it would be abelian.

The classification of groups of order 8 is left as an exercise. We will next show any group of order 9 (or order p^2) is abelian. (There are 2 such groups for each prime.)

Nilpotent groups. The class of nilpotent groups is the smallest class of groups containing the abelian groups and with the property that G is nilpotent if $G/Z(G)$ is nilpotent. For example, S_3 is *not* nilpotent (its center is trivial) but D_8 *is* nilpotent (the 'antipodal map' r^2 lies in the center and yields as quotient $\mathbb{Z}/2 \times \mathbb{Z}/2$). In the same way the quaternion group Q is nilpotent.

(The terminology is explained by examples of linear groups where every element has the form $I + N$, N a nilpotent matrix).

p-groups. If $|G| = p^e$ is a power of a prime, G is a *p-group*. The class equation enables one to analyze p-groups rather well:

Theorem 11.3 *Every p-group is nilpotent.*

Proof. The center of G must be nontrivial, else the class equation would give $|G| = 1 \bmod p$. And then $G/Z(G)$ is again a p-group. ∎

Theorem 11.4 *Every group of order p^2 is abelian.*

As we have seen, this implies G is isomorphic to $\mathbb{Z}/p \times \mathbb{Z}/p$ or $\mathbb{Z}/p^2$.

Proof. If $|Z(G)| = p^2$ we are done, so we may assume $|Z(G)| = p$. Now we use a basic fact: for any $x \in G$, x and $Z(G)$ together generated an abelian subgroup. So taking any $x \notin Z(G)$, we get an abelian subgroup with $|A| > p$, and hence $|A| = p^2 = |G|$. ∎

Note that D_8 and Q are nonabelian groups of order p^3. The upper-triangular matrices in $\mathrm{GL}_n(\mathbb{F}_p)$ with 1's on the diagonal also give nice examples of nonabelian p-groups.

The Sylow theorems. Recall that a p-group is nilpotent. Remarkably, every finite group contains a p-group that is as large as possible, and this p-subgroup is unique up to conjugacy.

Theorem 11.5 (First Sylow) *Let $|G| = p^e m$ where p is prime and $\gcd(p, m) = 1$. Then G contains a subgroup H with $|H| = p^e$.*

Such an H is called a *p-Sylow subgroup* of G.

Example. The 'prime' example is the general linear group $G = \mathrm{GL}_n(\mathbb{F}_p)$. By considering bases it is easy to see that

$$|G| = (p^n - 1)(p^n - p) \cdots (p^n - p^{n-1}) = p^e m,$$

where $e = 0 + 1 + 2 + \cdots + (n-1)$. A p-Sylow subgroup of G is given by the group of upper-triangular matrices with 1's along the diagonal.

Theorem 11.6 (Second Sylow) *Any p-subgroup is contained in a p-Sylow subgroup, and all p-Sylow subgroups are conjugate.*

Theorem 11.7 (Third Sylow) *Suppose $|G| = p^e m$, $\gcd(p, m) = 1$. Then the number of p-Sylow subgroups in G divides m, and is congruent to $1 \bmod p$.*

Proof of Sylow I. Let $\mathcal{S} = \mathcal{P}_{p^e}(G)$ denote the set of all $S \subset G$ such that $|S| = p^e$. Then one can show that $|\mathcal{S}| = \binom{p^e m}{p^e}$ is not divisible by p. Now let G act on $\mathcal{S}$ by $g(S) = g \cdot S$ (*not* by conjugation!). Since $\mathcal{S}$ is partitioned into orbits of G, some orbit has size $|G \cdot S| = |G|/\operatorname{Fix}(S)$ which is not divisible by p. It follows that $p^e || H|$, where $H = \operatorname{Fix}(S)$.

On the other hand, by decomposing S into orbits of the *pointwise* action of H, we see $|H|$ divides $|S| = p^e$. It follows that $|H| = p^e$. ∎

The next proof exploits the following general observation, already used in the proof that p-groups are nilpotent. Namely if a p-group G acts on a set S, and S^G denotes the set of points fixed by every element of G, then we have

$$|S^G| = |S| \bmod p.$$

Indeed, we can decompose S into G-orbits, and any orbit which is not a single point has size divisible by p. The orbits which are single points, on the other hand, are exactly S^G.

Proof of Sylow II. Let P be a Sylow subgroup of G, and let H be a p-subgroup. We will show H is contained in some conjugate of P. This is the same as showing that H fixes some point when it acts on G/P. But $|G/P|$ is not divisible by p, so this follows from the basic lemma on orbits. ∎

Proof of Sylow III. The calculation of the number of Sylow subgroups is the same as the calculation of the number of conjugates of P.

Let $N_G(P)$ denote the normalizer of P, i.e. the elements of G that conjugate P to itself. (E.g. $N_G(P) = G$ iff P is a normal subgroup.) It is a general fact, true for any subgroup, that the number of conjugates is just $|G|/|N_G(P)|$.

In the case at hand, $|P| = p^e || N_G(P)|$, and thus the number of conjugates of P divides m.

For the last part, we consider the action of P by conjugation on the set S of all p-Sylow subgroups. We will show that $S^P = \{P\}$. This shows by the lemma that $|S| = 1 \bmod p$.

To finish the proof suppose Q is a p-Sylow subgroup and $xQx^{-1} = Q$ for all $x \in P$. Then P is contained in $N_G(Q)$. But then Q and P are both subgroups of $N_G(Q)$, so P and Q are conjugate by Sylow II. But Q is normal in $N_G(Q)$, so this means $P = Q$. ∎

Groups of order 15. A well-known application of the Sylow theory is to show that the only group of order 15 is $\mathbb{Z}/3 \times \mathbb{Z}/5$. For this we let n_p denote the number of p-Sylow subgroups. We have $n_3 = 1(3)$ and $n_3|5$ so $n_3 = 1$. Similarly $n_5 = 1(5)$ and $n_5|3$ so $n_5 = 1$. Thus we have normal cycle subgroups A and B of size 3 and 5, which can only meet in the identity. This implies (as we have seen) that $G \cong A \times B$.

More generally we have:

Theorem 11.8 *Let $|G| = pq$ be a product of distinct primes with $p < q$. Then either G is abelian — and hence isomorphic to $\mathbb{Z}/p \times \mathbb{Z}/q$ — or $p|q-1$ and G is the unique nonabelian group of this order.*

Proof. This is a mild generalization of the case $pq = 15$. We have $n_q = 1$ or p, but also $n_q = 1(q)$, so $n_q = 1$. Thus there exists a normal subgroup $N \cong \mathbb{Z}/q$. Similarly $n_p = 1$ or q. If $n_p = 1$ then we have a product group as before.

If $n_p = q$ then $q = 1(p)$, i.e. $p|q-1$. In this case we still have an exact sequence
$$0 \to \mathbb{Z}/q \to G \to \mathbb{Z}/p \to 0.$$

Since G has an element of order p, this sequence splits and we have a semidirect product, easily seen to be unique up to isomorphism. ∎

12 Representation theory

In this section we discuss the classification of representations of finite groups using characters.

Basic definitions. Let G be a finite group and let V be a finite-dimensional complex vector space. A *representation* of G is a homomorphism

$$\rho : G \to \mathrm{GL}(V).$$

We write $\rho(g)(v) = gv$ with ρ understood. Of course representations of larger groups, on infinite-dimensional spaces, are also interesting, but for the purposes of this section we restrict to the case above.

Two representations are *equivalent* (consider the same) if there is a linear isomorphism $T : V \to V'$ such that $T(gv) = g'T(v)$.

The regular representation. We let $L^2(G)$ denote the space of all functions $\phi : G \to \mathbb{C}$ with the inner product

$$\langle \phi(g), \psi(g) \rangle = |G|^{-1} \sum_G \phi(g)\overline{\psi}(g).$$

The action of G on $V = L^2(G)$ is called the *regular representation* of G. Formally it is given by

$$(g \cdot \phi)(x) = \phi(g^{-1}x),$$

or

$$\left(g \cdot \sum a_h h\right) = \sum a_h gh = \sum a_{g^{-1}h} h.$$

This representation is *unitary*.

The group ring. We can extend the definition above to make $L^2(G)$ into a *ring* $\mathbb{C}[G]$. Any representation of G extends to a homomorphism of rings $\rho : \mathbb{C}[G] \to \mathrm{End}(V)$. Conversely, a representation of G is the same as a module over the group ring.

Complete irreducibility. A representation is *irreducible* if the only invariant subspaces are V and (0), and $V \neq (0)$. Its character is then an *irreducible character*.

A representation is *unitary* if there is a (Hermitian) inner product $\langle v, w \rangle$ on V preserved by ρ; i.e. satisfying $\langle gv, gw \rangle = \langle v, w \rangle$.

Theorem 12.1 *Any representation of a finite group preserves an inner product on V. Thus we can think of the target of ρ as the unitary group U_n.*

Corollary 12.2 *Any representation of a finite group is a direct sum of irreducible representations.*

Characters. The *character* of (ρ, V) is the function

$$\chi(g) = \operatorname{tr} \rho(g).$$

Evidentally isomorphic representations have the same character. One of the main point of representation theory is the converse. Thus characters give a concrete way to record and distinguish representations.

Here are some general facts about characters.

1. We have $\chi(e) = \dim V$.

2. We have $\chi(g^{-1}) = \overline{\chi}(g)$.

3. Similarly $\overline{\chi}(g)$ is the character of the adjoint representation, $g \mapsto \rho(g^{-1})^*$. (Using duality or an invariant Hermitian form.)

4. If $\dim V = 1$, the character is a homomorphism: we have

5. In general, χ is a *class function*: it satisfies

$$\chi(ghg^{-1}) = \chi(h),$$

so it is constant on conjugacy classes.

Examples of characters.

1. For the regular representation, $\chi(e) = |G|$ and $\chi(g) = 0$ for $g \neq e$.

2. For the action of S_n on $\mathbb{C}^n$, $\chi(g) = |\operatorname{Fix}(g)|$.

3. For the action of $\mathbb{Z}/n$ by rotations on $\mathbb{R}^2 \subset \mathbb{C}^2$, with k rotating by k/n, we have
$$\chi(k) = 2\cos(2\pi k/n).$$
It is useful to note that the trace is 1, 0 and -1 for rotations by $60°$, $90°$ and $120°$; and also that any reflection on $\mathbb{R}^2$ has $\chi(g) = 0$.

4. For the action A_4 on $\mathbb{R}^3$ by the symmetries of the cube we have:

A_4	(1)	(3)	(4)	(4)
	e	f	r	r^2
χ	3	-1	0	0.

Note that we have given one entry for each conjugacy class, and recorded the order of the conjugacy class along the top.

Duality for abelian groups. Now suppose G is a finite abelian group.

We define the *dual* of G by

$$\widehat{G} = \mathrm{Hom}(G, S^1).$$

Note that this is indeed a group. For $G = \mathbb{Z}/n$ we see $\widehat{G}$ is isomorphic to the nth roots of unity, and thus $\widehat{G} \cong G$. This holds for any abelian group; however the isomorphism is not canonical, it depends on a choice of generators for G.

Any character determines a 1-dimensional representation by $\rho(g)v = \chi(g)v$. Conversely, χ can be recovered as the trace character of ρ. Thus:

Theorem 12.3 *The set $\widehat{G}$ coincides with the characters of 1-dimensional representations of G.*

Theorem 12.4 *In the abelian case, every irreducible representation is 1-dimensional.*

Proof. It suffices to show that every $g \in G$ acts by a multiple of the identity. For this consider any $g \in G$ and, using the fact that $g^n = e$, decompose V into eigenspaces $\oplus V_\lambda$ of g. These spaces must be preserved by every $h \in G$ because of commutativity. Thus g has only one eigenvalue, so it acts by a multiple of the identity. ∎

Orthogonality. As functions the irreducible characters $\chi_i \in \widehat{G}$ can be thought of as elements of $L^2(G)$. Extending by linearity, we obtain a natural map

$$L^2(\widehat{G}) \to L^2(G)$$

defined by

$$\sum a_i \chi_i \to \sum a_i \chi_i(g).$$

Theorem 12.5 (Orthogonality relations) *The 1-dimensional characters of a finite Abelian G form an orthonormal basis for $L^2(G)$.*

Proof. Given $\alpha, \beta \in \widehat{G}$ we note that $\alpha \overline{\beta} = \gamma \in \widehat{G}$ as well, and

$$\langle \alpha, \beta \rangle = |G|^{-1} \sum_G \gamma(g) = 1$$

if $\gamma(g)$ is trivial. For orthogonality, multiply the lefthand side by $\gamma(g) \neq 1$ and note that the answer is unchanged.

Thus we have an injective map $L^2(\widehat{G}) \to L^2(G)$. To see it is surjective, one can use the fact that $|G| = |\widehat{G}|$. $\blacksquare$

Note that the map $L^2(\widehat{G}) \to L^2(G)$ becomes an isometric isomorphism if we define $\|\chi_i\| = 1$. However this definition is inconsistent with the one given already for $L^2(G)$, since the norm has no factor of $|G|^{-1}$.

For an abelian group the regular representation decomposes into 1-dimensional representations V_i, and a basis element $\chi_i \in V_i$ normalized so that $\chi_i(e) = 1$ is just a character. This gives another proof of surjectivity and shows:

Theorem 12.6 *If G is an abelian group, then the regular representation decomposes as a direct sum of $|G|$ distinct 1-dimensional representations, one for each character of g.*

We also note:

Theorem 12.7 *There is a natural isomorphism $G \cong \widehat{\widehat{G}}$, sending g to the 1-dimensional character of $\widehat{G}$ defined by $\alpha_g(\chi) = \chi(g)$.*

Fourier transform. The preceding discussion shows that every function on G can be written as a sum of characters:

$$f(g) = \sum_{\widehat{G}} \widehat{f}(\chi)\chi(g).$$

The coefficients are given by

$$\widehat{f}(\chi) = \langle f, \chi \rangle$$

and are called the Fourier transform of G.

In the 1-dimensional case, we know we can write $V \cong \sum m_i V_i$ where V_i is the 1-dimensional representation associated to the character α_i. Consequently:

$$\chi(g) = \sum m_i \chi_i.$$

This shows:

Theorem 12.8 *The decomposition of a representation of an abelian group into its irreducible components is given by the Fourier transform of the trace character.*

Corollary 12.9 *Two representations are equivalent iff they have the same character.*

Example. The regular representation V_n of $G = \mathbb{Z}/n$ has character $\chi(0) = n$, $\chi(k) = 0$ for $k \neq 0$.

If $d|n$ then the regular representation V_d of $\mathbb{Z}/d$ occurs as well, with character $\chi(k) = d$ if $k = 0 \bmod d$, and $\chi(k) = 0$ otherwise.

How does the representation $W = \mathrm{Sym}^2(V)$ decompose when n is even? This is the same as the permutation representation on pairs of elements of G. This $\chi(k)$ is the number of fixed pairs. We find $\chi(0) = n(n+1)/2$ and $\chi(n/2) = n/2$. Thus $\mathrm{Sym}^2(V)$ is the sum of $n/2$ copies of V_n and one copy of $V_{n/2}$.

E.g. when $n = 4$ we have the orbits $(00, 11, 22, 33), (01, 12, 23, 03)$ and $(02, 13)$.

More general groups. In general every locally compact abelian group G has a dual group $\widehat{G}$ of the same time, consisting of the continuous characters $\chi : G \to S^1$. The Fourier transform sends $L^2(G)$ to $L^2(\widehat{G})$.

Examples of dual pairs: S^1 and $\mathbb{Z}$; $\mathbb{R}$ and $\mathbb{R}$. For the first the Fourier transform goes back and forth between functions on S^1 and the coefficients of their powers series $f(z) = \sum a_n z^n$. For the second, the characters are $\exp(itx)$.

The Γ-function. Note that e^{-t} and t^s are homomorphism from $\mathbb{R}$ and $\mathbb{R}_+$ into $\mathbb{C}^*$. The Γ-function gives the Fourier transform of e^{-t} as a function on $\mathbb{R}^*$: it is given by

$$\Gamma(s) = \int_0^\infty e^{-t} t^s \frac{dt}{t}.$$

Its analogue for $\mathbb{Z}/p$ is a Gauss sum,

$$\sigma(\chi) = \sum_{(n,p)=1} \chi(n) e^{2\pi i n/p},$$

where $\chi : (\mathbb{Z}/p)^* \to S^1$.

Irreducible characters. We will shortly see that G has only a finite number of isomorphism classes of irreducible representations. Let us enumerate

them $V_1, \ldots, V_m$ with V_1 the trivial representation, and let $\chi_1, \ldots, \chi_n$ be the corresponding characters.

The following is the most fundamental result in representations of finite groups.

Theorem 12.10 *The irreducible characters χ_i form an orthonormal basis for the class functions in $L^2(G)$ (functions constant on conjugacy classes).*

Corollary 12.11 *The number of irreducible representations of G is the same as the number of conjugacy classes of G.*

Corollary 12.12 *Any representation V with character χ decomposes as $V \cong \sum m_i V_i$ where $m_i = \langle \chi, \chi_i \rangle$.*

Corollary 12.13 *A representation is irreducible iff its character satisfies $\langle \chi, \chi \rangle = 1$.*

Corollary 12.14 *The dimensions $d_i = \dim V_i$ satisfy*

$$|G| = \sum d_i^2.$$

More precisely, the regular representation decomposes as

$$\mathbb{C}[G] \cong \sum d_i V_i.$$

Proof. The character of the regular representation is $\chi(e) = |G|$ and otherwise $\chi(g) = 0$. Thus $\langle \chi, \chi_i \rangle = \chi_i(e) = d_i$. ∎

It can also be shown that d_i divides $|G|$ but this is more difficult.

Corollary 12.15 *Two representations V and V' of G are equivalent iff $\chi = \chi'$.*

Character tables. Note that the theorem gives (i) a quick way to check if a representation is irreducible and (ii) a quick way to check if we have a complete list of irreducible representations. This makes it easy to generate a character table in simple examples, as illustrated below for $G = \mathbb{Z}/3$ and $G = D_3$. The numbers on the top row give the size of each conjugacy class.

These weights must be taken in to account when checking the orthgonality relations.

$\mathbb{Z}/3$	(1)	(1)	(1)
	e	r	r^2
χ_1	1	1	1
χ_2	1	ω	ω^2
χ_3	1	ω^2	ω

D_6	(1)	(3)	(2)
	e	f	r
χ_1	1	1	1
χ_2	1	-1	1
χ_3	2	0	-1

The group A_4 has a 3-dimensional geometric representation as the symmetries of a cube. It also admits a map $A_4 \to \mathbb{Z}/3$, so the representations of $\mathbb{Z}/3$ must occur inside those of A_4. Using this fact, its character table can be easily computed below.

A_4	(1)	(3)	(4)	(4)
	e	f	r	r^2
χ_1	1	1	1	1
χ_2	1	1	ω	ω^2
χ_3	1	1	ω^2	ω
χ_4	3	-1	0	0

A lot of information is encoded in these tables. For example, the first column gives the *dimension* of the representations. The *commutator subgroup* is the intersection of the kernels of all 1-dimensional characters. More generally, the kernel of any character is *normal*, and the intersections of these give all normal subgroups.

Schur's lemma. As usual in linear algebra, the central idea is to work in a basis-independent fashion. Since we implicitly identify representations that are isomorphic, it is important to have a good method to produce these isomorphisms.

Theorem 12.16 *If $T : V \to V'$ is a G-map between irreducible representations, then $T = 0$ unless $V \cong V'$. If $V = V'$ then T is a multiple of the identity.*

Proof. Since $\operatorname{Ker} T$ and $\operatorname{Im} T$ are invariant, by irreducibility T is an isomorphism or 0. If $V = V'$ then then the eigenspaces of T give further invariant

subspaces, so there can be only one. ■

Averaging and matrix coefficients. Here are 2 good averaging procedures to produce G-invariant maps. Suppose $S : V \to V'$ is any linear map between G-spaces. Then we can set

$$T = \frac{1}{|G|} \sum_G g' S g^{-1}$$

and it is easily verified that T is a G-map. So if V and V' are irreducible, T is a multiple of the identity, which can only be nonzero if $V \cong V'$. In this case we can assume $V = V'$ and then we find

$$\mathrm{tr}(T) = \mathrm{tr}(\lambda I) = \mathrm{tr}(S) = \dim(V)\lambda,$$

so we can compute the value of T from $\mathrm{tr}(S)$.

Matrix coefficients. If V is a unitary representation of G, then (for any basis of V) the functions $g \mapsto \langle g e_i, e_j \rangle$ are called the *matrix coefficients* of V. They span a subspace $c(V) \subset L^2(G)$ that is well-defined independent of the choice of basis or inner product. The span contains $\langle gv, w \rangle$ for any v and w of course.

Within $c(V)$ we have the character χ of V, since this is a sum of matrix coefficients. Thus the orthogonality of different characters follows from:

Theorem 12.17 *If V and V' are distinct irreducible representations of G, then $c(V)$ and $c(V')$ are orthogonal in $L^2(G)$.*

Proof. Let $S : V \to V'$ be any linear map. Then we have $\sum (g')^{-1} S g = 0$, and thus applying to v and taking the inner product with v' we get

$$0 = \sum_G \langle S(gv), g'v' \rangle$$

using the unitarity to move g' over to v'. Now suppose $S(x) = \langle x, w \rangle w'$. Then this gives

$$\sum_G \langle gv, w \rangle \langle w', g'v' \rangle = \sum_G \langle gv, w \rangle \overline{\langle g'v', w' \rangle} = 0.$$

But this is just the inner product of 2 matrix coefficients, considered as functions on G. ■

Theorem 12.18 *If V is irreducible then $\langle \chi, \chi \rangle = 1$.*

Proof. Let $(e_i)_1^d$ be an orthonormal basis for V. Then we can write $\chi = \sum \langle ge_i, e_i \rangle = \sum \xi_i$. To complete the proof is it enough to show that $\langle \xi_i, \xi_j \rangle = \delta_{ij}/d$.

To this end, let $S(v) = \langle v, e_i \rangle e_j$. Then $\operatorname{tr}(S) = 1$ and so $\operatorname{tr}(T) = 1$, where $T = |G|^{-1} \sum_G g^{-1} S g$. Consequently $Tv = v/d$ by Schur's lemma. This shows:

$$\langle Te_i, e_j \rangle = \delta_{ij}/d = |G|^{-1} \sum_G \langle ge_i, e_i \rangle \overline{\langle ge_j, e_j \rangle} = \langle \xi_i, \overline{\xi}_j \rangle$$

as desired. ∎

Corollary 12.19 *The number of irreducible characters is at most the number of conjugacy classes in G.*

Class functions from G-maps. We have seen that the character χ of any representation (ρ, V) is a positive integral sum of class functions. Where do the other class functions come from?

A natural source is to consider G-maps $T : V \to V$ where $V \cong \sum m_i V_i$. That is, $Tg = gT$. Suppose we set

$$\phi(g) = \operatorname{tr}(Tg).$$

Then we have

$$\phi(gh) = \operatorname{Tr}(Tgh) = \operatorname{Tr}(gTh) = \operatorname{Tr}(Thg) = \phi(hg),$$

so ϕ is a class function. Moreover, by Schur's lemma, one can verify that

$$\operatorname{tr}(Tg|m_i V_i) = \lambda_i \chi_i(g)$$

(this is immediate when $m_i = 1$). Consequently

$$\phi(g) = \sum \lambda_i \chi_i$$

(and conversely all sums of characters can be obtained by this construction).

Averaging and class functions. We can also go backwards in a certain sense: given any G-space V and class function ϕ, we can form

$$T = \sum_G \phi(g)g.$$

This is a G-map because:

$$
\begin{aligned}
T(hv) &= \sum_G \phi(g)ghv = \sum \phi(g)h(h^{-1}gh)v \\
&= \sum h\phi(hgh^{-1})(hgh^{-1})v = h(Tv)
\end{aligned}
$$

for all $h \in G$. Using this averaging procedure we find:

Theorem 12.20 *The irreducible characters span the class functions.*

Proof. Let $\phi(g)$ be a class function, let $V = L^2(G)$, and form T as above. Then $\psi(g) = \operatorname{tr}(Tg)$ is a linear combination of characters. But we have

$$
\psi(h) = \sum_G \phi(g)\operatorname{Tr}(gh) = |G|\phi(g^{-1}),
$$

by properties of the regular representation. Since any function can be expressed as $\phi(g^{-1})$, the sums of characters span the class functions. ∎

The Monster. The largest sporadic finite simple group is the *monster*, M, with $|M| \approx 10^{54}$. One of the few ways of grappling with this group is via its character table, which is given in [CCN].

13 Group presentations

In this section we present a topological perspective on group theory: the Cayley graph, free groups, and group presentations.

The Cayley graph. Given generators a_i for G, we draw a directed graph with a vertex for each $g \in G$ and an edge from g to $a_i g$, colored by a_i. If a_i has order two, the arrow is dropped.

Examples: $\langle \mathbb{Z}, 1 \rangle$; $\langle \mathbb{Z}/n, 1 \rangle$; $\langle V_4, a, b \rangle$; generators i, j; the star, i.e. $\mathbb{Z}/5$ with generator 2.

Examples: (S_3, f, r) vs. $(\mathbb{Z}/6, 2, 3)$. Two triangles running opposite directions in one case, the same direction in the other. Visualizing commutativity.

Example: (S_4, a, b) where $a^3 = b^2 = e$. The graph is a cube with the corners trimmed off.

Groups of order 8 as graphs. The dihedral group (D_4, f, r). The group $\mathbb{Z}/2 \times \mathbb{Z}/4$.

The quaternion group Q with generators i, j; 8 points on S^3! (Put $\pm 1, \pm i, \pm j$ on the vertices of an octahedron. Then put k in the center of the octahedron and $-k$ at infinity!)

Example: the Cayley graph of $\mathbb{Z}/a \times \mathbb{Z}/b$ can be visualized as a torus.

The free group on 2 generators. Any one generator G group admits a *surjective* map $\phi : \mathbb{Z} \to G$. Thus $\mathbb{Z}$ is the 'biggest' group with one generator — every other cyclic group is a quotient of it.

What is the biggest group you can make with two generators?

Answer: Consider all finite words in the letters a, a', b and b'. A word is *reduced* if it does not contain aa', $a'a$, bb' or $b'b$. An arbitrary word can be reduced by repeatedly cancelling out (removing) these 2-letter subwords whenever they occur. Example: $ab'aa'ba \to ab'ba \to aa$.

Let G be the set of all reduced words. Multiplication is defined by concatenation followed by reduction. The inverse of a word is obtained by writing it down backwards and exchanging the primed and unprimed variables.

The result is the *free group* $\mathbb{Z} * \mathbb{Z} = \langle a, b \rangle$.

Trees. The Cayley graph of the free group is an infinite tree.

Theorem 13.1 *Let G' be a group that can be generated by two elements x and y. Then there is a unique surjective map $\phi : \mathbb{Z} * \mathbb{Z} \to G'$ with $\phi(a) = x$ and $\phi(b) = y$.*

This means G' can be described as $G / \operatorname{Ker} \phi$.

Example. Let $H \subset G = \mathbb{Z} * \mathbb{Z} = \langle a, b \rangle$ be the subgroup generated by *all commutators*, $[g, h] = ghg^{-1}h^{-1}$. Since the set of commutators is closed under conjugation, so is H. Therefore H is *normal*.

What is G/H? Construct a map $\phi : G \to \mathbb{Z}^2$ sending a to $(1, 0)$ and b to $(0, 1)$. Since every commutator is in the kernel, we actually get a map $\phi : G/H \to \mathbb{Z}^2$. Now construct a map $\psi : \mathbb{Z}^2 \to G/H$, sending (i, j) to $a^i b^j$. Since G/H is abelian, this map is a homomorphism. Clearly the compositions are the identity, so G/H is isomorphic to $\mathbb{Z}^2$.

Group presentations. By $G = \langle g_1, \ldots, g_n : r_1, \ldots, r_m \rangle$ we denote F/N where F is the free group on generators $g_1, \ldots, g_n$, and $N \subset F$ is the smallest normal subgroup containing $R = \{r_1, \ldots, r_m\}$.

Another way to put it is that N is the set of *consequences* of the relations R. Here a consequence of R means either

1. the identity e (the trivial consequence),

2. an element $r \in R$,

3. the product of two consequences,

4. the inverse of a consequence, or

5. fxf^{-1} where $f \in F$ and x is a consequence.

Clearly if we have a group in which the elements of R represent the identity, so do the consequences of R. Now it is easy to see that the consequences form a normal subgroup, and in fact they are exactly N, the smallest normal subgroup containing R.

Theorem 13.2 *To give a homomorphism $\psi : G \to H$, where $G = \langle g_1, \ldots, g_n : r_1, \ldots, r_m \rangle$, it suffices to give values $h_i = \psi(g_i) \in H$ for each generator, and check that $\psi(r_i) = e$ for each relations.*

Proof. Let F be the free group on generators $\langle g_1, \ldots, g_n \rangle$. The h_i determine a homomorphism $\phi : F \to H$. By assumption, $r_i \in \mathrm{Ker}(\phi)$ for each relation. Since the kernel is a normal subgroup, it *contains* the smallest normal subgroup N containing the relations r_i. But by definition, $F/N \cong G$, so ϕ descends to a map $\psi : F/N \to H$. ∎

Presenting $\mathbb{Z}^2$. Let $G = \langle a, b : ab = ba \rangle$. Then G is abelian, and there is a map to $\mathbb{Z}^2$, which is obviously surjective. There is also an inverse map, which is a homomorphism because G is abelian. Both compositions give the identity, so $G \cong \mathbb{Z}^2$.

The checkerboard. It is useful to compare the Cayley graphs of the free group $\langle a, b \rangle$ and of the free abelian group $\langle a, b : ab = ba \rangle$. The relation gives loops in the second graph.

Dihedral groups. As an example, let's consider the group

$$G = \langle f, r : f^2 = r^n = e, rf = fr^{-1} \rangle.$$

Using the relations, every word can be represented in the form $r^i f^j$ where $0 \leq i < n$ and $j = 0, 1$. Thus G has *at most* $2n$ elements.

Define the obvious homomorphism $\phi : G \to D_n$. Then ϕ is onto, so by counting we see it is an isomorphism.

The infinite dihedral group. Let $D_\infty = \langle f, r : rf = fr^{-1} \rangle$. We can think of D_∞ as acting on $\mathbb{R}$ as the boundary of an 'infinite polygon', with vertices at the integers, by $f(x) = -x$, $r(x) = x + 1$. Then $frf(x) = -((-x) + 1) = x - 1 = r^{-1}(x)$ as required.

Another presentation for D_∞. Let $G = \langle a, b : a^2 = b^2 = e \rangle = \mathbb{Z}/2 * \mathbb{Z}/2$. It is easy to draw the Cayley graph of G; it's a straight line, just like the boundary of an infinite polygon.

Theorem 13.3 D_∞ *and G are isomorphic.*

Proof. Define a map $\phi : G \to D_\infty$ by $\phi(a) = f$, $\phi(b) = rf$. Then clearly $\phi(a^2) = e$ and $\phi(b^2) = rfrf = rr^{-1} = e$, so ϕ is a homomorphism.

Now define a map $\psi : D_\infty \to G$ by $\psi(f) = a$ and $\psi(r) = ba$. Then $\psi(f^2) = a^2 = e$ and

$$\psi(fr^{-1}) = a(ba)' = aa'b' = b' = b = (ba)a = \psi(rf),$$

so ψ is a homomorphism. We then compute $\psi \circ \phi(a) = a$,

$$\psi \circ \phi(b) = \psi(rf) = baa = b,$$

so $\psi \circ \phi$ is the identity. Similarly $\phi \circ \psi$ is the identity, so these two groups are isomorphic. ■

Generators and relations of S_n.

Theorem 13.4 S_n *has generators* $\tau_i = (i, i+1)$, $i = 1, \ldots n-1$, *with relations*

$\tau_i^2 = e;$
$\tau_i \tau_{i+1} \tau_i = \tau_{i+1} \tau_i \tau_{i+1};$ *and*
$\tau_i \tau_j = \tau_j \tau_i$ *if* $|i - j| > 1$.

Proof. To check the main relation, let $(i, i+1, i+2) = (i, j, k)$; then we have: $(ij)(jk)(ij) = (ik) = (jk)(ij)(jk)$. So there is a map of the group above to S_n, and since adjacent permutations generate, it is *onto*.

Now by the picture of changing crossings, it is clear that any two diagrams of the same permutation differ by these relations. ■

Cor. The parity of an element in S_n is well-defined.

Proof. The relations preserve parity. Alternatively, define a map from S_n to $\mathbb{Z}/2$ by sending each τ_i to one, and observe that the relations are satisfied.

∎

Trivial groups. It is not always easy to tell whether or not a presentation is just giving the trivial group. For example, $\langle a : a^{12} = e, a^{25} = e \rangle$ is trivial.

14 Knots and the fundamental group

What is a knot? It is a smooth closed curve in 3-space. A knots is not allowed to cross itself. A knot can be moved a little so it is a polygon. We do not allow wild knots.

Two knots K_0 and K_1 are *equivalent* if you can make a smoothly moving family of knots K_t that connects them. You can imagine this motion taking place in discrete steps, $K_0, \ldots, K_n$, where K_i and K_{i+1} differ by a triangle move.

A *link* is defined similarly as a finite number of disjoint closed loops.

Knot projections. A useful way to discuss knots is by projections: you put the knot almost in a plane, with pairs of strands meeting at crossings.

Any knot can be given a knot projection; in fact a generic projection will work. You just have to avoid the directions tangent to the knot, and the directions of lines passing through the knot in 3 points (taken with multiplicities). Each forms a one-dimensional set.

Examples.

1. The unknot. There are several projections. Any knot projection with 0, 1 or 2 crossings is the unknot. Any knot projection you can draw without anticipating crossings is the unknot.

2. The trefoil knot, 3_1.

3. The figure-eight knot, 4_1.

4. The (p, q)-torus knot. You start with q strands and form a braid of the form β^p, where β is a cyclic permutation; then close. If p and q are relatively prime, you get a knot. The $(1, 3)$ torus knot is the unknot; $(2, 3)$ is the trefoil.

5. The unlink on two components.

6. The Hopf link.

7. The Borromean rings, after the Renaissance family crest of the Borromeas.

History. Lord Kelvin conjectured that atoms are knots in ether. Tait and Little undertook the tabulation of knots. In recent times biologists have discovered that DNA is often knotted. The classification of 3-dimensional spaces is intimately tied up with knot theory.

Showing two knots are the same. Suppose K_1 and K_2 are projections that happen to correspond to the same knot. Then you can transform K_1 to K_2 be a sequence of *Reidemeister moves* (or their inverses). These moves are:

I Transform one strand into a loop with one crossing. The singularity is a cusp.

II Transform two parallel strands by adding two crossings. The singularity is a tangency.

III Transform three strands preserving the number of crossings. The singularity is a triple-point.

The Reidemeister moves can be remembered by the number of strands they involve. Planar isotopy is also allowed. The Reidemeister moves also work for links.

Proof. One way to approach the proof is to consider what kinds of singularities can arise as you view a generic knot projection during isotopy. The generic singularities are cusps, tangencies and triple points, accounting for the 3 moves.

Example: Draw a trefoil with one crossing wrong. This can be undone by the sequence III, I, II.

Example: Tangle up the trefoil.

Example: For each crossing of 6_3, change it and simplify the result.

Orientation, linking number and tricoloring. A knot or link is *oriented* if we have chosen a direction (usually indicated by an arrow) to traverse each component. A link with n components has 2^n possible orientations.

Showing two links are different. Let $L = K_1 \cup K_2$ be a two component oriented link. The *linking number* $\ell(K_1, K_2)$ is defined as follows: at each

crossing between K_1 and K_2, count $+1$ if it is a right-hand turn to get onto the overpass, otherwise -1. Add up and divide by two; this is $\ell(K_1, K_2)$.

Theorem 14.1 *The linking number is an* invariant *of L. Even though it is defined using a projection, the answer for two different projections is the same.*

Proof. Type I moves don't involve both components. A type two moves creates a pair of crossings of opposite sign. And type III doesn't really change any pair of strands, only the configuration of all three. ∎

Examples: the unlink, Hopf link and Whitehead link.

Unoriented links. The *absolute value* of the linking number is independent of orientation.

Tricoloring. We still haven't shown there are any real knots. To do this, let's say a *tricoloring* of a knot or link projection is an assignment of colors R, G, B to the arcs such that:

1) at least 2 colors are used; and at each crossing either
2a) all three strands have the same color or
2b) all three strands have different colors.

Theorem 14.2 *If one projection of a knot or link can be tricolored, then they all can.*

Proof. We must check the Reidemeister moves.

(I) The entire loop must be a single color.
(II) If the parallel strands are colored R and G, then we color the underloop B.
(III) If 3 colors appear, then a crossing where 3 colors are the same either appears or disappears.

In all 3 cases, if 2 colors were used before, then 2 are used after (see especially 2, which can increase the number of colors from 2 to 3).

Example. The unknot cannot be tricolored.

Example. The trefoil *can* be tricolored. Thus the trefoil is really knotted!

Example. The figure 8 knot cannot be tricolored. Thus tricoloring is not powerful enough to detect all knots.

Connect sum and colorings. Example. The connect sum of 2 trefoils can be tricolored in 11 different ways. (!) The point is that we can use a 1-coloring on one factor so long as it hooks up with a 3-coloring on the other factor.

The *number* of tricolorings is also an invariant of the knot or link.

Example. Let L be the unlink on 2 components. Then L can be tricolored with 2 colors — if drawn with no crossings — and with 3 colors — if drawn with 2 crossings.

Fundamental groups. *Examples of $\pi_1(X)$.* The disk, the annulus, the circle, the figure eight, the torus, the torus with a hole, a surface of genus two.

The knot group. This is the group of flight plans for airplanes leaving from a base located outside the knot and flying around it.

Examples: For the unknot, $G(K) = \mathbb{Z}$. For the unlink on two components, $G(L) = Z * Z$.

Presenting a knot or link group. There is one generator for each arc of the knot projection. To each arc we associate a generator of $\pi_1(\mathbb{R}^3 - K)$ that makes a right-handed underpass for that arc. (This means as one moves along the generator, one can make a safe right-hand turn to get on the superhighway.)

Now write ab for the loop that does a first, then b. Then when c crosses over $a - b$, so that we have a right-hand underpass, we get $ac = cb$. At a left-handed underpass, we get $ca = bc$.
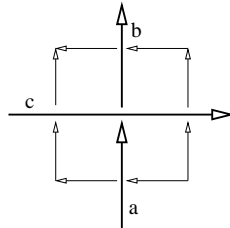

Figure 5. Proof that $ac = cb$ at safe underpass.

The trefoil knot group. We can now finally return to the trefoil knot. In the picture with all safe crossings, the group is

$$G(K) = \langle a, b, c \ : \ ab = bc, bc = ca, ca = ab \rangle.$$

Mapping $G(K)$ to $\mathbb{Z}$. This is easy: define $\phi(a) = \phi(b) = \phi(c) = 1$. It works for every knot group.

Mapping $G(K)$ to S_3. Notice that S_3 has 3 transpositions, call them A, B, C; they are all the odd permutations. They must satisfy $AB = CA$, since ABA is odd, it can't be A (else $A = B$) and it can't be B (else S_3 is commutative). Mapping (a, b, c) to (A, B, C) we see G admits S_3 as a quotient!

Corollary 14.3 *The trefoil knot really* is *knotted.*

Theorem 14.4 *The tricolorings of a knot correspond one-to-one to the surjective homomorphisms $\phi : G(K) \to S_3$.*

Proof. Let A, B, C be the three flips in S_3 as before. Since the generators of $G(K)$ correspond to arcs, we can use the three colors to define a map $\phi : G(K) \to S_3$ on the generators. Then the tricoloring condition shows that each relation in $G(K)$ is satisfied. So we can map the generators for strands of color A to flip a, etc. Since at least two colors are used, we get a surjective map.

Similarly, if we have a surjection, then the generators (all being conjugate) must go to flips, else the image would be in the abelian subgroup of rotations. We then obtain a coloring. ∎

Note: for a link the statement above is not quite true. A homomorphism that kills one component of a link does not correspond to a coloring. That is the tricolorings correspond to maps to S_3 that send all generators of the Wirtinger presentation to flips.

Changing presentations. To prove $G(K)$ is a knot invariant, not just an invariant of the knot projection, it is important to understand elementary (or Tietze) moves on a presentation. There are just two:

(1) $\langle g_i : r_i \rangle \iff \langle g_i : r_i, s \rangle$, where s is a consequence of the given relations. That means s is a product of conjugates of the r_i.

(2) $\langle g_i : r_i \rangle \iff \langle g_i, h : r_i, h = w(g_1, \ldots, g_n) \rangle$, where $w(\cdot)$ is a word in the generators g_i. This means we can add a new generator so long as we add a relation putting it in the group generated by the (g_i).

Example. The trefoil group can be simplified to 2 generators, one relation. *Invariance of $G(K)$.*

Theorem 14.5 *Equivalent projections of a knot (or link) give isomorphic groups $G(K)$.*

Proof. We must check the Reidemeister moves. (I) A loop gives $\langle a, b : aa = ba \rangle$, so we can use Tietze move (2) to eliminate b.

(II) Suppose the arc a is underpassed by (b, c, d). Then we get from the 2 crossings the relations $ba = ac, ac = da$. From this we derive $b = d$ (Tietze 1), then eliminate c, d (Tietze 2). We are left with a, b and no relations, which is the contribution of two parallel arcs.

(III) Let the topmost arc be c, NW to SE, over (a, d), SW to NE, with finally (b, f, e) passing W to E under a and c. The 3 crossings give the relations

$$R = \langle ac = cd, ab = fa, fc = ce \rangle.$$

See Figure 6.

We can obtain the relation $aba' = cec' = f$, and then eliminate f, giving the new set of relations

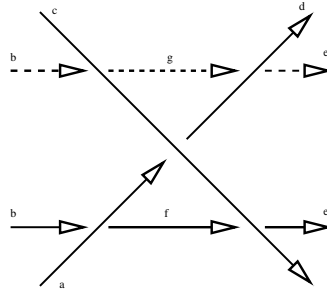$$R' = \langle ac = cd, aba' = cec' \rangle$$

Figure 6. Reidemeister move III.

After the Reidemeister move we get a new arc g and relations

$$S = \langle ac = cd, bc = cg, dg = ed \rangle.$$

We can obtain the relation $c'bc = d'ed = g$, and eliminate g, obtaining

$$S' = \langle ac = cd, c'bc = d'ed \rangle.$$

65

We now must show the two sets of relations R' and S' are equivalent. To see this, we will use the relation $ac = cd$, present in each group. Then the relation $c'bc = d'ed$ in S', implies the relation $(ac)c'bc(ac)' = (cd)d'ed(cd)'$, which simplifies to the relation $aba' = cec'$ in R'. The process can be reversed, so indeed the groups are isomorphic. ∎

Search for cycles. If we want to find a quotient $\phi : G(K) \to F$, it is *very useful* to note that the standard generators of a knot group are all conjugates. Thus the images of these generators must also be conjugate, and generate F (if ϕ is to be surjective.)

With this in mind, we can see that a natural choice for the trefoil group is S_3, the symmetries of a triangle, since it has three edges. We map each generator to the flip of the edge, and this gives ϕ.

The figure 8 group. We have the presentation

$$G = \langle a, b, c, d \ : \ da = bd, ba = ac, bc = db, dc = ca \rangle.$$

Now the elements a, b, c, d are *all conjugate* in G. So if we try to find a map $\phi : G \to F$, where F is a finite group, then the images of a, b, c, d all have the same order — and indeed they are all conjugate. So we need a group with 4 conjugate elements, that generate the group.

A nice candidate is $F = A_4$, the symmetries of a tetrahedron T. The motivation is that T has 4 faces, corresponding to the generators of G, just like the edges of the triangle did for the trefoil knot.

The clockwise rotations of a face are the permutations $(123), (134), (243)$ and (142). Any two faces stand in the same relation, so we may as well send a to (123) and d to (134).

Then the first relations gives $b = dad' = (432)$, and the second gives $c = a'ba = (421)$. We can then check that the last two relations are satisfied.

Geometric check. Another way to think about the relations in the figure 8 group is to let a, b, c, d correspond to faces A, B, C, D of the tetrahedron. Then an equation like $b = dad'$ is equivalent to $B = d(A)$. So we get the relations $B = d(A), C = a'(B), D = b(C)$ and $D = c(A)$. Once A and D are chosen, the first two relations specify C and B, and then we can check the last two.

Trefoil $\neq$ Figure Eight.

Theorem 14.6 *There is no surjection from the trefoil group to A_4. Thus the unknot, the trefoil and the figure eight knot are all distinct.*

66

Proof. Let $\phi : G \to A_4$ be a surjective homomorphism from the trefoil group to A_4. Since the generators are all conjugate, they all have the same order, which must be 1, 2 or 3. But the elements of order 2 in A_4 generate a proper V_4 subgroup, so the order must be 3. Moreover two generators must map to different elements, else the image would be $\mathbb{Z}/3$.

Up to symmetry we can then take the images of the generators to be $a = (123)$ and $b = (134)$. Then the relations $ab = bc = ca$ imply $c = b'ab = (432)$, but we then find $bc = (321) \neq ca = (431)$. ∎

Hopf link. The Hopf link L is *less knotted* than the unlink, since $G(L) \cong \mathbb{Z} \oplus \mathbb{Z}$. As a trick, one can weave the commutator through two unlinked carabiner, in such a way that the loop comes free when the carabiner are linked! (Cf. homework on computing $G(L)$.)

Theorem 14.7 *The linking number for $L = K_1 \cup K_2$ corresponds to the abelianization of the element of $\pi_1(\mathbb{R}^3 - K_1)$ represented by K_2.*

Proof. The proof is a little tricky. Working with a link projection, one can first change crossings of K_2 with itself so K_2 becomes the unknot. This change does not change our projection-based calculation of the linking number (obviously), nor does it change the image of K_2 in $\pi_1(\mathbb{R} - K_1)$ (obviously).

Now re-arrange the projection so K_2 is a counter-clockwise round circle. It is then clear that the number of outward-heading crossings of K_1 with K_2 is the same as the number of inward-heading crossings.

Count the crossings in 4 groups, TO, BO, TI, BI, where T/B means K_1 is on top/bottom, and O/I means it is headed out/in. Letting P be the link number in π_1, and L the linking number from the diagram, we have

$$
\begin{aligned}
TO + BO &= TI + BI \\
L &= (TO + BI - TI - BO)/2 \quad \text{and} \\
P &= TO - TI.
\end{aligned}
$$

Using the first equation we have $P = BI - BO$; averaging these two expressions for P, we obtain $P = L$. ∎

Theorem 14.8 (Gordon-Luecke) *Let $(G(K), H(K))$ be a knot group and a subgroup $H(K) \cong \mathbb{Z}$ generated by a meridian. Then K is equivalent to K', or its mirror image, iff there is an isomorphism $G(K) \cong G(K')$ sending $H(K)$ to $H(K')$.*

The $H(K)$ is needed for the square and granny knots, which have isomorphic groups. Often the $H(K)$ is unnecessary.

References

[Ar]   M. Artin. *Algebra.* Prentice-Hall, 1991.

[Ax]   S. Axler. *Linear Algebra Done Right.* Springer, 1997.

[CCN] J. H. Conway, R. T. Curtis, S. P. Norton, R. A. Parker, and R. A. Wilson. *Atlas of Finite Groups.* Oxford University Press, 1985.

[Hal]   P. Halmos. *Finite-Dimensional Vector Spaces.* Springer-Verlag, 1987.