Lectures on Representation Theory

Hjalmar Rosengren
Department of Mathematical Sciences
Chalmers University of Technology and University of Gothenburg

DRAFT October 17, 2019

Contents

Preface **4**

1 Review of group theory **5**
1.1 Motivation and definition 5
1.2 Subgroups and homomorphisms 7
1.3 Generating sets and cyclic groups 9
1.4 Cosets and quotients 9
1.5 Actions and representations 11
1.6 The symmetric group 13
1.7 Additional exercises 16

2 Review of ring theory **19**
2.1 Definition 19
2.2 Ideals and homomorphisms 21
2.3 Unique factorization 24
2.4 Additional exercises 27

3 Modules **29**
3.1 Definition 29
3.2 Submodules, homomorphisms and quotients 30
3.3 Generators and cyclic modules 32
3.4 Direct sums 33
3.5 Free modules 34
3.6 Associative algebras 38
3.7 Additional exercises 39

4 Tensor products **41**
4.1 Tensor products: three definitions 41
4.2 Properties of tensor products 44
4.3 Symmetric and antisymmetric tensors 46
4.4 Additional exercises 51

5 Modules over principal ideal domains **53**
5.1 Statement and overview 53
5.2 Proof of Theorem 5.1.1 54

5.3 The invariant factor decomposition 60
5.4 Canonical forms 61
5.5 Systems of differential equations 66
5.6 Additional exercises 68

6 Group representations **70**
6.1 Fundamental facts 70
6.2 Characters 75
6.3 The regular representation 78
6.4 The character table 80
6.5 Examples 82
6.6 Interpreting the character table 85
6.7 Group algebra and Fourier transform 89
6.8 Peter–Weyl Theorem 92
6.9 Frobenius divisibility 94
6.10 Additional exercises 96

7 Representations of the symmetric group **99**
7.1 Statement of results 99
7.2 Irreducible representations 103
7.3 Explicit realizations of representations 108
7.4 Schur–Weyl duality 113
7.5 Two determinant evaluations 116
7.6 Characters 118
7.7 Dimensions 124
7.8 Additional exercises 126

8 Representations of SU(2) **127**
8.1 Compact groups 127
8.2 Representations 130
8.3 Matrix elements and Krawtchouk polynomials 133
8.4 Jacobi polynomials 136

Preface

I would like to thank Vilhelm Agdur, Linnea Hietala, Erik Håkansson, Simon Jacobsson, Jules Lamers, Per Ljung, Carl Lundholm, Stepan Maximov and Michel Zoeteman for their valuable comments, leading to substantial improvement of these notes.

Chapter 1

Review of group theory

1.1 Motivation and definition

A group is an abstract version of the set of symmetries of an object. As an example, consider the plane motions preserving an equilateral triangle. There are six such symmetries: rotation 120° clockwise or anti-clockwise, reflection in one of the three altitudes and the identity symmetry that does not move any points at all. Note that the composition of two symmetries is again a symmetry. For instance, drawing the triangle as in Figure 1.1 and letting r denote rotation 120° clockwise and s reflection in the vertical altitude, the composition $s \circ r$ is reflection in the altitude through the lower right corner.

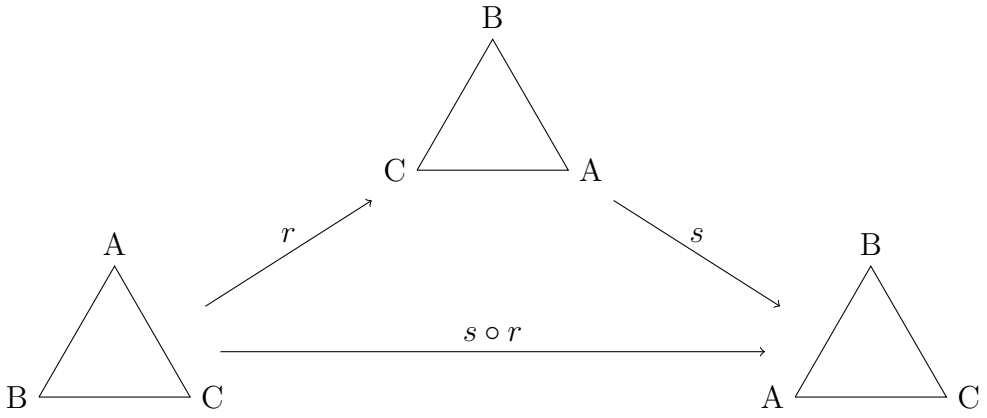

Figure 1.1: Composition of symmetries

Obviously, composition of symmetries satisfies the associative law $(a \circ b) \circ c = a \circ (b \circ c)$. Moreover, each symmetry is invertible and the inverse map is again a symmetry. That is, for each symmetry a there is a symmetry a^{-1} such that $a \circ a^{-1} = a^{-1} \circ a = \mathrm{id}$. The formal definition of a group is based on these properties of symmetries.

Definition 1.1.1. *A* group G *is a set equipped with a map* $G \times G \to G$ *(the*

multiplication*) that we denote* $(a, b) \mapsto a \cdot b = ab$*. The multiplication should satisfy the associative law*

$$a \cdot (b \cdot c) = (a \cdot b) \cdot c, \qquad a,\, b,\, c \in G. \tag{1.1a}$$

Moreover, there should be an identity element $1 \in G$ *such that*

$$a = 1 \cdot a = a \cdot 1, \qquad a \in G. \tag{1.1b}$$

Finally, each element should be invertible, that is, each $a \in G$ *should have an* inverse $a^{-1} \in G$ *such that*

$$a \cdot a^{-1} = a^{-1} \cdot a = 1. \tag{1.1c}$$

It is easy to see that the identity element and the inverse of each element are unique.

Note that we do not require the commutative law $a \cdot b = b \cdot a$. Indeed, it is easy to check that, in the example of Figure 1.1, $r \circ s \neq s \circ r$. However, there are many important examples of groups where the commutative law holds. A group with that property is called *abelian* (or *commutative*). In abelian groups, it is customary to use additive rather than multiplicative notation. That is, we denote multiplication by $(a, b) \mapsto a + b$ (and call it addition), inversion by $a \mapsto -a$ and the identity element by 0. Using the commutative law in the form

$$a + b = b + a,$$

the group axioms (1.1) simplify to

$$a + (b + c) = (a + b) + c, \qquad a = a + 0, \qquad a + (-a) = 0.$$

In an abelian group we may introduce subtraction as $a - b = a + (-b)$. By contrast, in non-abelian groups we never write $\frac{a}{b}$ as it is not clear if this would mean ab^{-1} or $b^{-1}a$.

Some examples of groups are:

- The set of all invertible maps from a set X to itself is a group under composition. We will denote it S_X.
- If X is a finite set, an invertible map from X to itself is called a *permutation*. We often take $X = \{1, \ldots, n\}$ and write S_n instead of S_X. The group S_n is called the *symmetric group* on n elements.
- If V is a vector space[1], the set of invertible linear maps from V to V is a group, again under composition. It is called the *general linear group* and is denoted $\mathrm{GL}(V)$.

[1]We will review the definition in §3.1; you may think of $\mathbb{R}^n$ or $\mathbb{C}^n$.

- The set of invertible complex $n \times n$-matrices is a group under matrix multiplication. It can be identified with $\mathrm{GL}(\mathbb{C}^n)$; another common notation is $\mathrm{GL}(n, \mathbb{C})$.
- Each of the number systems $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ is an abelian group under addition.
- None of the number systems $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ is a group under multiplication. However, $\mathbb{Q}^* = \mathbb{Q} \setminus \{0\}$ is a group and so is $\mathbb{R}^*$ and $\mathbb{C}^*$, but not $\mathbb{Z}^*$. Likewise, $\mathbb{Q}_{>0}$ and $\mathbb{R}_{>0}$ are groups under multiplication. All these groups are abelian.
- Any vector space is an abelian group under addition of vectors.
- For a fixed positive integer n, the numbers $0, 1, \ldots, n-1$ form an abelian group under addition modulo n. It is known as the cyclic group of order n and will be denoted $\mathbb{Z}_n$.

Thanks to the associative law, one can define positive powers a^k of an element unambiguously ($a^2 = a \cdot a$, $a^3 = a \cdot (a \cdot a) = (a \cdot a) \cdot a$, $a^4 = a \cdot (a \cdot (a \cdot a)) = \cdots$). One also defines $a^0 = 1$ and $a^k = (a^{-1})^{-k}$ for negative k. It is then easy to prove that $a^{k+l} = a^k a^l$ for k, $l \in \mathbb{Z}$. Of course, in an abelian group we write $ka = a + \cdots + a$ rather than $a^k = a \cdots a$.

Exercise 1.1.1. Write down the 6×6 multiplication table for the symmetry group of the triangle.

Exercise 1.1.2. Show that if $ab = 1$ in a group then $ba = 1$. (That is, any right inverse is also a left inverse.)

Exercise 1.1.3. Let $\mathbb{Z}_n^*$ be the subset of $\mathbb{Z}_n$ consisting of integers relatively prime to n. Show that $\mathbb{Z}_n^*$ is a group under multiplication modulo n.

1.2 Subgroups and homomorphisms

Many groups in the list of examples above are sitting inside each other; for instance, $\mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$. Another example is the symmetry group of the equilateral triangle, which sits inside the much larger group of all invertible maps from the triangle to itself.

Definition 1.2.1. *A* subgroup[2] *of a group G is a non-empty subset $H \subseteq G$ such that if $a, b \in H$ then $a \cdot b \in H$ and $a^{-1} \in H$.*

[2]Swedish: undergrupp, delgrupp. A unicorn dies each time you say subgrupp.

Equivalently, a subgroup is a subset $H \subseteq G$ which is a group with the same multiplication as G (or, to be hyper-correct, the restriction of the multiplication on $G \times G$ to $H \times H$).

We also need a notion of "sameness" for groups. For instance, if we know how a symmetry (still in the sense of planar motions) of the equilateral triangle acts on the corners, we know how it acts on the whole triangle. Thus, we can identify the symmetries with permutations of the three corners. Since all six permutations appear in this way, the symmetry group of the equilateral triangle can be identified with S_3. The technical term for being the same algebraic structure is *isomorphic*.

Definition 1.2.2. *We say that two groups G and H are* isomorphic *if there exists an invertible map $\phi : G \to H$ (called an* isomorphism*) such that*

$$\phi(a \cdot b) = \phi(a) \cdot \phi(b), \qquad a, b \in G. \tag{1.2}$$

Note that the multiplication dot on the left of (1.2) refers to the multiplication in G and the one on the right to the multiplication in H. We will use the notation $G \simeq H$ for isomorphic groups.

Relaxing the condition that ϕ is invertible gives another important type of map between groups.

Definition 1.2.3. *If G and H are groups, a map from G to H that satisfies* (1.2) *is called a* homomorphism.

An *endomorphism* is a homomorphism from a group G to itself. A bijective endomorphism is called an *automorphism*.

Exercise 1.2.1. Show that if $\phi : G \to H$ is a homomorphism, then $\phi(1_G) = 1_H$ and $\phi(a^{-1}) = \phi(a)^{-1}$ for all $a \in G$.

Exercise 1.2.2. Show that $\mathbb{Z}_n$ is isomorphic to a subgroup of $\mathbb{C}^*$.

Exercise 1.2.3. Show that the group $\mathbb{R}_{>0}$ (under multiplication) is isomorphic to the group $\mathbb{R}$ (under addition).

Exercise 1.2.4. Is $\mathbb{Q}_{>0}$ isomorphic to $\mathbb{Q}$?

Exercise 1.2.5. Show that any subgroup of $\mathbb{Z}$ is of the form $n\mathbb{Z} = \{nx;\, x \in \mathbb{Z}\}$, where n is a non-negative integer.

Exercise 1.2.6. Let $H \subseteq G$ be a subset such that $a, b \in H \Rightarrow ab \in H$. Show that if G is finite then H is a subgroup of G. Show that this need not be the case if G is infinite.

1.3 Generating sets and cyclic groups

It is easy to see that the intersection of any family of subgroups is itself a subgroup. Thus, if $S \subseteq G$ is an arbitrary subset, we can take the the intersection of all subgroups of G containing S, obtaining in this way the smallest such subgroup. It is called the group *generated by* S.

A group generated by a single element is called a *cyclic group*.

Proposition 1.3.1. *Any cyclic group is either isomorphic to* $\mathbb{Z}$ *or to* $\mathbb{Z}_n$ *for some positive integer* n.

Proof. Suppose G is generated by $a \in G$. With a^k as defined in §1.1, let $H = \{a^k;\, k \in \mathbb{Z}\} \subseteq G$. Since $a^k a^l = a^{k+l}$ and $(a^k)^{-1} = a^{-k}$, H is a subgroup containing a. Since G is generated by a it follows that $H = G$. There are now two possibilities. If the powers a^k are all distinct, then $k \mapsto a^k$ defines an isomorphism from $\mathbb{Z}$ to G. Else, $a^k = a^l$ for some $k < l$. With $n = l - k$ we then have $a^n = 1$. Let n be the smallest positive integer with this property. Then, $G = \{1, a, a^2, \ldots, a^{n-1}\}$ and $k \mapsto a^k$ is an isomorphism from $\mathbb{Z}_n$ to G. □

The *order* of a finite group is the number of elements in the group. The *order* of a group element a is the smallest positive integer n such that $a^n = 1$. By the proof of Proposition 1.3.1, any element in a finite group has finite order, which is equal to the order of the subgroup generated by a.

Exercise 1.3.1. Show that the group generated by $S \subseteq G$ consists of all finite words $x_1 \cdots x_n$, where $x_i \in S$ or $x_i^{-1} \in S$. If G is abelian, show that this simplifies to $k_1 x_1 + \cdots + k_n x_n$, where $k_j \in \mathbb{Z}$ and $x_j \in S$.[3]

Exercise 1.3.2. Show that the group S_3 is generated by two elements. Describe all pairs of elements that work.

1.4 Cosets and quotients

Definition 1.4.1. *If* $H \subseteq G$ *is a subgroup and* $a \in G$, *then* $aH = \{ah;\, h \in H\}$ *is called a (left) coset*[4] *of* H. *The set of all cosets of* H *is denoted* G/H.

As an example, if $n\mathbb{Z}$ is considered as a subgroup of $\mathbb{Z}$, the cosets are the residue classes $\{x \in \mathbb{Z};\, x \equiv k \mod n\}$. As another example, if the x-axis $\{(x, 0);\, x \in \mathbb{R}\}$ is considered as a subgroup of $\mathbb{R}^2$, the cosets are the horizontal lines.

[3] Note the similarity to the subspace spanned by a set S in linear algebra. As we will see in §3, abelian group theory can be viewed as linear algebra over $\mathbb{Z}$.

[4] Swedish: sidoklass.

It is easy to see that the cosets form a partition of G, that is, each element of G belongs to exactly one coset. Moreover, the map $h \mapsto ah$ is a bijection $H \to aH$, so all cosets of H have the same cardinality. In particular, if G is finite then the number of cosets $|G/H|$ is equal to $|G|/|H|$. (In general, we use $|S|$ to denote the number of elements in a finite set S.) The following simple consequence is fundamental for understanding the structure of finite groups.

Proposition 1.4.2 (Lagrange's Theorem)**.** *If H is a subgroup of a finite group G, then $|H|$ divides $|G|$.*

To exemplify the usefulness of this result we state two immediate consequences.

Corollary 1.4.3. *If G is a group of order n, then the order of each element in G divides n. In particular, $a^n = 1$ for all $a \in G$.*

This follows by applying Lagrange's theorem to the subgroup generated by a.

Corollary 1.4.4. *If G is a group with p elements for a prime p, then $G \simeq \mathbb{Z}_p$.*

Indeed, by Corollary 1.4.3, all elements have either order 1 or p. Since only the unit element has order 1, any other element generates a subgroup isomorphic to $\mathbb{Z}_p$, which must be the whole group.

In some situations, the cosets of a subgroup form a group in a natural way. For instance, adding residue classes modulo n gives the cyclic group $\mathbb{Z}_n$. In general, we would like to multiply cosets by the rule

$$aH \cdot bH = abH, \tag{1.3}$$

but this may depend on the choice of a. If we replace a by ah, with $h \in H$, the left-hand side of (1.3) does not change, so we must have $ahbH = abH$. Multiplying with $b^{-1}a^{-1}$ from the left gives $b^{-1}hbH = H$, which is equivalent to $b^{-1}hb \in H$. This motivates the following definition.

Definition 1.4.5. *A subgroup $H \subseteq G$ is called* normal *if $a^{-1}ha \in H$ for all $a \in G$ and $h \in H$.*

If G is abelian, then $a^{-1}ha = h$ always, so any subgroup of an abelian subgroup is normal. On the other hand, it is easy to check that the permutations of $\{1, 2, 3\}$ that fix the element 3 form a non-normal subgroup of S_3.

If H is a normal subgroup of G, we write $H \triangleleft G$. It is then easy to see that (1.3) defines a group structure on G/H, a so called *quotient group*. In particular, the unit element is $1_{G/H} = H$. As we have already indicated, an example of a quotient group is $\mathbb{Z}/n\mathbb{Z} \simeq \mathbb{Z}_n$. To state the main fact about quotient groups we need the following definition.

Definition 1.4.6. *The* kernel[5] *of a group homomorphism $\phi : G \to H$ is the inverse image of the identity, that is,* $\mathrm{Ker}(\phi) = \{a \in G;\ \phi(a) = 1\}$.

The following important result is often called the first isomorphism theorem. We will leave the proof to the reader. The second and third isomorphism theorem are given as Exercises 1.7.9–1.7.10.

Theorem 1.4.7. *If $\phi : G \to H$ is a homomorphism then* $\mathrm{Ker}(\phi)$ *is a normal subgroup of G and $G/\mathrm{Ker}(\phi) \simeq \mathrm{Im}(\phi)$. Conversely, if $N \lhd G$, then $\phi(a) = aN$ defines a homomorphism $\phi : G \to G/N$ with kernel N and image G/N.*

In particular, the normal subgroups of G are exactly the kernels of all homomorphisms from G to other groups. Likewise, the quotient groups of G are exactly the images of homomorphisms from G to other groups.

One way to think of Theorem 1.4.7 is in terms of information: $\mathrm{Ker}(\phi)$ describes the information about G that is forgotten by ϕ and $\mathrm{Im}(\phi)$ describes the information that is retained.

Exercise 1.4.1. Show that $\mathbb{R}/\mathbb{Z}$ is isomorphic to a subgroup of $\mathbb{C}^*$.

Exercise 1.4.2. Let G be the symmetry group of the square, consisting of four rotations (including the identity) and four reflections. Let a and b denote the reflections in the two diagonals. Now let $K = \{1, a\}$ and $H = \{1, a, b, ab\}$. Show that $K \lhd H$ and $H \lhd G$ but K is not normal in G.

1.5 Actions and representations

We started our lecture by saying that a group is an abstraction of the set of symmetries of an object. It is natural to ask how much more general these abstract groups are compared to concrete symmetry groups. The answer is that they are not a bit more general; if we define a *symmetry group* to be a group of invertible maps on some set X, then any group is isomorphic to a symmetry group. Indeed, if G is a group we can define $\phi : G \to \mathrm{S}_G$ by $\phi(a)(b) = ab$. It is easy to see that ϕ is an injective homomorphism, hence G is isomorphic to the subgroup $\mathrm{Im}(\phi) \subseteq \mathrm{S}_G$. In particular, for finite groups we have *Cayley's theorem*, saying that any group with n elements is isomorphic to a subgroup of S_n.

More generally, a homomorphism $\phi : G \to \mathrm{S}_X$ is called a *group action*[6]. The set X is then called a *G-set*. If ϕ is injective, so that G is isomorphic to a subgroup of S_X, it is called a *faithful* action. An example is the action of S_3 on the equilateral triangle. We will usually write $g(x)$ rather than $(\phi(g))(x)$ for the action of $g \in G$ on

[5]Swedish: kärna.

[6]Swedish: gruppverkan.

$x \in X$. An important type of group actions appear from quotients by subgroups. Namely, if H is a subgroup of G, not necessarily normal, then G acts on the left cosets G/H by $g_1(g_2H) = g_1g_2H$.

When X is a G-set, the *orbit* of $x \in X$ is the set $G(x) = \{g(x);\, g \in G\}$. It is easy to see that X splits as the disjoint union of orbits. Indeed, if $G(x) \cap G(y) \neq \emptyset$, then $g_1(x) = g_2(y)$ for some $g_1,\, g_2 \in G$. This implies $g(x) = (gg_1^{-1}g_2)(y)$ for all g, hence $G(x) \subseteq G(y)$ and by symmetry $G(x) = G(y)$. The structure of the orbit $G(x)$ is determined by the *stabilizer* $G_x = \{g \in G;\, g(x) = x\}$. Indeed, there is a bijection $G(x) \to G/G_x$ defined by $g(x) \mapsto gG_x$, which preserves the group action. More formally, two group actions (or G-sets) $\phi_X : G \to \mathrm{S}_X$ and $\phi_Y : G \to \mathrm{S}_Y$ are called *equivalent* if there is a bijection $\psi : X \to Y$ with $\phi_Y(g) = \psi \circ \phi_X(g) \circ \psi^{-1}$ for all $g \in G$. Then, $G(x) \simeq G/G_x$ as G-sets. We may summarize the result of this discussion as follows.

Proposition 1.5.1. *Any G-set is equivalent to a disjoint union*

$$X \simeq \bigcup_x G/G_x,$$

where x runs over a set of representatives for the orbits.

A particular case of group actions are *group representations*. These are homomorphisms $G \to \mathrm{GL}(V)$, where V is a vector space. (Since $\mathrm{GL}(V)$ is a subgroup of S_V, a representation is indeed an action.) Group representations are a powerful tool to investigate symmetries, both from an abstract and an applied viewpoint. For instance, in quantum mechanics a physical system is described by a Hilbert space. Symmetries of the system result in a representation of the corresponding group on the Hilbert space, which may lead to non-trivial physical information.

We will have much to say about group representations in §6–8, but let us note already that there is a version of "Cayley's theorem" for representations. Namely, let G be a group and V the complex vector space of all functions $G \to \mathbb{C}$. Then we can define $\phi : G \to \mathrm{GL}(V)$ by $(\phi(a)(f))(b) = f(a^{-1}b)$, where $a,\, b \in G$ and $f \in V$. It is easy to check that ϕ is an injective homomorphism (called the *regular representation*), so we have realized G as a group of invertible linear maps on a vector space. In particular, if G is a finite group with n elements, then $\dim(V) = n$. We conclude that any group with n elements can be realized as a group of complex $n \times n$-matrices.

Exercise 1.5.1. Show that $\phi(a)(b) = aba^{-1}$ defines an action of a group G on itself. Show that this action is faithful for $G = \mathrm{S}_3$, but not for groups in general.

Exercise 1.5.2. A *permutation matrix* is a matrix with exactly one 1 in each row and column, all other elements being 0. Show that S_n is isomorphic to the group

of all $n \times n$ permutation matrices. Conclude that any group with n elements is isomorphic a group of $n \times n$ permutation matrices.

Exercise 1.5.3. Two subgroups H_1, $H_2 \subseteq G$ are called *conjugate* if $H_1 = gH_2g^{-1}$ for some $g \in G$. Show that $G/H_1 \simeq G/H_2$ as G-sets if and only if H_1 and H_2 are conjugate.

Exercise 1.5.4. Let X be a G-set and V the vector space of all functions $X \to \mathbb{C}$. Define a natural representation of G on V (this is called a *permutation representation*).

1.6 The symmetric group

We will be particularly interested in representation theory of the symmetric group, and collect some facts that that will be needed in §7.

First of all, we introduce the *cycle notation*. It should be sufficient to explain this for an example. Let $\sigma \in \mathrm{S}_6$ be defined by the table

x	1	2	3	4	5	6
$\sigma(x)$	3	2	6	5	4	1

.

In cycle notation, we write

$$\sigma = (1\ 3\ 6)(4\ 5), \tag{1.4}$$

where the first factor stands for the 3-cycle $1 \mapsto 3$, $3 \mapsto 6$, $6 \mapsto 1$ and the second factor for the 2-cycle $4 \mapsto 5$, $5 \mapsto 4$. The 1-cycle or fix-point $2 \mapsto 2$ could have been denoted (2) but we usually omit 1-cycles from the notation. Clearly, the cycle notation is unique up to permuting the cycles and cyclically permuting the elements in each cycle. For instance,

$$(4\ 5)(1\ 3\ 6), \quad (3\ 6\ 1)(4\ 5), \quad (5\ 4)(6\ 1\ 3)$$

all represent the same permutation as (1.4). In particular, we can define the *cycle structure* of a permutation to be $(c_1, \ldots, c_n)$, where c_k is the number of k-cycles appearing in the cycle notation.

It is easy to multiply permutations in the cycle notation. For instance, let σ be as above and $\tau = (1\ 5\ 2\ 6)(3\ 4)$. To compute $(\sigma\tau)(1)$, we only need to look at the underlined symbols in

$$\sigma\tau = (1\ 3\ 6)(\underline{4}\ \underline{5})(\underline{1}\ \underline{5}\ 2\ 6)(3\ 4). \tag{1.5}$$

Starting from the right, we see that 1 is mapped to 5, which is then mapped to 4, so $(\sigma\tau)(1) = 4$. To get the cycle notation for $\sigma\tau$, we successively compute in the same way $(\sigma\tau)(4) = 6$ and so on. In this case we obtain a single cycle, namely,

$$\sigma\tau = (1\ 4\ 6\ 3\ 5\ 2\ 1). \tag{1.6}$$

To avoid a possible source of misunderstanding, we stress that a permutation can be written as a product of cycles in many ways. When we speak about the cycle notation or cycle structure, we always refer to a factorization into *disjoint* cycles. For instance, the right-hand sides of (1.5) and (1.6) represent the same permutation, but only (1.6) reveals the cycle structure.

Next, we consider conjugation of permutations, that is, the map $\sigma \mapsto \tau\sigma\tau^{-1}$ for fixed τ. If $\sigma(i) = j$, then $(\tau\sigma\tau^{-1})(\tau(i)) = \tau(j)$. In cycle notation, this means that any pair of symbols $(\cdots i\ j \cdots)$ is replaced by $(\cdots \tau(i)\ \tau(j) \cdots)$, so we simply need to apply τ to every symbol. For instance, if σ is as in (1.4), then

$$\tau\sigma\tau^{-1} = (\tau(1)\ \tau(3)\ \tau(6))(\tau(4)\ \tau(5)), \qquad \tau \in \mathrm{S}_6. \tag{1.7}$$

We can think of this as expressing the same permutation as σ after the "change of coordinates" $i \mapsto \tau(i)$. It follows that the permutations conjugate to σ are precisely those that have the same cycle structure as σ.

In general, when a group acts on itself by conjugation, the orbits $\{\tau\sigma\tau^{-1};\ \tau \in G\}$ are known as *conjugacy classes*. As we just explained, the conjugacy classes in S_n consist of permutations with the same cycle structure. Since the sum of the total length of the cycles (counting 1-cycles) is n, conjugacy classes in S_n can be labelled by *partitions* of n, where we define a partition of n to be an expression for n as a decreasing sum of positive integers. If the number of k-cycles is c_k, then the corresponding partition has c_k parts equal to k.

As an example, the seven partitions of 5 are

$$5, \quad 4+1, \quad 3+2, \quad 3+1+1, \quad 2+2+1, \quad 2+1+1+1, \quad 1+1+1+1+1.$$

A typical representative of the corresponding conjugacy class is, respectively,

$$(1\ 2\ 3\ 4\ 5), \quad (1\ 2\ 3\ 4), \quad (1\ 2\ 3)(4\ 5), \quad (1\ 2\ 3), \quad (1\ 2)(3\ 4), \quad (1\ 2), \quad \mathrm{id}\,.$$

We encourage the reader to check that the number of elements in each class is

$$24, \quad 30, \quad 20, \quad 20, \quad 15, \quad 10, \quad 1.$$

In general, we have the following result that we will need in §7.6.

Lemma 1.6.1. *The conjugacy class with cycle structure* $(c_1, \ldots, c_n)$ *consists of*

$$\frac{n!}{\prod_{k=1}^{n} k^{c_k} c_k!}$$

elements.

To see this, start with a representative of the conjugacy class written in the cycle notation, with 1-cycles included. For instance, if $n = 5$ and $(c_1, \dots, c_5) = (1, 2, 0, 0, 0)$ we may take

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

We can create the whole conjugacy class by conjugating with a general $\tau \in \mathrm{S}_n$, which gives

$$(\tau(1)\ \tau(2))(\tau(3)\ \tau(4))(\tau(5)).$$

We claim that, for each of the $n!$ choices for τ, there are always $\prod_{k=1}^{n} k^{c_k} c_k!$ choices that lead to distinct cycle notations for the same element. Namely, for each k-cycle we may pick any element that we write first, which gives k^{c_k} choices. Having made these choices we can still write the k-cycles in any order, which accounts for the factor $c_k!$.

We will need the *sign* of a permutation, which is a homomorphism $\mathrm{sgn} : \mathrm{S}_n \to \{\pm 1\}$. You may have encountered this in linear algebra in the definition of the determinant, which can be written[7]

$$\det_{1 \le i,j \le n}(a_{ij}) = \sum_{\sigma \in \mathrm{S}_n} \mathrm{sgn}(\sigma) \prod_{i=1}^{n} a_{i,\sigma(i)}. \tag{1.8}$$

An elegant definition of the sign is by the identity

$$\mathrm{sgn}(\sigma) = \prod_{1 \le i < j \le n} \frac{x_{\sigma(j)} - x_{\sigma(i)}}{x_j - x_i}, \tag{1.9}$$

where $x_1, \dots, x_n$ can be thought of as real or formal variables. Since each factor in the numerator is plus or minus a factor in the denominator, $\mathrm{sgn}(\sigma)$ takes values in $\{\pm 1\}$. It follows immediately that

$$\mathrm{sgn}(\sigma\tau) = \prod_{1 \le i < j \le n} \frac{x_{\sigma(\tau(j))} - x_{\sigma(\tau(i))}}{x_{\tau(j)} - x_{\tau(i)}} \prod_{1 \le i < j \le n} \frac{x_{\tau(j)} - x_{\tau(i)}}{x_j - x_i} = \mathrm{sgn}(\sigma)\,\mathrm{sgn}(\tau),$$

where in the first factor we used (1.9) with x_j replaced by $x_{\tau(j)}$.

Exercise 1.6.1. Pick two elements σ and τ in the group S_5. Define them by a table of values and then give their cycle notation. Compute $\sigma\tau$ and $\tau\sigma\tau^{-1}$ both using the tables and using cycle notation.

Exercise 1.6.2. Given $\sigma \in \mathrm{S}_n$, an *inversion* is a pair (i, j) with $1 \le i < j \le n$ and $\sigma(j) > \sigma(i)$. Show that $\mathrm{sgn}(\sigma)$ equals 1 if σ has an even number of inversions and -1 else.

[7] Actually, most authors of linear algebra textbooks make a complete mess of their chapter on determinants, precisely because they try very hard to avoid writing down this simple definition.

Exercise 1.6.3. Show that any permutation can be written as a product of transpositions (2-cycles). Show that $\operatorname{sgn}(\sigma)$ equals 1 if σ is the product of an even number of transpositions and -1 else.

1.7 Additional exercises

Exercise 1.7.1. Show that $\mathrm{GL}(2, \mathbb{Z}_2)$, the group of invertible matrices with elements in $\mathbb{Z}_2$, is isomorphic to S_3. (The definition of matrix multiplication should be obvious, but formally it is based on the fact that $\mathbb{Z}_2$ is a ring; see Chapter 2.)

Exercise 1.7.2. A proper subgroup H of G is a subgroup which is neither equal to $\{1\}$ nor to G. Show that if G has no proper subgroups, then $G \simeq \mathbb{Z}_p$, where p is a prime or $p = 1$.

Exercise 1.7.3. Show that the matrices

$$I = \begin{bmatrix} \mathrm{i} & 0 \\ 0 & -\mathrm{i} \end{bmatrix}, \qquad J = \begin{bmatrix} 0 & \mathrm{i} \\ \mathrm{i} & 0 \end{bmatrix}.$$

generate a subgroup of $\mathrm{GL}(2, \mathbb{C})$ with eight elements. Find a nice way to express the relations between these elements and carve them into a bridge.[8]

Exercise 1.7.4. If H, K and L are subgroups of a group G with $H \subseteq K \cup L$, show that either $H \subseteq K$ or $H \subseteq L$. In particular, $K \cup L$ is a subgroup if and only if $K \subseteq L$ or $L \subseteq K$.

Exercise 1.7.5. The *direct product* $G \times H$ of two groups is a group wih multiplication $(a, b) \cdot (c, d) = (ac, bd)$. Show that $\mathbb{Z}_m \times \mathbb{Z}_n \simeq \mathbb{Z}_{mn}$ if and only if m and n are relatively prime.

Exercise 1.7.6. Show that any automorphism of $\mathbb{Q}$ is given by $\phi(x) = ax$ for some fixed $a \in \mathbb{Q}^*$. Can you prove that this is true also for $\mathbb{R}$? (Hint: No, you can't.)

Exercise 1.7.7. Show that $\mathbb{Q} \not\simeq \mathbb{Q} \times \mathbb{Q}$ as groups. (It may surprise you to learn that $\mathbb{R} \simeq \mathbb{R} \times \mathbb{R}$ as groups and, more generally, $\mathbb{R}^m \simeq \mathbb{R}^n$ as groups for any m and n. However, it is impossible to give an isomorphism explicitly (unless $m = n$), since their existence depends on the axiom of choice.)

Exercise 1.7.8. Let G be a group, and let G' be the subgroup generated by all commutators $aba^{-1}b^{-1}$, a, $b \in G$. (It is called the *derived* group or the *commutator subgroup*.) Show that G' is a normal subgroup of G and that G/G' is abelian. More generally, show that a quotient G/N is abelian if and only if $G' \subseteq N$.

[8] The first person to do so was William Rowan Hamilton on 16 October 1843. The group is called the quaternion group and generates the famous quaternion algebra.

Exercise 1.7.9. If A and B are subsets of a group G, let $AB = \{ab;\ a \in A,\ b \in B\}$. Show that if H is a subgroup and N is a normal subgroup, then HN is a subgroup of G and $H \cap N$ a normal subgroup of H. Moreover, show that $HN/N \simeq H/(H \cap N)$.

Exercise 1.7.10. If M and N are normal subgroups in G with $M \subseteq N$, show that $G/N \simeq (G/M)/(N/M)$.

Exercise 1.7.11. Show that, for any $a \in G$, $\phi_a(b) = aba^{-1}$ defines an isomorphism from G to itself. Show that these maps, which are called *inner automorphisms*, form a group under composition.

Exercise 1.7.12. The *center* $\mathrm{Z}(G)$ of a group G is the set of elements $a \in G$ such that $ab = ba$ for all $b \in G$. Show that $\mathrm{Z}(G)$ is a normal subgroup and that $G/\mathrm{Z}(G)$ is isomorphic to the group of inner automorphisms of G.

Exercise 1.7.13. Show that if $G/\mathrm{Z}(G)$ is cyclic, then G is abelian and hence $\mathrm{Z}(G) = G$.

Exercise 1.7.14. If G acts on a set X and $x \in X$, show that $G_x = \{g \in G;\ gx = x\}$ is a subgroup of G. It is called the *stabilizer* or *isotropy group* of x. Let $Gx = \{gx;\ g \in G\}$ be the *orbit* of x. Show that there is a bijection between the orbit of x and the cosets of G_x. In particular, if G is a finite group,

$$|G| = |G_x| \cdot |Gx|, \qquad x \in G. \tag{1.10}$$

Exercise 1.7.15. If a permutation σ has cycle structure $(c_1, \ldots, c_n)$, what is the order of σ?

Exercise 1.7.16. Show that sgn is constant on each conjugacy class of S_n. Find an expression for $\mathrm{sgn}(\sigma)$ on permutations σ with cycle structure $(c_1, \ldots, c_n)$.

Exercise 1.7.17. Show that the group of rotations preserving a cube is isomorphic to S_4. (Hint: any rotation permutes the four long diagonals.) Explain the geometric interpretation of each of the five conjugacy classes in S_4 (for instance, one of them corresponds to rotations 120° around a long diagonal).

Exercise 1.7.18. Show that there are only two groups of order 4, namely, $\mathbb{Z}_4$ and $\mathbb{Z}_2^2 = \mathbb{Z}_2 \times \mathbb{Z}_2$ (see Exercise 1.7.5 for the definition of direct product).

Exercise 1.7.19. Show that if G has even order, then there is always an odd number of elements of order 2.

Exercise 1.7.20. Show that if all elements in a group G have order 1 or 2, then G is abelian and $|G| = 2^k$ for some k.[9]

[9]It follows from Theorem 5.1.1 below that $G \simeq \mathbb{Z}_2^k$.

Exercise 1.7.21. Classify all groups of order 6 by the following method. First use Exercises 1.7.19 and 1.7.20 to show that G has an element a of order 2 and an element b of order 3.[10] Then show that $G = \{1, b, b^2, a, ab, ab^2\}$. Show that either $ba = ab$ or $ba = ab^2$. Finally, show that $G = \mathbb{Z}_6$ or $G = \mathrm{S}_3$.

[10]This also follows from a theorem of Cauchy: if p is a prime that divides $|G|$, then G has an element of order p.

Chapter 2

Review of ring theory

2.1 Definition

Structures equipped with both addition and multiplication are ubiquitous in mathematics. Familiar examples include numbers, square matrices and polynomials. The notion of a ring is an abstract version of such structures.

Definition 2.1.1. *A* ring *R is a set equipped with two maps $R \times R \to R$, called addition and multiplication and denoted $(a, b) \mapsto a + b$ and $(a, b) \mapsto a \cdot b = ab$. It should be an abelian group under addition, that is,*

$$a + (b + c) = (a + b) + c, \qquad a + b = b + a, \quad a,\, b,\, c \in R$$

and there exists an element $0 \in R$ and a map $R \to R$ denoted $a \mapsto -a$ such that

$$a + 0 = a, \qquad a + (-a) = 0, \quad a \in R.$$

Moreover, the multiplication should be associative,

$$a \cdot (b \cdot c) = (a \cdot b) \cdot c, \qquad a,\, b,\, c \in R,$$

and satisfy the distributive laws

$$a \cdot (b + c) = a \cdot b + a \cdot c, \qquad (a + b) \cdot c = a \cdot c + b \cdot c, \quad a,\, b,\, c \in R.$$

We will also assume that multiplication has a unit element, that is, there is an element $1 \in R$ such that

$$a \cdot 1 = 1 \cdot a = a, \qquad a \in R.$$

The reader should be aware that the precise definition of a ring varies; in particular, many authors do not assume the existence of a multiplicative unit.

Some examples:

- The number systems $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ are rings.

- The even integers $2\mathbb{Z}$ satisfy all axioms except that it does not have a multiplicative unit. With our conventions, it is not a ring.
- Sets of polynomials such as $\mathbb{Z}[x]$, $\mathbb{R}[x, y]$ and, more generally, $R[x_1, \ldots, x_n]$ when R is a ring, are all rings.
- If G is an abelian group, then the endomorphisms of G, that is, the group homomorphisms from G to itself, forms a ring with the operations $(\phi \cdot \psi)(a) = \phi(\psi(a))$ and $(\phi + \psi)(a) = \phi(a) + \psi(a)$. We will denote it $\mathrm{End}(G)$.
- The $n \times n$-matrices with entries in a ring R form a ring.
- Addition and multiplication modulo n give a ring structure to $\mathbb{Z}_n$.
- The ring $\mathbb{Z}_1$ containing only one element is known as the zero ring or the trivial ring. It is the only ring such that $0 = 1$.
- The space $\mathbb{R}^3$ with vector addition and vector multiplication ("cross product") satisfies most axioms but the product is not associative and does not have a unit. It is not a ring.

Additional axioms give special classes of rings.

- If the commutative law $ab = ba$ holds, then R is called a *commutative ring*.
- A non-zero ring where all non-zero elements are invertible, that is, if for each $a \neq 0$ there exists an element a^{-1} with $aa^{-1} = a^{-1}a = 1$, is called a *division ring* or a *skew field*.
- A commutative division ring is called a *field*[1].
- A non-zero ring without zero divisors, that is, $ab = 0$ implies $a = 0$ or $b = 0$, is called a *domain*. It is useful to know that this is equivalent to the *cancellation property*, see Exercise 2.1.2. All division rings are domains.
- A commutative domain is called an *integral domain*[2]. All fields are integral domains.

As examples, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ are fields, whereas $\mathbb{Z}$ is only an integral domain. The ring $\mathbb{Z}_n$ is a field when n is prime, but otherwise it is not even a domain.

Exercise 2.1.1. Show that in a ring with more than one element, $0 \neq 1$.

[1]Swedish: kropp. Fields are often denoted K or k, after the German counterpart Körper.
[2]Swedish: integritetsområde.

Exercise 2.1.2. Show that, in a ring R, the left cancellation property $ab = ac \Rightarrow a = 0$ or $b = c$ and the right cancellation property $ac = bc \Rightarrow a = b$ or $c = 0$ are both equivalent to R being a domain.

Exercise 2.1.3. Prove that the ring of 2×2-matrices over a ring R is never a domain. Prove that it is commutative only if R is the zero ring.

Exercise 2.1.4. Give an example of a ring with two elements a and b such that $ab = 1$ but $ba \neq 1$.

Exercise 2.1.5. Let R be the power set (set of all subsets) of a set S. Show that R is a ring with the operations

$$a + b = (a \cup b) \setminus (a \cap b), \qquad ab = a \cap b.$$

2.2 Ideals and homomorphisms

As for groups, we need a notion of homomorphisms.

Definition 2.2.1. *A ring homomorphism is a map $\phi : R \to S$ between two rings, such that*

$$\phi(x + y) = \phi(x) + \phi(y), \quad \phi(xy) = \phi(x)\phi(y), \quad \phi(1) = 1, \qquad x, y \in R.$$

A bijective homomorphism is called an isomorphism.

Note that $\phi(0) = 0$ follows by taking $y = 0$ in $\phi(x + y) = \phi(x) + \phi(y)$ and adding $-\phi(x)$ to both sides. We must add $\phi(1) = 1$ as a separate axiom as we cannot always cancel $\phi(x)$ from the identity $\phi(x)\phi(1) = \phi(x)$.

In groups, kernels of homomorphisms are the same as normal subgroups. In rings, the corresponding notion is *ideal*.

Definition 2.2.2. *An* ideal *is a non-empty subset I of a ring R such that*

$$a, b \in I \quad \Rightarrow \quad a - b \in I,$$

$$a \in I,\ r \in R \quad \Rightarrow \quad ar,\ ra \in I.$$

It follows that $0 \in I$, $-a \in I$ and $a + b \in I$, for a, $b \in I$. Note that if $1 \in I$ then the second property gives $I = R$. Thus, non-trivial ideals do not contain 1 and are, in particular, not subrings.

Later in the course, the following weaker concept will be important. Clearly, the two notions are equivalent for commutative rings.

Definition 2.2.3. *A* left ideal *is a non-empty subset I of a ring R such that*

$$a, b \in I \quad \Rightarrow \quad a - b \in I,$$

$$a \in I,\ r \in R \quad \Rightarrow \quad ra \in I.$$

Lemma 2.2.4. *The ideals in $\mathbb{Z}$ are precisely the subsets of the form $n\mathbb{Z}$, with $n \in \mathbb{Z}$.*

Proof. Let I be an ideal in $\mathbb{Z}$. Either $I = \{0\} = 0\mathbb{Z}$ or I contains a smallest positive integer n. Suppose $m \in I$ and apply the division algorithm to write $m = nq + r$, where $0 \leq r \leq n - 1$. By the definition of an ideal, $r \in I$ and, by our choice of n, we must have $r = 0$. Hence $I \subseteq n\mathbb{Z}$. The reverse inclusion follows from $n \in I$ and the definition of an ideal. □

The intersection of any family of ideals is an ideal, and we can therefore define the ideal generated by a set $S \subseteq R$ as the smallest ideal containing S. The ideal generated by a finite set $\{a_1, \ldots, a_n\}$ will be denoted $(a_1, \ldots, a_n)$. An ideal (a) generated by a single element is called a *principal ideal*. In a commutative ring, we can write

$$(a_1, \ldots, a_n) = \{a_1x_1 + \cdots + a_nx_n;\ x_1, \ldots, x_n \in R\} \tag{2.1}$$

and, in particular, $(a) = aR = \{ar;\ r \in R\}$ (see Exercise 2.2.2 for the case of non-commutative rings).

Lemma 2.2.4 shows that all ideals in $\mathbb{Z}$ are principal, so they correspond more or less to numbers (except for the fact that n and $-n$ generate the same ideal). Some rings also contain non-principal ideals. An example is the ideal $\{a + ib\sqrt{3};\ a,\ b \in 2\mathbb{Z}\}$ in the ring $\mathbb{Z}[\sqrt{-3}] = \{a+ib\sqrt{3};\ a,\ b \in \mathbb{Z}\}$. In work that predated ring theory, Kummer developed a theory of "ideal numbers" to deal with problems caused by such non-principal ideals. This is the historical origin of the term "ideal".

Definition 2.2.5. *A* principal ideal domain[3] *or* PID *is an integral domain where all ideals are principal.*

Examples of PIDs are $\mathbb{Z}$ and (with almost the same proof as Lemma 2.2.4) the polynomial ring $k[x]$ for any field k. Examples of integral domains that are not PIDs are $\mathbb{Z}[x]$ and $\mathbb{R}[x, y]$.

If I is an ideal, then the cosets $r + I = \{r + x;\ x \in I\}$ form a ring with the obvious operations $r + I + s + I = r + s + I$, $(r + I)(s + I) = rs + I$. This *quotient ring* is denoted R/I. An example is $\mathbb{Z}/n\mathbb{Z}$, which is isomorphic to $\mathbb{Z}_n$ not only as an additive group but also as a ring.

[3]Swedish: huvudidealområde.

Just as for groups, we have the following first isomorphism theorem for rings. (For the second and third isomorphism theorems, see Exercises 2.4.2–2.4.3.) Again, we leave the proof to the reader.

Theorem 2.2.6. *If $\phi : R \to S$ is a ring homomorphism, then the kernel* $\operatorname{Ker}(\phi) = \{x \in R;\, \phi(x) = 0\}$ *is an ideal in R and $R/\operatorname{Ker}(\phi) \simeq \operatorname{Im}(\phi)$ as rings. Conversely, if $I \subseteq R$ is an ideal, then $\phi(x) = x + I$ defines a surjective ring homomorphism $R \to R/I$ with* $\operatorname{Ker}(\phi) = I$.

In a nutshell: ideals are precisely kernels of homomorphisms and quotients are precisely images of homomorphisms.

To give a familiar example, consider the principal ideal $I = (x^2+1)$ in $R = \mathbb{R}[x]$. The ring R/I formally consists of polynomials in x where x^2+1 is identified with 0. After this identification, any such polynomial can be written $a+bx$ with a, $b \in \mathbb{R}$, where the multiplication is determined by the rule $x^2 = -1$. This is precisely the complex numbers, though one usually writes i rather than x. To relate this to Theorem 2.2.6, consider the ring homomorphism $\phi : R \to \mathbb{C}$ given by $\phi(p) = p(\mathrm{i})$. It is surjective with kernel I, so we can indeed deduce from Theorem 2.2.6 that $R/I \simeq \mathbb{C}$.

Exercise 2.2.1. Show that, for any ring R, there is a unique ring homomorphism $\mathbb{Z} \to R$.

Exercise 2.2.2. Let R be a ring and $S \subseteq R$ an arbitrary subset. Show that the smallest left ideal containing S is

$$\{x_1 s_1 + \cdots + x_m s_m;\, m \in \mathbb{Z}_{>0},\, x_j \in R,\, s_j \in S,\, s_j \neq s_k \text{ for } j \neq k\}.$$

In particular, when S is finite with n elements, we can take $m = n$ and recover (2.1) for ideals in commutative rings. Show that the smallest ideal containing S is

$$\{x_1 s_1 y_1 + \cdots + x_m s_m y_m;\, m \in \mathbb{Z}_{>0},\, x_j,\, y_j \in R,\, s_j \in S\},$$

where the s_j are not assumed to be distinct. In particular, a principal ideal in a non-commutative ring has the form

$$(a) = \{x_1 a y_1 + \cdots + x_m a y_m;\, m \in \mathbb{Z}_{>0},\, x_j,\, y_j \in R\}.$$

Exercise 2.2.3. Let R be the ring of 2×2-matrices over some ring S and let $I \subseteq R$ be the matrices whose second column is zero. Show that I is a left ideal in R. Show that it is an ideal only if S is the zero ring.

Exercise 2.2.4. Show that all ideals in $\mathbb{Z}_n$ are principal. For which n is $\mathbb{Z}_n$ a principal ideal domain?

Exercise 2.2.5. Show that (x, y) is a non-principal ideal in $\mathbb{R}[x, y]$.

Exercise 2.2.6. Let I be the set of polynomials p in $\mathbb{Z}[x]$ such that $p(0)$ is even. Show that I is not a principal ideal.

2.3 Unique factorization

Some of the most important theorems in mathematics are about unique factorization in rings. For instance, the fundamental theorem of arithmetic states that any positive integer can be written uniquely as a product of primes (up to reordering the factors). The fundamental theorem of algebra implies that any complex polynomial in one variable has a unique factorization of the form $C(x-a_1)\cdots(x-a_n)$, again up to reordering. We want to put such results in a general framework. In particular, we will show a unique factorization result in a general principal ideal domain (PID).

Let R be an integral domain. We recall that this means that R is a non-zero commutative ring with the cancellation law

$$ab = ac \quad \Longrightarrow \quad a = 0 \ \text{ or } \ b = c.$$

By a *unit* in R we mean an invertible element. (For instance, the units in $\mathbb{Z}$ are ± 1 and the units in $\mathbb{C}[x]$ are non-zero constants.) We will sometimes write $a \sim b$ if $a = eb$ for some unit e.

Lemma 2.3.1. *If a and b are elements in an integral domain R, then $(a) = (b)$ if and only if $a \sim b$.*

Proof. If $(a) = (b)$, then $a = bc$ and $b = ad$ for some $c, d \in R$. It follows that $a = acd$. Applying the cancellation law gives $a = 0$ or $cd = 1$. In the first case, we must have $b = 0$ and in the second case $c = d^{-1}$ is a unit. In any case we may conclude that $a \sim b$. The converse is easy. □

By an *irreducible element* we mean a non-zero non-unit p such that if $p = ab$ then either a or b is a unit. Equivalently, p is *not* irreducible if we can write $p = ab$ with neither a nor b a unit.

We write $a \mid b$ if $b = ac$ for some $c \in R$. By a *prime element* we mean a non-zero non-unit p such that if $p \mid ab$ then $p \mid a$ or $p \mid b$. It is easy to see that any prime element is irreducible.

In $\mathbb{Z}$ there is no difference between prime elements and irreducible elements; both notions mean $\pm p$ with p a prime number. Likewise, in $\mathbb{C}[x]$ both notions mean a polynomial of degree 1 (to prove this one needs the fundamental theorem of algebra).

By contrast, in the ring $\mathbb{Z}[\sqrt{-5}]$, it is easy to check that the elements 3 and $2 \pm \sqrt{-5}$ are all irreducible. However, in view of the identity

$$3 \cdot 3 = (2 + \sqrt{-5}) \cdot (2 - \sqrt{-5}),$$

none of these elements are prime. For instance, $2 + \sqrt{-5} \mid 3 \cdot 3$ but $2 + \sqrt{-5} \nmid 3$. It would seem from this example that non-prime irreducible elements are related to non-unique factorizations. Indeed, we have the following result.

Proposition 2.3.2. *Let R be an integral domain such that any non-zero non-unit can be written as a product of irreducible elements. Then, the following two conditions are equivalent:*

(A) *Every irreducible element is prime.*

(B) *Factorization into irreducible elements is unique up to reordering and multiplication by units.*

More precisely, condition (B) means that if

$$x = p_1 \cdots p_k = q_1 \cdots q_l$$

with all p_j and q_j irreducible, then $k = l$ and, after reordering the elements, $p_j \sim q_j$ for all j.

A domain such that the conditions in Proposition 2.3.2 hold (including the statement in the first sentence) is called a *unique factorization domain* (UFD).

Proof of Prop. 2.3.2. To prove that (A) implies (B), we use induction on the minimal number of irreducible factors of x. If $x = p_1$ is itself irreducible, then the statement in (B) follows from the definition of irreducibility. Suppose now that (B) holds for all products of k irreducible factors and consider a product $x = p_1 \cdots p_{k+1}$ with all p_j irreducible. If we also have $x = q_1 \cdots q_l$, then since p_1 is prime it divides one of the factors q_j. After reordering, we may assume $p_1 \mid q_1$. Since q_1 is irreducible, $p_1 = eq_1$ with e a unit. Cancelling p_1 we get

$$p_2 \cdots p_{k+1} = eq_2 \cdots q_l.$$

We may now apply the induction hypothesis to deduce that (B) holds for the element x.

To prove that (B) implies (A), let p be irreducible and suppose that $p \mid ab$, that is, $ab = pc$ for some c. We factor each of the elements a, b and c into irreducibles and plug into the equation $ab = pc$ (if the element is a unit, we leave it as it is). By (B), p is equivalent to one of the irreducible factors on the left-hand side. If that factor comes from a we have $p \mid a$ and else $p \mid b$. □

The following result will be important in §5.

Theorem 2.3.3. *Every* PID *is a* UFD.

To prove Theorem 2.3.3 we need the following lemma.

Lemma 2.3.4. *Any* PID *is a* Noetherian ring, *that is, it does not contain any infinite strictly increasing chain of ideals*

$$I_1 \subsetneq I_2 \subsetneq I_3 \subsetneq \cdots .$$

Proof. Suppose there is such a chain and let $I = \cup_k I_k$. It is easy to see that I is an ideal. (For instance, if $x \in I$ and $y \in I$, then $x \in I_k$ and $y \in I_l$ for some k and l. If $l \geq k$, then $x \in I_l$ so, by the fact that I_l is an ideal, $x - y \in I_l \subseteq I$.) By the definition of a PID, $I = aR$ for some $a \in R$. Since $a \in I$, there exists a k with $a \in I_k$. But then, $I = aR \subseteq I_k$, so $I = I_k$. This contradicts the assumption that $I_k \subsetneq I_{k+1} \subseteq I$. □

We can now prove that a PID satisfies the first requirement for being an UFD.

Lemma 2.3.5. *In a* PID*, any non-zero non-unit can be factored as a product of irreducible elements.*

Proof. Take a non-zero non-unit element a. If it is irreducible we are done. If not, we can write $a = bc$ with b and c non-units. This implies $(a) \subsetneq (b)$. For if $(a) = (b)$, then Lemma 2.3.1 gives $a = be$ for some unit e. Then, $ac = ae$ implies $c = e$, which contradicts c being a non-unit. This argument can be repeated. If b and c are both irreducible we are done; otherwise one of them, say b, can be factored. The factors of b generate ideals strictly larger than (b). Continuing in this way we will either end up with a factorization of a into irreducibles or with an infinite strictly increasing chain of ideals. However, the second scenario is impossible in view of Lemma 2.3.4. □

Next, we will verify condition (A) in Proposition 2.3.2. It will be useful to know some facts about greatest common divisors. Let a and b be elements of a PID. Then $(a, b) = (c)$ for some element c, which is uniquely determined up to multiplication by a unit. It is clear that $c \mid a$, $c \mid b$ and that if d divides both a and b then $d \mid c$. Thus, it make sense to call c the *greatest common divisor* of a and b, and write $c = \gcd(a, b)$. Since $c \in (a, b) = aR + bR$ we can write $c = as + bt$ for some $s, t \in R$. In particular, if $\gcd(a, b) = 1$ we have Bezout's identity $as + bt = 1$.

Lemma 2.3.6. *In a* PID*, every irreducible element is prime.*

Proof. Let p be an irreducible element and assume that $p \mid ab$. Let $c = \gcd(p, a)$. Then, $p = cd$ for some element d. Since p is irreducible either c or d is a unit. If d is a unit then $p \sim \gcd(p, a)$ and hence $p \mid a$. If c is a unit then Bezout's identity gives $1 = ps + at$ for some elements s and t. Then $b = pbs + abt$ and since $p \mid ab$ we may conclude that $p \mid b$. □

This completes the proof of Theorem 2.3.3.

Finally, we mention without proof the following fact.

Theorem 2.3.7. *If R is a* UFD *then so is the polynomial ring $R[x]$.*

By iteration, $R[x_1, \dots, x_n]$ is a UFD. This shows that the class of UFDs is much larger than the class of PIDs; for instance, $\mathbb{Z}[x]$ and $\mathbb{R}[x, y]$ are UFDs but not PIDs.

To summarize some relations between various classes of rings, we have

$$\text{fields} \subseteq \text{PIDs} \subseteq \text{UFDs} \subseteq \text{integral domains} \subseteq \text{commutative rings} \subseteq \text{rings}. \quad (2.2)$$

Exercise 2.3.1. Where do the rings $\mathbb{Z}_3$, $\mathbb{Z}_6$, $\mathbb{Z}$, $\mathbb{Z}[x]$, $\mathbb{Z}[x, y]$, $\mathbb{R}[x]$ and $\mathbb{R}[x, y]$ fit in the chain of inclusions (2.2)?

Exercise 2.3.2. The real numbers $\mathbb{R}$ is a PID. What is the prime factorization of a real number?

Exercise 2.3.3. Let $R = \mathbb{R} + x^2\mathbb{R}[x] \subseteq \mathbb{R}[x]$. Show that R is not a UFD.

Exercise 2.3.4. Let $R = \mathbb{R}[x_1, x_2, \dots]$ (the elements in this ring are normal polynomials like $7 + 5x_1x_5 + 13x_6^2x_{1000}$, but there is no bound on the number of variables). Show that R is not Noetherian.

2.4 Additional exercises

Exercise 2.4.1. A ring is called *Boolean* if $a^2 = a$ for all a. Show that any Boolean ring is commutative and that $-a = a$ for all a. (One can prove that any Boolean ring is isomorphic to a subring of a ring of sets defined as in Exercise 2.1.5.)

Exercise 2.4.2. Let I be an ideal and S a subring of a ring R. Show that $S + I = \{s + i;\ s \in S,\ i \in I\}$ is a subring of R and $S \cap I$ an ideal in S. Moreover, show that $(S + I)/I \simeq S/(S \cap I)$.

Exercise 2.4.3. If I and J are ideals in a ring R with $I \subseteq J$, show that J/I is an ideal in R/I and that $R/J \simeq (R/I)/(J/I)$. Also show that any ideal in R/I has this form.

Exercise 2.4.4. Let K be a field and consider the unique homomorphism $\phi : \mathbb{Z} \to K$ (see Exercise 2.2.1). Show that $\mathrm{Ker}(\phi) = p\mathbb{Z}$, where p is either zero or a prime p. The number p is called the *characteristic* of K. (Slightly illogically, the case $p \neq 0$ is often called *finite characteristic*.)

Exercise 2.4.5. In a PID, one may define the *least common multiple* as $\mathrm{lcm}(a, b) = ab/\gcd(a, b)$ (it is defined up to multiplication by a unity). Prove that $aR \cap bR = \mathrm{lcm}(a, b)R$.

Exercise 2.4.6. Let R be a PID and $a, b \in R$ with $\gcd(a, b) = 1$. Show that $R/abR \simeq R/aR \times R/bR$.

Exercise 2.4.7. By Exercise 2.4.6, $\mathbb{Z}_{12} \simeq \mathbb{Z}_4 \times \mathbb{Z}_3$ both as groups and as rings. How many group isomorphisms are there between them? How many ring isomorphisms are there?

Exercise 2.4.8. Show that if R is an integral domain but not a field, then $R[x]$ is not a PID.

Chapter 3

Modules

3.1 Definition

We will give two equivalent definitions of modules. In the first, we view modules as a generalization of vector spaces, where the field of scalars is replaced by a general ring.

Definition 3.1.1. *If R is a ring, an R-*module *M (or a module M over R) is an abelian group equipped with a map $R \times M \to M$, denoted $(r, m) \mapsto rm$, such that*

$$r(m_1 + m_2) = rm_1 + rm_2, \qquad (r_1 + r_2)m = r_1 m + r_2 m, \tag{3.1a}$$

$$(r_1 r_2)m = r_1(r_2 m), \qquad 1m = m \tag{3.1b}$$

for all r, r_1, $r_2 \in R$ and m, m_1, $m_2 \in M$.

As we have indicated, a module over a field is called a *vector space*. Just as in linear algebra, we sometimes refer to the elements of R as *scalars* and to the product $(r, m) \mapsto rm$ as *scalar multiplication*.

In our second definition, we think of modules as representations of rings. Recall that a group representation is a homomorphism $G \to \mathrm{GL}(V)$ for a vector space V. In other words, the abstract group G is realized concretely by invertible linear maps on a vector space. For rings, we will instead consider realizations by endomorphisms of an abelian group.

Definition 3.1.2. *If R is a ring, an R-module M is an abelian group equipped with a ring homomorphism $\phi : R \to \mathrm{End}(M)$.*

It is straight-forward to check that the two definitions of a module are equivalent, via the correspondence $\phi(r)(m) = rm$.

We will now give some examples of modules.

- If M is a $\mathbb{Z}$-module, then it is easy to see that

$$kx = \begin{cases} \underbrace{x + \cdots + x}_{k}, & k > 0, \\ 0, & k = 0, \\ -(\underbrace{x + x + \cdots + x}_{k}), & k < 0. \end{cases}$$

  Conversely, these relations define a $\mathbb{Z}$-module structure on any abelian group. Thus, $\mathbb{Z}$-modules are essentially the same as abelian groups (if you insist on a more precise statement, they are equivalent as categories). This might seem uninteresting, but in §5 we will derive quite non-trivial facts on abelian groups as special cases of results for modules.

- If $I \subseteq R$ is a left ideal, then I is an R-module, with the module structure given by the ring multiplication. In particular, any ring is a module over itself.

- The direct product $R^n = R \times \cdots \times R$ is an R-module with componentwise addition $(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n)$ and scalar multiplication $r(x_1, \ldots, x_n) = (rx_1, \ldots, rx_n)$.

- Let X be any set and R^X be the set of all functions from X to a ring R. Then, R^X is an R-module with addition $(f + g)(x) = f(x) + g(x)$ and scalar multiplication $(rf)(x) = r\, f(x)$. When $X = \{1, \ldots, n\}$ is finite we can identify R^X with R^n, simply by identifying f with $(f(1), \ldots, f(n)) \in R^n$.

Exercise 3.1.1. Prove that, in an R-module, $0m = 0$ and $(-1)m = -m$ (on the left, 0 is the zero element in R and -1 the additive inverse of 1 in R; on the right, 0 is the zero element in M and $-m$ the additive inverse of m in M).

Exercise 3.1.2. Let I be a left ideal in a ring R. Show that R/I is an R-module in a natural way.

Exercise 3.1.3. Let $\phi : R \to S$ be a homomorphism of rings. Show that S is an R-module with scalar multiplication $rs = \phi(r)s$.

3.2 Submodules, homomorphisms and quotients

In analogy to the cases of groups and rings, we will use the following terminology.

Definition 3.2.1. *An* R-module homomorphism *is a map* $\phi : M \to N$ *between* R*-modules such that* $\phi(x + y) = \phi(x) + \phi(y)$ *and* $\phi(rx) = r\phi(x)$ *for all* $r \in R$ *and* $x, y \in M$*. If* ϕ *is also bijective it is called an* R-module isomorphism*. We then call* M *and* N isomorphic *and write* $M \simeq N$*.*

When R is a field, an R-module homomorphism is called a linear map. We will sometimes write *R-linear map* instead of R-module homomorphism. The set of all R-module homomorphisms from M to N will be denoted $\mathrm{Hom}_R(M, N)$.

Definition 3.2.2. *A subset $N \subseteq M$ of an R-module is called an* R-submodule *if it is a subgroup and $rn \in N$ for all $r \in R$ and $n \in N$.*

When R is a field, a submodule is called a linear subspace.

If $N \subseteq M$ is a submodule, we can define the *quotient module* M/N as the set of all cosets $m+N = \{m+n;\ n \in N\}$, where $m \in M$. We know from §1.4 that M/N is an abelian group. More generally, it is an R-module with scalar multiplication $r(m + N) = rm + N$. To see that this is well-defined, we must check that if $m + N = m' + N$ then $rm + N = rm' + N$. Equivalently, if $m - m' \in N$ then $r(m - m') \in N$. This follows from N being a submodule. It is straight-forward to check that the axioms (3.1) hold in M/N.

It is easy to check that, if $\phi : M \to N$ is an R-module homomorphism, then the kernel $\mathrm{Ker}(\phi) = \{x \in M;\ \phi(x) = 0\}$ and image $\mathrm{Im}(\phi) = \{\phi(x) \in N;\ x \in M\}$ are submodules. We then have the following fundamental isomorphism theorem. (This is often called the first isomorphism theorem for modules; the second and third are given as Exercise 3.7.1–3.7.2.) Again, we leave the proof as an exercise.

Theorem 3.2.3. *If $\phi : M \to N$ is an R-module homomorphism, then $M/\mathrm{Ker}(\phi) \simeq \mathrm{Im}(\phi)$ as R-modules. Conversely, if N is an R-submodule of M, then the map $m \mapsto m + N$ from M to M/N is an R-module homomorphism with kernel N and image M/N.*

This means that the R-submodules of M are exactly the kernels of R-module homomorphisms defined on M, and the quotients of M are precisely the images of such homomorphisms.

For vector spaces, Theorem 3.2.3 is closely related to the dimension theorem of linear algebra. Namely, let A be a real $(n \times m)$-matrix. Then A can be identified with a linear map $\mathbb{R}^m \to \mathbb{R}^n$. The kernel, or nullspace, of A is a subspace of some dimension k and can thus be identified with $\mathbb{R}^k$. Similarly, the image, or range, of A can be identified with $\mathbb{R}^l$, where l is known as the rank of A. Theorem 3.2.3 then gives an isomorphism $\mathbb{R}^m/\mathbb{R}^k \simeq \mathbb{R}^l$. Using basic facts on dimension of vector spaces one may conclude that $m - k = l$, which is indeed the content of the dimension theorem.

Note that any ring is an R-module over itself. To avoid a common source of confusion, one must understand that the R-module structure is quite different from the ring structure. For instance, a ring homomorphism is required to satisfy $\phi(xy) = \phi(x)\phi(y)$ and a module homomorphism $\phi(xy) = x\phi(y)$. As an example, the map $\phi(x) = 2x$ from $\mathbb{Z}$ to $\mathbb{Z}$ is a $\mathbb{Z}$-module homomorphism but not a ring homomorphism. Also note that an R-submodule of a ring is precisely a left ideal.

Exercise 3.2.1. Show that, for any ring R, $R[x] \simeq R[x, y]$ as R-modules.

Exercise 3.2.2. Describe $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}, \mathbb{Z})$, $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Q}, \mathbb{Z})$ and $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}, \mathbb{Q})$.

Exercise 3.2.3. If M and N are modules over a commutative ring R, show that $\mathrm{Hom}_R(M, N)$ is an R-module with the obvious operations $(\phi+\psi)(x) = \phi(x)+\psi(x)$, $(r\phi)(x) = r\phi(x)$. In particular, $M^* = \mathrm{Hom}_R(M, R)$ is called the *dual module* of M.

Exercise 3.2.4. A *right module* over a ring R is an abelian group M equipped with a map $(m, r) \mapsto mr$ satisfying the obvious modifications of (3.1). Show that if M is a left module, then M^* is a right module in a natural way.

3.3 Generators and cyclic modules

It is easy to see that the intersection of any family of submodules is a submodule. Thus, if $S \subseteq M$ is any subset, the intersection of all submodules of M containing S is the smallest submodule with that property. We call it the submodule *generated by* S. More explicitly, it is given by $\{r_1x_1 + \cdots + r_mx_m;\ m \in \mathbb{N},\ r_i \in R,\ x_i \in S\}$, where we can take the x_i as distinct. In particular, if $S = \{x_1, \ldots, x_n\}$ is finite, then the module generated by S is $Rx_1 + \cdots + Rx_n = \{r_1x_1 + \cdots + r_nx_n;\ r_i \in R\}$. A module generated by a finite set is called *finitely generated*. A module Rx generated by a single element is called *cyclic*.

Proposition 1.3.1 can be reformulated as saying that a cyclic module over $\mathbb{Z}$ is isomorphic to a quotient of $\mathbb{Z}$. We will generalize this to modules over any ring. It will be useful to define the *annihilator* of an element x in an R-module M to be the left ideal

$$\mathrm{Ann}(x) = \{r \in R,\ rx = 0\}.$$

Lemma 3.3.1. *Let M be a module over a ring R. Then, the following are equivalent:*

(A) *M is cyclic, that is, $M = Rx$ for some $x \in M$,*

(B) *$M \simeq R/I$ for some left ideal $I \subseteq R$.*

More precisely, a module Rx generated by x is isomorphic to $R/\mathrm{Ann}(x)$.

Proof. If $M = Rx$, then $\phi(r) = rx$ gives a surjective homomorphism $R \to M$ with kernel $I = \mathrm{Ann}(x)$, so $Rx \simeq R/I$. Conversely, if we let $x = 1 + I \in R/I$, then x generates R/I and $\mathrm{Ann}(x) = I$. □

Exercise 3.3.1. Show that an R-module M is finitely generated if and only if $M \simeq R^n/N$ for some n and some R-submodule $N \subseteq R^n$.

3.4 Direct sums

When K and L are modules, $K \times L$ is a module with $(k_1, l_1) + (k_2, l_2) = (k_1 + k_2, l_1 + l_2)$ and $r(k, l) = (rk, rl)$. Often, we have a module M and ask whether it is isomorphic to $K \times L$. Since $K \simeq K \times \{0\} \subseteq K \times L$ and $L \simeq \{0\} \times L$, one may assume that K and L are submodules of M. The following criteria are then useful.

Lemma 3.4.1. *Let K, L be submodules of some module M. Then, the following are equivalent:*

(A) *The map $(k, l) \mapsto k + l$ is a module isomorphism $K \times L \to M$.*

(B) *Any element $x \in M$ can be written uniquely as $x = k + l$, $k \in K$, $l \in L$.*

(C) *$K + L = M$ and $K \cap L = \{0\}$.*

(D) *There exists a module homomorphism $P : M \to M$ such that $P^2 = P$, $\mathrm{Im}(P) = K$ and $\mathrm{Ker}(P) = L$.*

The proof is straight-forward and left to the reader. If these conditions hold, we write $M = K \oplus L$ and call M the *direct sum* of K and L. We will use the same notation $K \oplus L$ for the module $K \times L$, when K and L are not a priori defined as submodules of some module.

If K is a submodule of M, it is natural to ask whether there is a submodule L of M such that $M = K \oplus L$. If that is the case, we call K a *direct summand* of M and L a *complement* of K.

Not all submodules are direct summands; for instance, you may check that the $\mathbb{Z}$-module $\mathbb{Z}_4$ contains a submodule isomorphic to $\mathbb{Z}_2$, which is not a direct summand.

We will now give criteria for determining whether a submodule is a direct summand. We need some terminology. An *exact sequence* is a finite or infinite sequence of modules and homomorphisms

$$\cdots \to L \xrightarrow{f} M \xrightarrow{g} N \to \cdots$$

such that $\mathrm{Im}(f) = \mathrm{Ker}(g)$ at each step. A *short* exact sequence has the form

$$0 \to L \xrightarrow{f} M \xrightarrow{g} N \to 0, \tag{3.2}$$

where $0 = \{0\}$ is the trivial R-module. This means that f is injective, $\mathrm{Im}(f) = \mathrm{Ker}(g)$ and g is surjective. Thus, by Theorem 3.2.3, up to isomorphism any short exact sequence has the form

$$0 \to L \to M \to M/L \to 0, \tag{3.3}$$

where L is a submodule of M.

Lemma 3.4.2. *Given a short exact sequence* (3.2)*, the following are equivalent:*

(A) *There is a homomorphism* $\phi : M \to L$ *with* $\phi \circ f = \mathrm{id}$.

(B) *There is a homomorphism* $\psi : N \to M$ *with* $g \circ \psi = \mathrm{id}$.

(C) *The submodule* $\mathrm{Im}(f) = \mathrm{Ker}(g)$ *in* M *is a direct summand.*

If these conditions hold, we say that (3.2) is a *split exact sequence* and that ϕ or ψ *splits* the sequence. Note that since f is injective, $\mathrm{Im}(f) \simeq L$. If (C) holds then $M \simeq L \oplus K$ for some R-module K. But then $K \simeq M/L \simeq N$. Thus, if the sequence (3.2) splits we have

$$M \simeq L \oplus N.$$

Proof. We prove the equivalence of (B) and (C). The equivalence of (A) and (C) is similar and left as an exercise.

Assuming (B), we show that $M = \mathrm{Ker}(g) \oplus \mathrm{Im}(\psi)$. By Lemma 3.4.1, it is enough to check that $M = \mathrm{Ker}(g) + \mathrm{Im}(\psi)$ and $\mathrm{Ker}(g) \cap \mathrm{Im}(\psi) = \{0\}$. For the first part, we write $x = x_1 + x_2$, where $x_1 = x - \psi(g(x))$ and $x_2 = \psi(g(x))$. It is then clear that $x_1 \in \mathrm{Ker}(g)$ and, obviously, $x_2 \in \mathrm{Im}(\psi)$. For the second part, suppose $y \in \mathrm{Ker}(g) \cap \mathrm{Im}(\psi)$, so $y = \psi(b)$. Then $0 = g(y) = g(\psi(b)) = b$ so $y = \psi(0) = 0$.

To show that (C) implies (B), let $M_1 = \mathrm{Ker}(g)$ and suppose there exists M_2 with $M = M_1 \oplus M_2$. Let h be the restriction of g to M_2. We claim that $h : M_2 \to N$ is an isomorphism. Indeed, it is injective since $\mathrm{Ker}(h) = \mathrm{Ker}(g) \cap M_2 = M_1 \cap M_2 = \{0\}$. To see that h is surjective, take $y \in N$ arbitrary and choose $x \in M$ with $g(x) = y$. Decomposing $x = x_1 + x_2$, $x_i \in M_i$, we have $y = g(x_1) + g(x_2) = 0 + h(x_2)$ so $y \in \mathrm{Im}(h)$. Since h is an isomorphism, so is $\psi = h^{-1}$, and it is clear that $g \circ \psi = \mathrm{id}$. □

Exercise 3.4.1. What can you say about a "very short" exact sequence $0 \to L \to M \to 0$?

3.5 Free modules

Two fundamental notions of linear algebra that we have not yet extended to modules are *basis* and *dimension*.

Definition 3.5.1. *A* basis *for an* R*-module* M *is a set* $(e_i)_{i \in \Lambda} \subseteq M$ *such that any* $x \in M$ *can be written uniquely as* $x = \sum_{i \in \Lambda} x_i e_i$*, where* $x_i \in R$ *and only finitely many* x_i *are non-zero. A module is called* free *if it has a basis.*

We give some examples.

- The module R^n is a free R-module with basis $e_1 = (1, 0, \dots, 0), \dots, e_n = (0, \dots, 0, 1)$.

- Conversely, if M has a finite basis, then $M \simeq R^n$ for some n. Indeed, if $e_1, \dots, e_n$ is a basis, then $(x_1, \dots, x_n) \mapsto x_1 e_1 + \dots + x_n e_n$ gives an isomorphism $R^n \to M$. More generally, if two R-modules have bases with the same cardinality, then they are isomorphic.

- If X is any set, we define the *free R-module generated by X* to be the module of finite sums $\sum_{x \in X} r_x x$, where $r_x \in R$ and all but finitely many r_x are non-zero. Such sums can be added and multiplied by scalars in an obvious way. This is a free module with basis X. An equivalent definition is obtained by considering $x \mapsto r_x$ as a function on X. In this language, the free module generated by X is the space of all functions $f : X \to R$ such that $f(x) = 0$ for all but finitely many x.

- The cyclic group $\mathbb{Z}_n$ is not free as a $\mathbb{Z}$-module. Indeed, if e is an element in a basis then $0 = ne = 0e$ are two distinct expressions for the element $0 \in \mathbb{Z}_n$ as a $\mathbb{Z}$-linear combination of basis elements.

- Any vector space has a basis. This statement is equivalent to the axiom of choice, which means that it may be impossible to give a basis explicitly. Nobody can write down a basis for, say, the real vector space of continuous functions on $[0, 1]$, or for $\mathbb{R}$ considered as a vector space over $\mathbb{Q}$.

An important difference between vector spaces and general R-modules is that a module can have two bases with different cardinalities. We will give an example of this phenomenon.

Let V be the vector space of real sequences $(x_1, x_2, \dots)$ with finitely many non-zero entries, and let $R = \mathrm{End}_{\mathbb{R}}(V)$. Explicitly, R is the ring of real matrices $(a_{jk})_{j,k=1}^{\infty}$ such that only finitely many elements in each column are non-zero. Consider R as an R-module. As for any non-zero ring, the identity element forms a basis of R with one element. Now consider the two matrices

$$f_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & 0 & \\ 0 & 0 & 0 & 0 & 1 & \\ \vdots & & & & & \ddots \end{bmatrix}, \qquad f_2 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 1 & 0 & 0 & \\ 0 & 0 & 0 & 0 & 0 & 1 & \\ \vdots & & & & & & \ddots \end{bmatrix}.$$

It is easy to see that any $A \in R$ can be written uniquely as $A = A_1 f_1 + A_2 f_2$, where A_1 and A_2 are obtained from A by deleting all even and odd columns, respectively. Thus, f_1 and f_2 form a basis for R with two elements. For this ring, $R \simeq R^2$ as

R-modules, an explicit isomorphism being $A \mapsto (A_1, A_2)$. By a variation of the same argument, $R \simeq R^n$ for any n.

For modules over non-zero[1] *commutative* rings, one cannot construct such examples. Any two bases for a free module over a non-zero commutative ring have the same cardinality. We will be content with proving this for finitely generated modules.

Proposition 3.5.2. *Any two bases for a finitely generated free module M over a non-zero commutative ring R have the same number of elements. This number is called the* rank *(in the case of vector spaces,* dimension*) of the module.*

Proof. First note that any basis of a finitely generated free module is finite. To see this, let $x_1, \ldots, x_n$ be generators and $(e_j)_{j\in\Lambda}$ a basis. Expressing each x_j in terms of the basis elements, all coefficients outside a finite subset of Λ vanish. Consequently, taking $x \in M$ arbitrary, x can be expressed as a combination of basis elements indexed by a finite set. By the definition of basis, this set must be all of Λ.

Assume now that we have two bases $(e_j)_{j=1}^m$ and $(f_k)_{k=1}^n$ for M. We can write

$$e_j = \sum_{k=1}^{n} a_{jk} f_k, \qquad f_j = \sum_{k=1}^{m} b_{jk} e_k, \quad a_{jk},\ b_{jk} \in R.$$

If we plug one of these expressions into the other one, we find that the matrices $A = (a_{jk})$ and $B = (b_{jk})$ necessarily satisfy $AB = I_m$ and $BA = I_n$, where I_k is the identity matrix of order k. We want to prove that $m = n$. For a contradiction, assume $m > n$. We then create $(m \times m)$-matrices A' and B' by filling out the matrices A and B by zeroes, that is,

$$A' = \begin{bmatrix} A & 0 \end{bmatrix}, \qquad B' = \begin{bmatrix} B \\ 0 \end{bmatrix}.$$

It is now easy to compute

$$A'B' = I_m, \qquad B'A' = \begin{bmatrix} I_n & 0 \\ 0 & 0 \end{bmatrix}.$$

Taking determinants we get

$$\det(A')\det(B') = 1, \qquad \det(B')\det(A') = 0,$$

which is a contradiction as $1 \neq 0$ in a non-zero ring. □

[1] Any module over the zero ring is isomorphic to the zero module, which has two bases with different cardinalities, namely, $\emptyset$ and $\{0\}$.

In the proof we used that

$$\det(AB) = \det(A)\det(B) \tag{3.4}$$

for matrices over commutative rings. Indeed, looking at the proof for real matrices found in any linear algebra textbook, it should be clear that it only uses general properties of commutative rings. It is also possible to deduce the identity from the fact that it holds for, say, real numbers. As this is a very general and useful technique, we sketch the argument. We consider $\det(AB) - \det(A)\det(B) = p(a_{11}, \ldots, a_{nn}, b_{11}, \ldots, b_{nn})$ as a polynomial in all the matrix elements of A and B. We know that p is identically zero as a function on $\mathbb{R}^{2n^2}$. As it is a polynomial function, it follows that it is the zero polynomial. That is, $p = 0$ as an element of $\mathbb{Z}[x_1, \ldots, x_{2n^2}]$. But if $a_1, \ldots, a_N$ are any elements of a commutative ring, then $p \mapsto p(a_1, \ldots, a_N)$ defines a ring homomorphism $\mathbb{Z}[x_1, \ldots, x_N] \to R$. In particular, if p is the zero polynomial then $p(a_1, \ldots, a_N) = 0$. Thus, any polynomial identity that holds for the real numbers holds in any commutative ring. This is not true for non-commutative rings; $xy - yx$ is the zero polynomial in $\mathbb{Z}[x, y]$ but $ab - ba = 0$ does not hold as a general identity in rings.

The following result is used in the proof of Lemma 5.2.5 below.

Lemma 3.5.3. *If M is a module over a ring R and $L \subseteq M$ is a submodule such that M/L is free, then L is a direct summand of M.*

Proof. Consider the exact sequence (3.3). We will apply criterion (B) in Lemma 3.4.2. Take a basis $(e_j)_{j\in\Lambda}$ for M/L and pick for each j a representative $f_j \in M$ of the coset e_j. Define $\psi : M/L \to M$ by

$$\psi\left(\sum_j x_j e_j\right) = \sum_j x_j f_j, \qquad x_j \in R.$$

It is then easy to check that ψ satisfies the appropriate condition. □

Exercise 3.5.1. Show that $\mathbb{Z}[x]$ is a free module over $\mathbb{Z}[x^3]$. Calculate its rank.

Exercise 3.5.2. Show that $\mathbb{Q}$ is not a free module over $\mathbb{Z}$.

Exercise 3.5.3. Assume for simplicity that R is commutative. If V is a free R-module with a finite basis $(e_i)_{i=1}^n$, show that the dual module V^* is free with basis vectors $(e_i^*)_{i=1}^n$ determined by $e_i^*(e_j) = \delta_{ij}$. Morover, show that if V is a free module with an infinite basis $(e_i)_{i\in\Lambda}$, then the dual basis vectors $(e_i^*)_{i\in\Lambda}$ do *not* form a basis for V^*, unless the ring of scalars is the zero ring.

Exercise 3.5.4. The elements $e_1, \ldots, e_n$ in an R-module V are called *R-linearly independent* if

$$x_1 e_1 + \cdots + x_n e_n = 0 \quad \Rightarrow \quad x_1 = \cdots = x_n = 0,$$

where $x_i \in R$. Show that $e_1, \ldots, e_n$ are a basis for V if and only if they generate V and are R-linearly independent.

Exercise 3.5.5. Give an explicit counterexample to the following false statement: "If V is a free R-module of rank n and $e_1, \ldots, e_n$ are R-linearly independent, then $e_1, \ldots, e_n$ form a basis for V".

Exercise 3.5.6. If M and N are finitely generated free modules of rank m and n, respectively, show that $M \oplus N$ is again free with rank $m + n$.

Exercise 3.5.7. Let R be the ring of infinite matrices defined in the text. Considering R as a module over itself, give a basis with infinitely many elements.

3.6 Associative algebras

Associative algebras play a relatively minor role in this text, but it will be useful to know the basic definitions.

If R is a commutative ring, an *associative algebra* A over R is a ring that is also an R-module, such that the multiplication $A \times A \to A$ is R-bilinear. Since the multiplication is automatically additive (the distributive laws) it is enough to postulate

$$r(ab) = (ra)b = a(rb), \qquad r \in R, \quad a, b \in A.$$

For the remainder of this section, we will understand "algebra" to mean associative algebra over a commutative ring R.

By our definition of a ring, algebras have a multiplicative unit. We stress that not all authors make this assumption. We also mention that to define algebras over non-commutative rings is more complicated as one needs to distinguish scalar multiplication from the left and right.

Note that we are considering three multiplications at once: the algebra multiplication $A \times A \to A$, the ring multiplication $R \times R \to R$ and the scalar multiplication $R \times A \to A$. They are all denoted by concatenation of symbols or, occasionally, by a dot. Fortunately, if an expression is ambiguous it is independent of the interpretation. For instance, it does not matter if

$$rsab, \qquad r, s \in R, \quad a, b \in A \tag{3.5}$$

is interpreted as $((rs)(ab))$, as $((rs)a)b$ or in any other way, as the axioms guarantee that the result is the same.

Algebras may seem complicated but you have already seen many examples.

- The polynomial ring $R[x]$ is an algebra with scalar multiplication

$$b(a_0 + \cdots + a_n x^n) = ba_0 + \cdots + ba_n x^n, \qquad b,\, a_j \in R.$$

  More generally, $R[x_1, \ldots, x_n]$ is an algebra.

- If V is an R-module over a commutative ring R, then $\mathrm{End}_R(V) = \mathrm{Hom}_R(V, V)$ is an algebra, with the operations

$$(f+g)(x) = f(x) + g(x), \quad (fg)(x) = f(g(x)), \quad (rf)(x) = r \cdot f(x),$$

  where $x \in V$, $r \in R$ and $f \in \mathrm{End}_R(V)$.

- The ring of $(n \times n)$-matrices with coefficients in a ring R is an algebra with usual addition, matrix multiplication and multiplication of a matrix by a scalar. If R is commutative it can be identified with $\mathrm{End}_R(R^n)$.

Homomorphisms and isomorphisms of algebras are defined in an obvious way. If A is an algebra and $I \subseteq A$ is an ideal (in the usual sense of ring theory), then I is automatically an R-submodule since $rx = (r1_A)x \in I$ for all $r \in R$ and $x \in I$. Then, the quotient A/I is an algebra in a natural way. It is also straight-forward to define direct sums of algebras.

3.7 Additional exercises

Exercise 3.7.1. Show that, if M and N are submodules of the same module, then $M + N = \{m + n;\, m \in M,\ n \in N\}$ and $M \cap N$ are modules, and $(M + N)/M \simeq N/(M \cap N)$.

Exercise 3.7.2. Let $L \subseteq M \subseteq N$ be three modules. Show that M/L is a submodule of N/L and that $(N/L)/(M/L) \simeq N/M$.

Exercise 3.7.3. Let I be an ideal in R, M be an R-module and let IM denote the set of finite sums $a_1x_1 + \cdots + a_nx_n$, where $a_j \in I$ and $x_j \in M$. Prove that M/IM is a module over R/I with the natural scalar product $(r+I)(m+IM) = rm+IM$.

Exercise 3.7.4. Let M be an R-module. Suppose $P_1, \ldots, P_n \in \mathrm{End}_R(M)$ satisfy $P_1 + \cdots + P_n = \mathrm{Id}_M$ and $P_jP_k = \delta_{jk}P_j$. Show that

$$M = \mathrm{Im}(P_1) \oplus \cdots \oplus \mathrm{Im}(P_n).$$

Conversely, if $M = M_1 \oplus \cdots \oplus M_n$, show that there exist P_j as above with $M_j = \mathrm{Im}(P_j)$.

Exercise 3.7.5. A module P is called *projective* if, given a homomorphism $f : P \to N$ and a surjective homomorphism $g : M \to N$, there exists a homomorphism $h : P \to M$ with $f = g \circ h$. (It is understood that all modules are over the same ring R and homomorphism means R-module homomorphism.) Show that a free module is projective.

Exercise 3.7.6. Let R be a commutative ring and $I \subseteq R$ a non-principal ideal. Show that when considered as R-modules, R is free but the submodule I is not free.

Exercise 3.7.7. Let R be a commutative ring containing a zero-divisor a. Show that the ideal aR is not free as an R-module.[2]

Exercise 3.7.8. Let $R = \mathbb{R}[x_1, x_2, \dots]$ be as in Exercise 2.3.4 and let I be the ideal consisting of polynomials with zero constant term. Show that R is finitely generated as an R-module, but the submodule I is not finitely generated.

[2]This and the previous exercise show that if a non-zero commutative ring R is such that the submodule of any free module is free, then R is a PID. The converse is also true, see Proposition 5.2.2 below for the case of finitely generated modules. For general modules one needs the axiom of choice.

Chapter 4

Tensor products

4.1 Tensor products: three definitions

Tensor product is a fundamental operation on vector spaces and, more generally, modules over commutative rings. The idea behind tensor products is to put multilinear and linear maps on equal footing. Let us start with the simplest non-trivial case: bilinear forms on $\mathbb{R}^2$. Let $f : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ be such a map. Bilinearity means that

$$f(au+bv, w) = af(u, w)+bf(v, w), \qquad f(u, av+bw) = af(u, v)+bf(u, w), \quad (4.1)$$

where a, $b \in \mathbb{R}$ and u, v, $w \in \mathbb{R}^2$. Let $e_1 = (1, 0)$, $e_2 = (0, 1)$ be standard basis vectors. It follows from (4.1) that

$$\begin{aligned} &f(a_1e_1 + a_2e_2, b_1e_1 + b_2e_2) \\ &\qquad = a_1b_1f(e_1, e_1) + a_1b_2f(e_1, e_2) + a_2b_1f(e_2, e_1) + a_2b_2f(e_2, e_2). \end{aligned} \quad (4.2)$$

Thus, f is uniquely determined by the four values $f(e_i, e_j)$. This is reminiscent of *linear* forms on a four-dimensional space. Let us introduce such a space, denoted $\mathbb{R}^2 \otimes \mathbb{R}^2$, with basis vectors denoted

$$e_1 \otimes e_1, \quad e_1 \otimes e_2, \quad e_2 \otimes e_1, \quad e_2 \otimes e_2.$$

(At this point, the symbol "$\otimes$" has no independent meaning; we could as well call the vectors e_{11}, e_{12} and so on.) There is then a bijection between *bilinear* forms $f : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ and *linear* forms $g : \mathbb{R}^2 \otimes \mathbb{R}^2 \to \mathbb{R}$, given by

$$f(e_i, e_j) = g(e_i \otimes e_j), \qquad i, j = 1, 2. \quad (4.3)$$

This example suggests the following definition. If U is a vector space with basis $(e_j)_{j\in\Lambda}$ and V a vector space with basis $(f_j)_{j\in\Gamma}$, then $U \otimes V$ is the vector space with basis vectors denoted $(e_i \otimes f_j)_{i\in\Lambda,\, j\in\Gamma}$. This definition is perfectly fine, and is actually used in some textbooks. However, it has three problems:

- It is not so elegant to have the definition depend on a choice of basis for the vector spaces. This is no big problem, as it is easy to describe the effects of a change of basis.
- For general vector spaces, the existence of a basis depends on the axiom of choice. Many mathematicians prefer a constructive definition when possible.
- The definition extends to free modules, but not to modules in general. This is too restrictive for many purposes.

To obtain a better definition, we try to make sense of (4.3) when e_j are not assumed to be elements of a fixed basis. That is, we want to write

$$f(u, v) = g(u \otimes v) \tag{4.4}$$

for general vectors, or elements of a module. In general, let U, V and W be modules over a commutative ring R.[1] We want to define an R-module $U \otimes V$ so that (4.4) gives a bijection between R-bilinear maps $f : U \times V \to W$ and R-linear maps (that is, R-module homomorphisms) $g : U \otimes V \to W$. Here, R-bilinear means that

$$f(a_1u_1 + a_2u_2, v) = a_1f(u_1, v) + a_2f(u_2, v), \tag{4.5a}$$
$$f(u, a_1v_1 + a_2v_2) = a_1f(u, v_1) + a_2f(u, v_2), \tag{4.5b}$$

where $a_j \in R$, u, $u_j \in U$ and v, $v_j \in V$. We get an important hint on how to do this by choosing $W = U \otimes V$ and $g = \mathrm{Id}$. In this case, $f(u, v) = u \otimes v$, so (4.5) reads

$$(a_1u_1 + a_2u_2) \otimes v = a_1(u_1 \otimes v) + a_2(u_2 \otimes v), \tag{4.6a}$$
$$u \otimes (a_1v_1 + a_2v_2) = a_1(u \otimes v_1) + a_2(u \otimes v_2). \tag{4.6b}$$

Apparently, these relations must be valid in our module $U \otimes V$.

It turns out that the relations (4.6) are all that we need. We define the tensor product $U \otimes V$ to be the R-module spanned by all expressions $u \otimes v$, $u \in U$, $v \in V$, subject to the relations (4.6). In more formal language, let F be the free module generated by $U \times V$, that is, the module consisting of finite sums $\sum_i a_i(u_i, v_i)$, $a_i \in R$, $u_i \in U$, $v_i \in V$. Let S be the submodule of F spanned by all elements of the form

$$(a_1u_1 + a_2u_2, v) - a_1(u_1, v) - a_2(u_2, v), \tag{4.7a}$$
$$(u, a_1v_1 + a_2v_2) - a_1(u, v_1) - a_2(u, v_2), \tag{4.7b}$$

[1] We will not actually use commutativity, but as indicated in Exercises 4.4.7–4.4.8 the definition that will follow is adapted to the commutative case. To obtain useful tensor products over non-commutative rings one should consider *bi-modules* with scalar multiplication from both the left and the right.

where again $a_j \in R$, u, $u_j \in U$ and v, $v_j \in V$. Then we define $U \otimes V = F/S$ and write the coset $(u, v) + S$ as $u \otimes v$.[2] To stress the dependence on the ring of scalars one often writes $U \otimes_R V$.

A word of warning: The tensor product $U \otimes V$ does *not* consist only of elements $u \otimes v$, where $u \in U$ and $v \in V$. In general, it consists of finite sums of such *pure tensors*.

Another word of warning: Do not confuse tensor product with direct product. The vector space $\mathbb{R}^m \times \mathbb{R}^n = \mathbb{R}^m \oplus \mathbb{R}^n$ has dimension $m + n$; the vector space $\mathbb{R}^m \otimes \mathbb{R}^n$ has dimension mn.

Let us now show that our tensor product does the intended job, namely, that R-bilinear maps $f : U \times V \to W$ can be identified with R-linear maps $g : U \otimes V \to W$. The identification should have the form (4.4), that is, $f = g \circ \pi$, where $\pi : U \times V \to U \otimes V$ is the bilinear map $\pi(u, v) = u \otimes v$. We need to show that for any f there is a unique g with this property. By construction, any element $x \in U \otimes V$ can be written as a finite sum $x = \sum_i u_i \otimes v_i$. Since we require g to be R-linear, (4.4) gives $g(x) = \sum_i f(u_i, v_i)$. This shows that g is unique. To show that it is well-defined on $U \otimes V$ we must show that it preserves the relations (4.6). That follows from f being bilinear.

The correspondence between bilinear maps on $U \times V$ and linear maps on $U \otimes V$ is referred to as the *universal property* of the tensor product.

Definition 4.1.1. *Let U, V and T be R-modules and let $\pi : U \times V \to T$ be a bilinear map. We say that π has the universal property for tensor products if, for any R-module W and R-bilinear map $f : U \times V \to W$, there is a unique R-linear map $g : T \to W$ such that $f = g \circ \pi$.*

It is standard abuse of terminology to say that T has the universal property. As we have seen, $U \otimes V$ has the universal property. By the following result, any module with the universal property can be identified with $U \otimes V$.

Proposition 4.1.2. *If $\pi_1 : U \times V \to T_1$ and $\pi_2 : U \times V \to T_2$ both have the universal property, then there is a unique isomorphism of R-modules $\phi : T_1 \to T_2$ such that $\phi \circ \pi_1 = \pi_2$.*

Proof. By Definition 4.1.1 with (π, f) replaced by (π_1, π_2), there is a unique R-linear map $\phi : T_1 \to T_2$ with $\pi_2 = \phi \circ \pi_1$. It remains to prove that ϕ is bijective. To this end, we reverse the role of π_1 and π_2, finding a homomorphism $\psi : T_2 \to T_1$ with $\pi_1 = \psi \circ \pi_2$. We then have two R-linear maps id and $\psi \circ \phi$ from T_1 to T_1 such that $\pi_1 = \mathrm{id} \circ \pi_1 = (\psi \circ \phi) \circ \pi_1$. Applying the uniqueness part of Definition 4.1.1 to $f = \pi_1$, we find that $\psi \circ \phi = \mathrm{id}$. Again reversing the role of π_1 and π_2, we have $\phi \circ \psi = \mathrm{id}$. Thus, ϕ is an isomorphism with inverse ψ. □

[2] My apologies to those readers whose reaction is "Why didn't you just say this from the start?".

Proposition 4.1.2 might seem rather abstract, but it is actually quite useful. For instance, it implies the following important fact, which shows that our first attempt at a definition leads to the same result when it makes sense.

Proposition 4.1.3. *Let U and V be free modules with bases $(e_i)_{i\in\Lambda}$ and $(f_i)_{i\in\Gamma}$, respectively. Then, $U \otimes V$ is free with basis $(e_i \otimes f_j)_{i\in\Lambda,\, j\in\Gamma}$.*

Proof. Let T be the free module with basis $e_i \otimes f_j$, which is now just a notation for some elements labelled by $\Lambda \times \Gamma$. If $u \in U$ and $v \in V$ are arbitrary, then we can write $u = \sum_i r_i e_i$ and $v = \sum_j s_j f_j$ for unique r_i, $s_j \in R$ (all but finitely many being zero). Thus, we can define $\pi : U \times V \to T$ by $\pi(u, v) = \sum_{i,j} r_i s_j (e_i \otimes f_j)$. It is easy to check that π is bilinear. We will show that it has the universal property. Let $h : U \times V \to W$ be bilinear. That $h = g \circ \pi$ implies that $g(e_i \otimes f_j) = h(e_i, f_j)$. If g is an R-module homomorphism, it is then uniquely determined on T by $g(\sum_{i,j} r_{ij} e_i \otimes f_j) = \sum_{i,j} r_{ij} h(e_i, f_j)$, $r_{ij} \in R$. We must show that $h = g \circ \pi$ holds on the whole space $U \times V$. Indeed, with u and v as above we have $h(u, v) = \sum_{i,j} r_i s_j h(e_i, f_j) = \sum_{i,j} r_i s_j g(e_i \otimes f_j) = (g \circ \pi)(u, v)$. □

To summarize, we have indicated three definitions of the tensor product $U \otimes V$. The first one, given in terms of basis elements, only works for free modules (and, in particular, for vector spaces). The second one, as a quotient of the free vector space spanned by $U \times V$, works for any modules (but is of interest mainly for modules over commutative rings). Third, we could very abstractly define "tensor product" as "an object that has the universal property for tensor products". This still calls for a proof that an object with the universal property exists, for which one has to go back to one of the first two definitions (or perhaps some other construction).

Exercise 4.1.1. Let $x = (2, 1, 0) \otimes (0, 1, 3) + (0, 2, 0) \otimes (1, 2, 0) \in \mathbb{R}^3 \otimes \mathbb{R}^3$. Express x in the standard basis $(e_i \otimes e_j)_{i,j=1}^3$ (where $e_1 = (1, 0, 0)$ and so on).

Exercise 4.1.2. If $f : U_1 \to V_1$ and $g : U_2 \to V_2$ are R-module homomorphisms, show that there is a naturally defined R-module homomorphism $f \otimes g : U_1 \otimes U_2 \to V_1 \otimes V_2$.

4.2 Properties of tensor products

The tensor product satisfies properties such as

$$U \otimes V \simeq V \otimes U, \tag{4.8a}$$

$$(U \otimes V) \otimes W \simeq U \otimes (V \otimes W), \tag{4.8b}$$

$$U \otimes (V \oplus W) \simeq (U \otimes V) \oplus (U \otimes W), \tag{4.8c}$$

where in each case the isomorphism is the obvious one, e.g. $u \otimes v \mapsto v \otimes u$ in the first case. Each of these statements can be proved either from our definition or using Proposition 4.1.2. For instance, to prove that $\phi(u \otimes v) = v \otimes u$ is an isomorphism $U \otimes V \to V \otimes U$, the first method amounts to observing that applying ϕ to the relations (4.6) gives the same relations with U and V interchanged. For an "abstract nonsense" proof, consider the map $f : U \times V \to V \otimes U$ defined by $f(u, v) = v \otimes u$. Then, f is bilinear, so by the universal property there exists an R-linear map $g : U \otimes V \to V \otimes U$ such that $g(u \otimes v) = v \otimes u$. Interchanging the roles of U and V we find a homomorphism in the other direction, which gives the inverse of g.

Having understood the two-fold tensor product $U \otimes V$, it is no great step to consider multiple tensor products $V_1 \otimes \cdots \otimes V_n$ of R-modules. One can either define it as an iterated two-fold tensor product, for instance, defining $U \otimes V \otimes W$ as either side of (4.8b), or by a straight-forward modification of the definitions in §4.1. For the latter approach, to define $U \otimes V \otimes W$ (4.7) is replaced by three expressions, one of which is

$$(u, a_1v_1 + a_2v_2, w) - a_1(u, v_1, w) - a_2(u, v_2, w).$$

Propositions 4.1.2 and 4.1.3 hold with straight-forward modifications of the statements and proofs. We will use the notation

$$V^{\otimes n} = \underbrace{V \otimes \cdots \otimes V}_{n}.$$

For the reader who is used to tensor products of vector spaces, some aspects of tensor products over rings may seem surprising. As an example, let m and n be relatively prime positive integers. We claim that

$$\mathbb{Z}_m \otimes_{\mathbb{Z}} \mathbb{Z}_n = 0. \tag{4.9}$$

Indeed, if p and q are integers with $mp + nq = 1$, then

$$x \otimes y = (mp + nq)(x \otimes y) = mpx \otimes y + x \otimes nqy = 0$$

for all x and y. The identity (4.9) simply reflects the fact that any $\mathbb{Z}$-bilinear map on $\mathbb{Z}_m \times \mathbb{Z}_n$ is identically zero.

Tensor products can often be thought of as *change of scalars* for modules. In general, let M be an R-module and $\phi : R \to S$ a ring homomorphism. Then, S can be viewed as an R-module with scalar multiplication $rs = \phi(r)s$, $r \in R$, $s \in S$ (see Exercise 3.1.3). In particular, $S \otimes_R M$ is an R-module. More interestingly, $S \otimes_R M$ can be considered as an S-module through $s(t \otimes x) = st \otimes x$. We can think of this S-module as obtained from M by changing the scalars from R to S. The

most familiar example is probably when V is a real vector space; then $\mathbb{C} \otimes_{\mathbb{R}} V$ is the *complexification* of V. One can also consider (4.9) from this perspective. We consider $\mathbb{Z}_n$ as a $\mathbb{Z}$-module and restrict the scalars to the quotient $\mathbb{Z}_m$; this can only be done in a trivial way when $\gcd(m, n) = 1$.

Exercise 4.2.1. If M is an R-module, show that $R \otimes_R M \simeq M$.

Exercise 4.2.2. Show that if M is a free R-module with basis $(e_i)_{i \in \Lambda}$ and R is a subring of S, then the S-module $S \otimes_R M$ is also free with the same basis. In particular, show that the complexification of $\mathbb{R}^n$ can be identified with $\mathbb{C}^n$.

Exercise 4.2.3. Let U and V be finitely generated and free modules over a ring R. Since we assume in this chapter that R is commutative, $U^* = \mathrm{Hom}_R(U, R)$ is again an R-module, see Exercise 3.2.3. Show that $\mathrm{Hom}_R(U, V) \simeq U^* \otimes V$, where $\phi \otimes v \in U^* \otimes V$ corresponds to the homomorphism $u \mapsto \phi(u)v$.

4.3 Symmetric and antisymmetric tensors

We will need some basic facts on symmetric and antisymmetric tensors. For our purposes, it would be enough to take the underlying ring to be $\mathbb{R}$ or $\mathbb{C}$. It is not hard to generalize this to general commutative rings, but some technical problems appear if some elements of the form $1 + \cdots + 1$ do not have multiplicative inverses. For simplicitly, we will assume that these problems do not appear. Thus, throughout this section we assume that R is a commutative ring such that $\mathbb{Q} \subseteq R$. For instance, R could be a field of characteristic zero (see Exercise 2.4.4) or a polynomial ring over such a field.

Let V and W be R-modules and $f : V^n \to W$ be a multilinear map. We say that f is *symmetric* if

$$f(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) = f(x_1, \ldots, x_n), \qquad \sigma \in \mathrm{S}_n, \quad x_1, \ldots, x_n \in V.$$

We call f *antisymmetric* if

$$f(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) = \mathrm{sgn}(\sigma) f(x_1, \ldots, x_n), \qquad \sigma \in \mathrm{S}_n, \quad x_1, \ldots, x_n \in V. \tag{4.10}$$

Here σ denotes the sign of a permutation, see §1.6. By Exercise 1.6.3, an equivalent condition is that f changes sign when two variables are interchanged.

Two familiar examples from linear algebra are the standard scalar product $(u, v) \mapsto u \cdot v$, which is a symmetric bilinear form $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, and the cross product $(u, v) \mapsto u \times v$, which is an antisymmetric bilinear map $\mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$.

As we will consider the two cases in parallel, let $\varepsilon \in \{\mathrm{id}, \mathrm{sgn}\}$ and call a form ε-symmetric if

$$f(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) = \varepsilon(\sigma) f(x_1, \ldots, x_n), \qquad \sigma \in \mathrm{S}_n, \quad x_1, \ldots, x_n \in V.$$

By similar arguments that led us from bilinear forms to the definition of the tensor product $U \otimes V$, but starting from ε-symmetric multilinear forms, one is led to supplement the relations in $V^{\otimes n}$ with

$$x_1 \otimes \cdots \otimes x_n = \varepsilon(\sigma) x_{\sigma(1)} \otimes \cdots \otimes x_{\sigma(n)}, \qquad \sigma \in \mathrm{S}_n. \tag{4.11}$$

It suffices to postulate this for some generating subset of S_n, such as the transpositions. Let us introduce the notation $V_\varepsilon^{\otimes n} = V^{\otimes n}/M_\varepsilon$, where M_ε is the submodule spanned by the elements

$$x_1 \otimes \cdots \otimes x_n - \varepsilon(\sigma) x_{\sigma(1)} \otimes \cdots \otimes x_{\sigma(n)}, \qquad \sigma \in \mathrm{S}_n, \quad x_1, \ldots, x_n \in V.$$

This is non-standard notation that we only introduce in order to discuss the two cases in parallel. The case $\varepsilon = \mathrm{id}$ is called the *symmetric power* and will be denoted $V^{\odot n}$ (another common notation is $\mathrm{S}^n(V)$). The case $\varepsilon = \mathrm{sgn}$ is the *exterior power* $V^{\wedge n}$. The coset containing $x_1 \otimes \cdots \otimes x_n$ is denoted $x_1 \odot \cdots \odot x_n$ and $x_1 \wedge \cdots \wedge x_n$, respectively. To compute with these products one only needs to know that they are both linear in each argument x_i and that the symmetric product is invariant under reordering the x_i, whereas the exterior product changes sign when two of the variables x_i are interchanged. It is also good to know that $x_1 \wedge \cdots \wedge x_n = 0$ if $x_i = x_j$ for some $i \neq j$. Indeed, applying the relation (4.11) with $\varepsilon = \mathrm{sgn}$ and $\sigma = (i\ j)$ gives $x_1 \wedge \cdots \wedge x_n = -x_1 \wedge \cdots \wedge x_n$. Using our assumption that $1 + 1$ is invertible, it follows that $x_1 \wedge \cdots \wedge x_n = 0$.

It is straight-forward to prove that there is a bijection between ε-symmetric multilinear maps $V^n \to W$ and R-linear maps $V_\varepsilon^{\otimes n} \to W$. One can also prove that $V_\varepsilon^{\otimes n}$ is characterized by a corresponding universal property.

It is often useful to consider $V_\varepsilon^{\otimes n}$ as a submodule of $V^{\otimes n}$ rather than a quotient. To this end, let $\pi : \mathrm{S}_n \to \mathrm{End}_R(V^{\otimes n})$ be defined on a spanning set by

$$\pi(\sigma)(v_1 \otimes \cdots \otimes v_n) = v_{\sigma^{-1}(1)} \otimes \cdots \otimes v_{\sigma^{-1}(n)}. \tag{4.12}$$

It is easy to check that

$$\pi(\sigma\tau) = \pi(\sigma)\pi(\tau).$$

When R is a field, this means that π is a representation of S_n. Let

$$N_\varepsilon = \{x \in V^{\otimes n};\ \pi(\sigma)(x) = \varepsilon(\sigma)x,\ \sigma \in \mathrm{S}_n\}.$$

This is the space of ε-symmetric (that is, *symmetric* or *antisymmetric*) *tensors*. Note that M_ε can be defined as the span of all elements of the form

$$x - \varepsilon(\sigma)\pi(\sigma)(x), \qquad x \in V^{\otimes n}, \quad \sigma \in \mathrm{S}_n. \tag{4.13}$$

There is a natural operation on tensors known as symmetrization ($\varepsilon = \mathrm{id}$) or antisymmetrization ($\varepsilon = \sigma$). It is a special case of a construction that will be

central later in the course, namely, averaging over group actions. In the case at hand, we define

$$P = \frac{1}{n!} \sum_{\sigma \in \mathrm{S}_n} \varepsilon(\sigma)\pi(\sigma). \tag{4.14}$$

Note that we are using that $n!$ has a multiplicative inverse in our ring.

Lemma 4.3.1. *The operator P satisfies*

$$\pi(\tau) \circ P = P \circ \pi(\tau) = \varepsilon(\tau)P, \qquad \tau \in \mathrm{S}_n. \tag{4.15}$$

Moreover, $\mathrm{Ker}(P) = M_\varepsilon$, $\mathrm{Im}(P) = N_\varepsilon$ *and* $P^2 = P$.

Proof. We have

$$\pi(\tau) \circ P = \frac{1}{n!} \sum_{\sigma \in \mathrm{S}_n} \varepsilon(\sigma)\pi(\tau\sigma) = \frac{1}{n!} \sum_{\sigma \in \mathrm{S}_n} \varepsilon(\tau^{-1}\sigma)\pi(\sigma) = \varepsilon(\tau)P.$$

Here, we first used that π is a group homomorphism, then we replaced σ by $\tau^{-1}\sigma$ in the summation and finally used that ε is a group homomorphism, with $\varepsilon(\tau^{-1}) = \varepsilon(\tau)$. The identity $P \circ \pi(\tau) = \varepsilon(\tau)P$ is proved in the same way.

To see that $M_\varepsilon \subseteq \mathrm{Ker}(P)$, it suffices to check that $Py = 0$ for all elements y of the form (4.13). This follows from (4.15); namely,

$$Px - P\varepsilon(\sigma)\pi(\sigma)x = Px - \varepsilon(\sigma)^2 Px = 0.$$

For the reverse inclusion, suppose $Px = 0$. Then, we can write

$$x = x - Px = \frac{1}{n!} \sum_{\sigma \in \mathrm{S}_n} (x - \varepsilon(\sigma)\pi(\sigma)x),$$

which is visibly in M_ε.

That $\mathrm{Im}(P) \subseteq N_\varepsilon$ is again an easy consequence of (4.15). To complete the proof, we note that if $x \in N_\varepsilon$ then each term in (4.14) acts trivially on x, so that $Px = x$. This implies that $N_\varepsilon \subseteq \mathrm{Im}(P)$. Moreover, replacing x by Px (which is always in N_ε) we find that $P^2 = P$. □

Applying condition (D) in Lemma 3.4.1 we can deduce the following fact. Alternatively, one can use Lemma 3.4.2 as it follows that

$$0 \to M_\varepsilon \to V^{\otimes n} \xrightarrow{P} N_\varepsilon \to 0$$

is an exact sequence, which is split by the inclusion map $N_\varepsilon \to V^{\otimes n}$.

Proposition 4.3.2. *We have $V^{\otimes n} = M_\varepsilon \oplus N_\varepsilon$. In particular, the space of ε-symmetric tensors N_ε is isomorphic to the ε-symmetric tensor power $V_\varepsilon^{\otimes n} = V^{\otimes n}/M_\varepsilon$.*

We now give the following analogue of Proposition 4.1.3. Although one could give an "abstract nonsense" proof based on appropriate universal properties, we choose to base our proof on Proposition 4.3.2.

Proposition 4.3.3. *Suppose that V is a free module with basis $(e_i)_{i\in\Lambda}$. Let $\leq$ be a total ordering of the index set Λ. Then, $V^{\odot n}$ has a basis consisting of the elements*

$$e_{i_1} \odot \cdots \odot e_{i_n}, \quad i_1 \leq \cdots \leq i_n \tag{4.16a}$$

and $V^{\wedge n}$ a basis consisting of the elements

$$e_{i_1} \wedge \cdots \wedge e_{i_n}, \quad i_1 < \cdots < i_n. \tag{4.16b}$$

Proof. It follows from Proposition 4.1.3 that the elements $e_{i_1} \otimes \cdots \otimes e_{i_n}$ span $V^{\otimes n}$. Using the relations (4.11), each of these elements is ± 1 times a similar element with $i_1 \leq \cdots \leq i_n$. If we are in the alternating case and have equality between two indices, the corresponding element is 0. This proves that the elements (4.16) form a spanning set. To show that they are independent, suppose there is a relation between them, that is,

$$\sum_{i_1 \prec \cdots \prec i_n} a(i_1, \ldots, i_n) e_{i_1} \otimes \cdots \otimes e_{i_n} \in M_\varepsilon \tag{4.17}$$

for some $a(i_1, \ldots, i_n) \in R$. Here, $\prec$ means $\leq$ and $<$, respectively. We need to show that a vanishes identically. Since $M_\varepsilon = \operatorname{Ker}(P)$, (4.17) is equivalent to the relation

$$\begin{aligned} 0 &= n! \sum_{i_1 \prec \cdots \prec i_n} a(i_1, \ldots, i_n) P(e_{i_1} \otimes \cdots \otimes e_{i_n}) \\ &= \sum_{i_1 \prec \cdots \prec i_n} \sum_{\sigma \in \mathrm{S}_n} \varepsilon(\sigma) a(i_1, \ldots, i_n) (e_{i_{\sigma^{-1}(1)}} \otimes \cdots \otimes e_{i_{\sigma^{-1}(n)}}) \\ &= \sum_{j_1, \ldots, j_n} \sum_{\sigma \in \mathrm{S}_n} \varepsilon(\sigma) a(j_{\sigma(1)}, \ldots, j_{\sigma(n)}) (e_{j_1} \otimes \cdots \otimes e_{j_n}). \end{aligned}$$

Since this is an expansion in a basis for $V^{\otimes n}$, it follows that

$$\sum_{\sigma \in \mathrm{S}_n} \varepsilon(\sigma) a(j_{\sigma(1)}, \ldots, j_{\sigma(n)}) = 0$$

for all indices j_k. In the alternating case, only σ with $j_{\sigma(1)} < \cdots < j_{\sigma(n)}$ give a non-zero contribution. Since there is at most one such permutation σ, we conclude that

a vanishes identically. In the symmetric case, there may be several terms. (For instance, if $n = 3$ and $j_1 = j_2 < j_3$ then both $\sigma = \mathrm{id}$ and $\sigma = (1\ 2)$ contribute.) However, they all lead to the same value for $\varepsilon(\sigma)a(j_{\sigma(1)}, \dots, j_{\sigma(n)})$, so a must vanish identically. □

In the finite rank case, it is a standard exercise in combinatorics to count the elements in (4.16).

Corollary 4.3.4. *If V is a free module with rank d, then $V^{\odot n}$ has rank $\binom{d+n-1}{n}$. If $0 \le n \le d$, then $V^{\wedge n}$ has rank $\binom{d}{n}$. If $n > d$, then $V^{\wedge n} = \{0\}$.*

It seems interesting that if V has rank n, then $V^{\wedge n}$ has rank 1. Let us have a closer look at this case. Let $e_1, \dots, e_n$ be a basis for V and consider a general exterior product $u_1 \wedge \cdots \wedge u_n$, where $u_i = \sum_j a_{ij} e_j$, $a_{ij} \in R$. If we use the multilinearity to expand

$$u_1 \wedge \cdots \wedge u_n = (a_{11}e_1 + \cdots + a_{1n}e_n) \wedge \cdots \wedge (a_{n1}e_1 + \cdots + a_{nn}e_n)$$

we get n^n terms, but they vanish if the same basis vector e_j appears twice. Thus, we only get a non-zero result if in the j-th factor we choose the term $a_{j,\sigma(j)}e_{\sigma(j)}$, for some $\sigma \in \mathrm{S}_n$. That is,

$$\begin{aligned} u_1 \wedge \cdots \wedge u_n &= \sum_{\sigma \in \mathrm{S}_n} \prod_{j=1}^{n} a_{j,\sigma(j)}\, e_{\sigma(1)} \wedge \cdots \wedge e_{\sigma(n)} = \sum_{\sigma \in \mathrm{S}_n} \operatorname{sgn}(\sigma) \prod_{j=1}^{n} a_{j,\sigma(j)}\, e_1 \wedge \cdots \wedge e_n \\ &= \det(A) e_1 \wedge \cdots \wedge e_n, \end{aligned} \tag{4.18}$$

where $A = (a_{ij})_{1 \le i,j \le n}$. We have rediscovered the determinant! Up to multiplication by a constant, the determinant is the only n-linear antisymmetric form on an n-dimensional space.

Let us conclude with a remark that will be followed up in the last part of the course. The module $V^{\otimes n}$ contains the two submodules N_{id} and N_{sgn}, isomorphic to $V^{\odot n}$ and $V^{\wedge n}$, respectively. If V is free with rank d, then $V^{\otimes n}$ has rank d^n, whereas the rank of the two submodules add up to $\binom{d+n-1}{n} + \binom{d}{n}$. If $n = 2$, we have $\binom{d+1}{2} + \binom{d}{2} = d^2$, which corresponds to the fact that any bilinear form f can be written uniquely

$$f(x,y) = \frac{f(x,y) + f(y,x)}{2} + \frac{f(x,y) - f(y,x)}{2}$$

as the sum of a symmetric and an antisymmetric form. If $n = 3$, we have $\binom{d+2}{3} + \binom{d}{3} = d(d^2+2)/3 \approx d^3/3$ for large d. Thus, the symmetric and antisymmetric forms only account for about one third of all trilinear forms. As we will see later in the course, the remaining forms correspond to a two-dimensional representation of S_3 (id and sgn are one-dimensional representations). Understanding general n-linear forms requires a good understanding of the representation theory of S_n.

Exercise 4.3.1. Show that f is antisymmetric as defined in the text if and only if $f(x_1, \ldots, x_n) = 0$ whenever $x_i = x_j$ for some indices with $i \neq j$.[3]

Exercise 4.3.2. Express $u \wedge v \wedge w$ in the standard basis for $(\mathbb{R}^4)^{\wedge 3}$, where $u = (1, 0, 0, 1)$, $v = (1, 1, 1, 0)$ and $w = (0, 1, 1, -1)$.

Exercise 4.3.3. Find a general expression for the exterior product $(x_1, x_2, x_3) \wedge (y_1, y_2, y_3)$ on $\mathbb{R}^3$. What is the relation to the "cross product" in linear algebra?

Exercise 4.3.4. Prove directly that $P^2 = P$, where P is given by (4.14).

Exercise 4.3.5. Is it true that

$$(V \wedge V) \wedge (V \wedge V) \simeq V \wedge V \wedge V \wedge V$$

for general R-modules V?

Exercise 4.3.6. Let V be a free module of rank d. Give a natural module isomorphism between $V^{\odot n}$ and the submodule of $R[x_1, \ldots, x_d]$ consisting of homogeneous polynomials of degree n.

Exercise 4.3.7. Let P_1 and P_2 be the two projections obtained by choosing $\varepsilon = \mathrm{id}$ and $\varepsilon = \mathrm{sgn}$ in (4.14). Show that $P_1 P_2 = P_2 P_1 = 0$. Deduce that with $U = \mathrm{Im}(\mathrm{id} - P_1 - P_2)$, one has the decomposition

$$V^{\otimes n} \simeq V^{\odot n} \oplus V^{\wedge n} \oplus U.$$

Exercise 4.3.8. Let $(e_i)_{i=1}^n$ be a basis for a module V and let $u_i = \sum_{j=1}^n a_{ij} e_j$ be arbitrary elements of V. Show that

$$u_1 \wedge \cdots \wedge u_m = \sum_{1 \leq s_1 < \cdots < s_m \leq n} \det_{1 \leq i,j \leq m} (a_{i,s_j})\, e_{s_1} \wedge \cdots \wedge e_{s_n}.$$

4.4 Additional exercises

Exercise 4.4.1. Compute the tensor product (4.9) when m and n are not assumed to be relatively prime.

Exercise 4.4.2. Let V be a module over a commutative ring R. Using the universal property, show that $V \otimes_R R^n \simeq V^n$.

Exercise 4.4.3. Let G be the group $\mathbb{Z}^2 \times \mathbb{Z}_3$. What is the real vector space $\mathbb{R} \otimes_{\mathbb{Z}} G$? (Hint: Use Exercise 4.4.2.)

[3] You need to use that $2 = 1 + 1$ is not a zero-divisor in R. If that would be the case, then one normally uses the definition suggested by this exercise.

Exercise 4.4.4. Let G be an abelian group and let $nG = \{na;\ a \in G\}$ for $n \in \mathbb{Z}$. Show that $\mathbb{Z}_n \otimes_{\mathbb{Z}} G \simeq G/nG$.

Exercise 4.4.5. Let M be an R-module over a commutative ring R and $I \subseteq R$ an ideal. Then, $R/I \otimes_R M$ can be viewed as an R/I-module. Show that it is equivalent to the module M/IM described in Exercise 3.7.3.

Exercise 4.4.6. Let A be an $(m \times m)$-matrix and B an $(n \times n)$-matrix over a commutative ring R. We can consider $A \otimes B$ as an R-module endomorphism of $R^m \otimes R^n$, so it can be identified with an $(mn) \times (mn)$-matrix. Show that $\det(A \otimes B) = \det(A)^n \det(B)^m$. (Hint: Do it first for $B = \mathrm{Id}$.)

Exercise 4.4.7. Show that if f is bilinear in the sense of (4.5), with R not necessarily commutative, then $rsf(u, v) = srf(u, v)$ for all r, $s \in R$. Moreover, show that if we define the tensor product $U \otimes V$ as in the text, then $rsx = srx$ for all r, $s \in R$ and $x \in U \otimes V$.

Exercise 4.4.8. Let R be the ring of (2×2) real matrices considered as a module over itself. Show that matrix multiplication $f(A, B) = AB$ is not bilinear in the sense of (4.5). Which properties of f seem analogous to bilinearity?

Exercise 4.4.9. Is it true or false that

$$\mathrm{Hom}(V \wedge V, W \wedge W) \simeq \mathrm{Hom}(V, W) \wedge \mathrm{Hom}(V, W),$$

where V and W are finite-dimensional complex vector spaces?

Chapter 5

Modules over principal ideal domains

5.1 Statement and overview

We will classify all finitely generated modules over a PID R. One important case is $R = \mathbb{Z}$, when we obtain a classification of finitely generated abelian groups. We will show that any such group is isomorphic to

$$\mathbb{Z}^s \times \mathbb{Z}/p_1^{k_1}\mathbb{Z} \times \cdots \times \mathbb{Z}/p_m^{k_m}\mathbb{Z}, \tag{5.1}$$

where p_j are primes. This is a good example to keep in mind throughout the discussion. Another important case gives the "Jordan normal form" of a complex matrix, which generalizes diagonalization to non-diagonalizable matrices. This is useful for many applications of linear algebra, for instance, to differential equations.

To formulate the main theorem, consider the equivalence relation on prime elements of R, defined by $p \sim q$ if $p = eq$ with e a unit. Let $\mathfrak{P}$ be a set of representatives for these equivalence classes. (For instance, in $\mathbb{Z}$ the equivalence classes are $\{\pm p\}$ with p a prime number. Choosing always the positive representative, we can identify $\mathfrak{P}$ with the set of prime numbers.)

Theorem 5.1.1. *Any finitely generated module over a* PID R *is isomorphic to*

$$R^s \times R/p_1^{k_1}R \times \cdots \times R/p_m^{k_m}R, \qquad p_j \in \mathfrak{P}. \tag{5.2}$$

Moreover, this decomposition is unique up to reordering the factors.

The decomposition (5.2) is called the *primary decomposition* of M (an ideal of the form $p^k R$ with p prime is called a primary ideal). Note that some of the primes p_j in Theorem 5.1.1 may be equal. As an example, if G is a finite abelian group of order 24, then G is a product as in (5.1), where the factor $\mathbb{Z}^s$ is absent ($s = 0$) and where $p_1^{k_1} \cdots p_m^{k_m} = 24 = 2^3 \cdot 3$. Up to reordering the factors, the only possibilities are

$$\mathbb{Z}_8 \times \mathbb{Z}_3, \qquad \mathbb{Z}_2 \times \mathbb{Z}_4 \times \mathbb{Z}_3, \qquad \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_3. \tag{5.3}$$

By the uniqueness part of the statement, these groups are mutually non-isomorphic.

The proof of Theorem 5.1.1 is rather long. The proof of existence of the decomposition (5.2) will be divided into three steps. To illustrate these steps we consider the example

$$M = \mathbb{Z}^2 \times \mathbb{Z}_2 \times \mathbb{Z}_4 \times \mathbb{Z}_3.$$

In the first step, we split M as

$$M = \mathbb{Z}^2 \times (\mathbb{Z}_2 \times \mathbb{Z}_4 \times \mathbb{Z}_3),$$

where the first factor is a free module and the second one a so called torsion module. In the next step, the torsion part is split into factors corresponding to distinct primes. In the example,

$$M = \mathbb{Z}^2 \times (\mathbb{Z}_2 \times \mathbb{Z}_4) \times \mathbb{Z}_3.$$

In the final step, the factor related to each prime p is split into modules of the form R/p^kR; in the example, we arrive at

$$M = \mathbb{Z}^2 \times \mathbb{Z}_2 \times \mathbb{Z}_4 \times \mathbb{Z}_3.$$

Exercise 5.1.1. Write down all abelian groups of order 360.

5.2 Proof of Theorem 5.1.1

A *torsion element* in an R-module M is an element x such that $ax = 0$ for some non-zero $a \in R$. If R is commutative, the torsion elements form a submodule, which we denote M_{tor}. A module with $M_{\text{tor}} = \{0\}$ is called *torsion free* and a module with $M_{\text{tor}} = M$ is called a *torsion module*. It is easy to see that M/M_{tor} is torsion free.

Lemma 5.2.1. *A free module M over an integral domain R is torsion-free.*

Proof. Let $(e_j)_{j\in\Lambda}$ be a basis for M. Suppose that $ax = 0$, where $x = \sum_j x_j e_j$. Then $\sum_j ax_j e_j = 0$ so, by the definition of a basis, $ax_j = 0$ for each j. Since R is a domain we have either $a = 0$ or $x_j = 0$ for all j, that is, $x = 0$. □

In the context of (5.2), Lemma 5.2.1 shows that the factor R^s is torsion free. The remaining factors form a torsion module, since they are annihilated by $p_1^{k_1} \cdots p_m^{k_m}$.

The condition that R is a PID has two important consequences, which makes the classification of modules tractable. Namely, any submodule of a free module is free and any submodule of a finitely generated module is finitely generated. In

fact, for non-zero commutative rings, the first of these properties is equivalent to R being a PID, whereas the second property is equivalent to R being Noetherian (see also Exercises 3.7.6–3.7.8). We will only need the first statement for finitely generated modules.

Proposition 5.2.2. *If F is a finitely generated free module over a* PID *R and M is a submodule, then M is again finitely generated and free. Moreover,* $\operatorname{rank}(M) \leq \operatorname{rank}(F)$.

Proof. Let $e_1, \ldots, e_n$ be a basis for F. Let

$$M_k = M \cap (Re_1 + \cdots + Re_k).$$

We will prove by induction on k that M_k is free and $\operatorname{rank} M_k \leq k$. Our starting point is the trivial case $M_0 = \{0\}$ and the endpoint case $k = n$ is the statement of the theorem.

Consider the map $\pi : M_k \to R$ defined by

$$\pi(x_1e_1 + \cdots + x_ke_k) = x_k.$$

Then π is an R-module homomorphism so $\operatorname{Im}(\pi)$ is an ideal. By the PID property, $\operatorname{Im}(\pi) = aR$ for some $a \in R$. If $a = 0$, clearly $M_k = M_{k-1}$ and we are done.

Assuming from now on that $a \neq 0$, pick $x \in M_k$ with $\pi(x) = a$. We claim that $M_k = M_{k-1} \oplus Rx$ or, equivalently, $M_k = M_{k-1} + Rx$ and $M_{k-1} \cap Rx = \{0\}$. For the first identity, pick any $y \in M_k$. We have $\pi(y) = ab$ for some $b \in R$. Then, $\pi(y - bx) = 0$, so $y - bx \in M_{k-1}$, which gives $y \in M_{k-1} + Rx$. For the second identity, if $rx \in M_{k-1}$ then $0 = \pi(rx) = r\pi(x) = ra$. Since $a \neq 0$ we get $r = 0$ (R is a domain) and thus $rx = 0$. Finally, we note that Rx is a free module with basis x. Otherwise, we would have $rx = 0$ for some $r \neq 0$. By Lemma 5.2.1, that would contradict F being free. Thus, it follows from the induction hypothesis that M_k is free with $\operatorname{rank}(M_k) = \operatorname{rank}(M_{k-1}) + 1 \leq k$. □

Corollary 5.2.3. *Let M be a finitely generated module over a* PID*. Then any submodule of M is finitely generated.*

Proof. Suppose M is generated by $v_1, \ldots, v_n$. Define $f : R^n \to M$ by

$$f(x_1, \ldots, x_n) = x_1v_1 + \cdots + x_nv_n.$$

Then, f is a homomorphism. If N is a submodule of M then $f^{-1}(N) = \{x \in R^n;\ f(x) \in N\}$ is a submodule of R^n. By Prop. 5.2.2, $f^{-1}(N)$ is finitely generated. Acting by f on a set of generators we obtain a finite set of generators for N. □

By Lemma 5.2.1, a free module over an integral domain is torsion free. The converse holds for finitely generated modules over a PID.

Lemma 5.2.4. *If M is a finitely generated module over a* PID *and M is torsion free then M is free.*

Proof. Let $v_1, \ldots, v_n$ be generators for M and let $e_1, \ldots, e_k$ be a maximal set of R-linearly independent elements among these generators. Then, $e_1, \ldots, e_k$ generate a free module $F \subseteq M$. We claim that we can find non-zero elements $a_j \in R$ such that $a_j v_j \in F$. If v_j is one of the generators e_k this is true with $a_j = 1$. Else, $\{v_j, e_1, \ldots, e_k\}$ are linearly dependent, so we can write

$$a_j v_j + x_1 e_1 + \cdots + x_k e_k = 0, \qquad a_j, x_1, \ldots, x_k \in R,$$

where not all the coefficients are zero. Since $e_1, \ldots, e_k$ are linearly independent, $a_j \neq 0$. We now let $a = a_1 \cdots a_n$. Since $a_j v_j \in F$ we have that $a v_j \in F$ for each j. It follows that $f(x) = ax$ is a homomorphism from M to F. Since M is torsion-free, f is injective. Thus, M is isomorphic to a submodule of a finitely generated free module. The conclusion now follows from Proposition 5.2.2. □

We can now complete the first step in the proof of Theorem 5.1.1.

Lemma 5.2.5. *If M is a finitely generated module over a* PID*, then $M \simeq R^s \times M_{\text{tor}}$ for some s.*

Proof. As we remarked above, M/M_{tor} is torsion free. Thus, by Lemma 5.2.4, it is free. By Lemma 3.5.3, M_{tor} is a direct summand of M, so $M = F \oplus M_{\text{tor}}$ for some submodule F. It follows that $F \simeq M/M_{\text{tor}}$, so F is free. Finally, by Cor. 5.2.3, F is finitely generated, so $F \simeq R^s$ for some s. □

We now turn to the second step in the proof, where we split the torsion module into terms corresponding to distinct prime elements.

Lemma 5.2.6. *If M is a finitely generated torsion module over a commutative ring R, then there exists a non-zero element $a \in R$ so that $ax = 0$ for all $x \in M$.*

Proof. Let $v_1, \ldots, v_n$ be generators for M. Since M is a torsion module, $a_j v_j = 0$ for some non-zero $a_j \in R$. We can then take $a = a_1 \cdots a_n$. □

In general, when M is a module and $a \in R$, we write $M_a = \{x \in M;\, ax = 0\}$. Clearly, M_a is a submodule. By Lemma 5.2.6, any finitely generated torsion module has the form $M = M_a$ for some $a \neq 0$. We claim that if the prime factorization of a is $p_1^{k_1} \cdots p_m^{k_m}$, then we can split M as

$$M_a = M_{p_1^{k_1}} \oplus \cdots \oplus M_{p_m^{k_m}}. \tag{5.4}$$

This follows by repeated use of the following result.

Lemma 5.2.7. *If M is a module over a* PID *R and $a, b \in R$ with $\gcd(a, b) = 1$, then*

$$M_{ab} = M_a \oplus M_b.$$

Proof. By Bezout's identity, $1 = as + bt$ for some $s, t \in R$. We need to show that $M_{ab} = M_a + M_b$ and that $M_a \cap M_b = \{0\}$. Both statements follow from writing $x = asx + tbx$. Indeed, if $x \in M_{ab}$, then $asx \in M_b$ and $tbx \in M_a$. Moreover, if $x \in M_b \cap M_c$ then $acx = tbx = 0$. □

As explained above, combining Lemma 5.2.6 and Lemma 5.2.7 gives the following conclusion.

Corollary 5.2.8. *If M is a finitely generated torsion module over a* PID*, then*

$$M = M_{p_1^{k_1}} \oplus \cdots \oplus M_{p_m^{k_m}} \tag{5.5}$$

for certain distinct primes $p_j \in \mathfrak{P}$ and positive integers k_j.

We should think of each summand in (5.5) as gathering all factors in (5.2) that involve a fixed prime. Let us stress that, although our proof of Corollary 5.2.8 is short, it depends crucially on Theorem 2.3.3 (every PID is a UFD).

In the third step of the proof, we decompose the individual terms in (5.5). When S is a subset of an R-module, its *annihilator* is defined by

$$\mathrm{Ann}(S) = \{r \in R;\ rs = 0 \text{ for all } s \in S\}.$$

When R is commutative, $\mathrm{Ann}(S)$ is an ideal (in general, it is a left ideal).

Lemma 5.2.9. *If M is a module over a* PID *R and p is a prime, then $\mathrm{Ann}(M_{p^k}) = p^l R$ for some $l \leq k$.*

Indeed, by the principal ideal property, $\mathrm{Ann}(M_{p^k}) = aR$ for some $a \in R$. Since $p^k \in \mathrm{Ann}(M_{p^k})$, a must be a divisor of p^k.

We may then use the following result to decompose the summands in (5.5).

Lemma 5.2.10. *Let M be a finitely generated module over a* PID *R, such that $\mathrm{Ann}(M) = p^k R$, where $p \in \mathfrak{P}$ and $k \geq 0$. Then, M is isomorphic to a product*

$$R/(p^{l_1}R) \times \cdots \times R/(p^{l_m}R) \tag{5.6}$$

for some positive integers l_j.

We will prove Lemma 5.2.10 by induction on the "size" of M. It is not immediately obvious how the size should be measured, but it turns out that the following idea works. It is easy to see that, for any module M, M/pM is a module over R/pR (cf. Exercise 3.7.3 and Exercise 4.4.5). Note that R/pR is a field, so M/pM is in fact a vector space. Moreover, if M is finitely generated, then the cosets containing the generators generate M/pM. Thus, $|M|_p = \dim_{R/pR}(M/pM)$ is a non-negative integer. We will prove Lemma 5.2.10 by induction on $|M|_p$. It is easy to check that if M is given by (5.6), then $|M|_p = m$ (cf. Lemma 5.2.11 below). Thus, we are actually performing induction over the number of factors in (5.6), but of course that does not make sense until we have proved the lemma.

Proof of Lemma 5.2.10. If M is the zero module, we interpret (5.6) as an empty product. Otherwise, $k \geq 1$ and there is an element $x \in M$ such that $p^{k-1}x \neq 0$. Note that $x \notin pM$ since $x = py$ would give $p^k y \neq 0$. This means in particular that $pM \subsetneq M$ so that $|M|_p \geq 1$. Thus, $|M|_p = 0$ only for the zero module, which we can use as the starting case for the induction.

Assume that the statement of the lemma holds for all modules N with $|N|_p < |M|_p$. Choose x as above and let $N = M/Rx$. We claim that $|N|_p < |M|_p$. To see this, note that the natural projection $M \to N/pN$ maps pM to 0, so there is a surjective homomorphism $\phi : M/pM \to N/pN$. As we have observed above, $x \notin pM$, so x is a non-trivial element in $\mathrm{Ker}(\phi)$. By the dimension theorem, it follows that $|M|_p = |N|_p + \dim_{R/pR}(\mathrm{Ker}(\phi)) > |N|_p$.

It now follows from our induction hypothesis that N is isomorphic to a product like (5.6). We will apply condition (B) in Lemma 3.4.2 to the sequence

$$0 \to Rx \to M \to N \to 0. \tag{5.7}$$

If this condition holds, then $M \simeq N \times Rx$. Since, by Lemma 3.3.1, $Rx \simeq R/p^k R$, this would complete the proof. Using again Lemma 3.3.1, if N has the form (5.6) then we can write

$$N = R\bar{x}_1 \oplus \cdots \oplus R\bar{x}_n, \tag{5.8}$$

where $\bar{x}_j \in M/Rx$ are such that $\mathrm{Ann}(\bar{x}_j) = p^{l_j}R$ (we write $\bar{x}_j$ for cosets). To split the sequence, we need to find representatives y_j for the coset $\bar{x}_j$ so that

$$\psi(a_1\bar{x}_1 + \cdots + a_m\bar{x}_m) = a_1 y_1 + \cdots a_m y_m \tag{5.9}$$

is a well-defined homomorphism from N to M. The only potential problem with this definition is that the coefficients $a_j \in R$ are not uniquely determined by $\sum a_j\bar{x}_j$. If $\sum a_j\bar{x}_j = \sum_j b_j\bar{x}_j$ then, since the sum (5.8) is direct, $b_j - a_j \in \mathrm{Ann}(\bar{x}_j) = p^{l_j}R$ for each j and thus $b_j = a_j + p^{l_j}r_j$ for some $r_j \in R$. The right-hand side of (5.9) is invariant under $a_j \mapsto b_j$ provided that $p^{l_j}y_j = 0$. If we can find such representatives y_j for $\bar{x}_j$, then ψ splits the sequence and the proof is complete.

Let x_j be arbitrary representatives of $\bar{x}_j$. We need to find $y_j \in \bar{x}_j = x_j + Rx$ with $p^{l_j} y_j = 0$. Since we know that $p^{l_j} x_j \in Rx$, we can write $p^{l_j} x_j = p^s cx$, where $\gcd(c, p) = 1$. If $s \geq l_j$, we have $p^{l_j}(x_j - p^{s-l_j} cx) = 0$, so $y_j = x_j - p^{s-l_j} cx$ works. Since $p^k x = 0$, we can assume that $s \leq k$. If $s = k$, then $y_j = x_j$ works. The only case that remains is when $s < l_j$ and $s < k$. If this is so, then we can write $p^{k-1} cx = p^{k-1-s} p^s cx = p^k p^{l_j - s - 1} x_j = 0$ since $p^k \in \mathrm{Ann}(M)$. By Bezout's identity, we have $tc + up = 1$ for some t, $u \in R$. This gives $0 = p^{k-1} tcx = p^{k-1} x - up^k x = p^{k-1} x$, which contradicts our choice of x. This completes the proof. □

Together, Lemma 5.2.5, Corollary 5.2.8, Lemma 5.2.9 and Lemma 5.2.10 prove the existence part of Theorem 5.1.1. For the uniqueness part, the following result is useful.

Lemma 5.2.11. *If R is a* PID*, p and q are prime elements and $M = R/q^k R$, then*

$$p^l M / p^{l+1} M \simeq \begin{cases} R/pR, & p \sim q \text{ and } l \leq k-1, \\ 0, & \text{else} \end{cases}$$

as R-modules (and hence also as R/pR-vector spaces).

Proof. By Lemma 3.3.1, we can write $M = Rx$ with $\mathrm{Ann}(x) = q^k R$. Then, $p^l M$ is generated by $p^l x$ and $N = p^l M / p^{l+1} M$ is generated by the coset $y = p^l x + p^{l+1} M$. Clearly $py = 0$ so, again by Lemma 3.3.1, we have either $N \simeq R/pR$ (when $p^l x \notin p^{l+1} M$ so that $y \neq 0$) or $N = \{0\}$ (when $p^l x \in p^{l+1} M$). If $p = q$ and $l \geq k$, then $p^l x = 0$ so we are in the second case. If $p = q$ and $l \leq k-1$, then $p^l x = p^{l+1} z$ would give $p^{k-1} x = p^k z = 0$, which contradicts $\mathrm{Ann}(x) = p^k R$. Then we are in the first case. Finally, if $p \not\sim q$ then $1 = tq^k + up^{l+1}$ for some t and u. Then, $z = up^{l+1} z \in p^{l+1} M$ for any $z \in M$, so we are in the second case. □

We can now prove the uniqueness part of Theorem 5.1.1. We need to show that if M denotes the module (5.2), then s, p_j and k_j can be constructed uniquely from M (up to reordering). It is clear that $s = \mathrm{rank}_R(M/M_{\mathrm{tor}})$ so it is enough to consider the torsion part. If M is a torsion module (that is, there is no factor R^s in (5.2)) let us compute the numbers

$$d_l(p) = \dim_{R/pR}(p^l M / p^{l+1} M) \tag{5.10}$$

for $p \in \mathfrak{P}$ and $l \geq 0$. We can compute $d_l(p)$ by adding up the contribution from each factor (cf. Exercise 3.5.6). Thus, by Lemma 5.2.11, $d_l(p)$ is the number of indices j such that $p_j = p$ and $l \leq k_j - 1$. Consequently,

$$d_{l-1}(p) - d_l(p) \tag{5.11}$$

is the number of indices j such that $p_j = p$ and $k_j = l$. This shows that p_j and k_j are uniquely determined by M. The proof of Theorem 5.1.1 is finally complete.

Exercise 5.2.1. Give an example of a torsion module over $\mathbb{Z}$ with trivial annihilator.

Exercise 5.2.2. Give the primary decomposition of $\mathbb{Z}_{21}^*$ (the group of integers relatively prime to 21 under multiplication mod 21. Also find an explicit isomorphism to that additive group. How many such "logarithms mod 21" are there?

Exercise 5.2.3. In Exercise 5.1.1 you were asked to list all abelian groups of order 360. Compute $\dim_{\mathbb{Z}_2}(2^l G/2^{l+1}G)$ for each such group and each non-negative integer l.

5.3 The invariant factor decomposition

It is often useful to rewrite the decomposition (5.2) in the following alternative form.

Theorem 5.3.1. *Any finitely generated module M over a* PID *R is isomorphic to*

$$R^s \times R/q_1R \times R/q_2R \times \cdots \times R/q_mR, \tag{5.12}$$

where q_j are non-zero non-units of R and $q_1|q_2|\cdots|q_m$. Moreover, this decomposition is unique up to multiplying q_j by units.

The elements q_j are called *invariant factors* and (5.12) the *invariant factor decomposition*. In particular, a finitely generated torsion module is isomorphic to (5.12), where the factor R^s is absent ($s = 0$). In this case, $\mathrm{Ann}(M) = q_mR$.

To prove Theorem 5.3.1 we need the following fact.

Lemma 5.3.2. *If R is a* PID *and $a, b \in R$ with $\gcd(a, b) = 1$, then we have the isomorphism of R-modules*

$$R/abR \simeq R/aR \times R/bR.$$

This looks very similar to Exercise 2.4.6, but neither result is a logical consequence of the other. If you have solved that exercise, you probably showed that the map $\phi(x+abR) = (x+aR, x+bR)$ is a ring isomorphism. Lemma 5.3.2 then follows from the observation that ϕ is also an R-module homomorphism. Alternatively, it can be obtained as the special case $M = R/abR$ of Lemma 5.2.7.

By iteration, it follows from Lemma 5.3.2 that if a has prime factorization $p_1^{k_1} \cdots p_m^{k_m}$, then

$$R/aR \simeq R/p_1^{k_1}R \times \cdots \times R/p_m^{k_m}R \tag{5.13}$$

as R-modules.

To prove the existence part of Theorem 5.3.1, we write M as in (5.2). We may assume that $s = 0$. We then rearrange the product in a rectangular array so that all factors involving a fixed prime are placed in the same row, with the entries in each row ordered by increasing exponent of p. Moreover, the rows are aligned to the right so that the right-most entry of each row ends up in the same column. We then define q_k as the product of all prime powers in the k-th column.

An example should clarify the definition of q_k. If

$$R = \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_{16} \times \mathbb{Z}_3 \times \mathbb{Z}_5 \times \mathbb{Z}_{25},$$

then we write

$$\begin{aligned} R = \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_{16} \\ \times \mathbb{Z}_3 \\ \times \mathbb{Z}_5 \times \mathbb{Z}_{25}. \end{aligned}$$

In this case $q_1 = 2$, $q_2 = 2 \cdot 5 = 10$ and $q_3 = 16 \cdot 3 \cdot 25 = 1200$. Thus, the invariant factor decomposition of the same group is

$$R = \mathbb{Z}_2 \times \mathbb{Z}_{10} \times \mathbb{Z}_{1200}.$$

Returning to the general situation, it is obvious from the construction that $q_k \mid q_{k+1}$. Moreover, by (5.13), each R/q_kR is isomorphic to the direct product of those factors in (5.2) that were put in the k-th column. This shows the existence part.

To prove uniqueness, we reverse the argument. Given a factorization (5.12) with $q_j \mid q_{j+1}$, we can factorize q_j into primes and construct a rectangular array of modules R/p^kR as above. This leads to a decomposition of M as in (5.2). Since that factorization is unique, so is the factorization (5.12).

Exercise 5.3.1. In Exercise 5.1.1 you were asked to give the primary decomposition of all abelian groups of order 360. Give the invariant factor decomposition of these groups.

5.4 Canonical forms

We now come to an important application of the theory described above: canonical forms of matrices. Any linear operator on a finite-dimensional vector space can be represented by a matrix, which expresses how it acts on a basis. However, changing the basis leads to a different matrix. We would like to find a basis such that the matrix has an especially simple form. Moreover, the form should be canonical in

the sense that this matrix is unique (possibly up to reordering the corresponding basis vectors). As a consequence, if we want to know whether two matrices are *similar* in the sense that they express the same linear map in different bases, we can check whether their canonical forms agree or not. One familiar example is diagonalization, but not any matrix can be diagonalized. The Jordan canonical form explained below can be viewed as an analogue of diagonalization for arbitrary complex matrices.

Let K be a field, V an n-dimensional vector space over K and $A \in \mathrm{End}_K(V)$, that is, A is a linear map from V to itself. The main idea is to study A by viewing V as a module over $R = K[x]$. The module structure is obtained from the ring homomorphism $\Phi : K[x] \to \mathrm{End}_K(V)$ given by $\Phi(p) = p(A)$. More explicitly,

$$\Phi(k_0 + k_1 x + \cdots + k_m x^m)v = k_0 v + k_1 Av + k_2 A^2 v + \cdots + k_m A^m v, \quad k_i \in K, \quad v \in V.$$

It is natural to ask what properties of A are reflected by the module structure of V. The following simple lemma answers that question.

Lemma 5.4.1. *If V_A denotes the $K[x]$-module associated to $A \in \mathrm{End}_K(V)$, then $V_A \simeq V_B$ if and only if A and B are similar over K, that is, $A = T^{-1}BT$ for some invertible element $T \in \mathrm{End}_K(V)$.*

The proof is trivial: a $K[x]$-module isomorphism $T : V_A \to V_B$ is an invertible element of $\mathrm{End}_K(V)$ such that $p(A) = T^{-1}p(B)T$ for all polynomials p. With $p(x) = x$, this gives $A = T^{-1}BT$. Conversely, if $A = T^{-1}BT$, then $A^k = T^{-1}B^kT$ and consequently $p(A) = T^{-1}p(B)T$ for all polynomials p.

Let us now start investigating V as a $K[x]$-module. We know that R is a PID and V is finitely generated. Note also that, as vector spaces over K, $K[x]$ is infinite-dimensional and $\mathrm{End}_K(V)$ is n^2-dimensional. It follows that $\mathrm{Ker}(\Phi) = \mathrm{Ann}(V)$ is non-zero. In particular, V is a torsion module.

We can now apply Theorem 5.1.1 and Theorem 5.3.1. We have

$$V \simeq R/q_1R \times \cdots \times R/q_mR \tag{5.14}$$

for some polynomials q_j. We may either choose $q_j = p_j^{k_j}$ with p_j irreducible polynomials or q_j as non-constant polynomials with $q_1|q_2|\cdots|q_m$. The units in R are the non-zero constant polynomials, so if we normalize q_j to be monic (leading coefficient 1) then they are unique (in the first case up to reordering).

If ϕ is an R-module isomorphism from V to the right-hand side of (5.14), then $\phi(Av) = x\phi(v)$. That is, to understand how A acts on the left we must understand how multiplication by x acts on the right. Consider first the action on R/qR, where $q(x) = a_0 + a_1x + \cdots + a_{k-1}x^{k-1} + x^k$ is a monic polynomial of degree k. We choose 1, x, $x^2, \ldots, x^{k-1}$ as a basis for R/qR. Then, multiplication by x maps each basis

element to the next, except that x^{k-1} is mapped to $x^k = -a_0 - a_1 x - \cdots - a_{k-1}x^{k-1}$. Thus, the matrix for multiplication by x on R/qR is given by

$$M_q = \begin{bmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & & 0 & -a_1 \\ 0 & 1 & & 0 & -a_2 \\ \vdots & & & & \vdots \\ 0 & 0 & \cdots & 1 & -a_{k-1} \end{bmatrix}, \tag{5.15}$$

the so called *companion matrix* to q. We can then interpret (5.14) as saying that there is a basis for V where A is expressed by the block matrix

$$\begin{bmatrix} M_{q_1} & 0 & \cdots & 0 \\ 0 & M_{q_2} & & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & M_{q_m} \end{bmatrix}. \tag{5.16}$$

In the first case ($q_j = p_j^{k_j}$), we call (5.16) the *primary canonical form* and in the second case the *rational canonical form*. The primary canonical form is unique up to reordering the blocks whereas the rational canonical form is unique. We remark that, by expanding (5.15) along the right column, it is easy to see that $\det(xI - M_q) = q(x)$. Thus, the characteristic polynomial for A is in both cases given by $\det(xI - A) = q_1 \cdots q_m$.

One important difference between these two forms is that the primary canonical form depends heavily on the base field. Given, say, a matrix with integer entries, the primary canonical form will typically look different if we work over $\mathbb{Q}$, $\mathbb{R}$ or $\mathbb{C}$. This is because these fields have different irreducible polyomials: $x^2 + 1$ is irreducible over $\mathbb{R}$ but not over $\mathbb{C}$ and $x^2 - 2$ is irreducible over $\mathbb{Q}$ but not over $\mathbb{R}$. By contrast, the rational canonical form does not change if we extend the field. This follows from uniqueness; if $K \subseteq K'$ the canonical form over K is also a canonical form over K'. One interesting consequence is that the notion of similarity does not depend on the field. For instance, if A and B are integer matrices such that $A = TBT^{-1}$ for a complex matrix T, then there is such a matrix T with rational entries.

Let us look more closely at the primary decomposition in the case when K is algebraically closed. (In fact, it is enough to assume that K contains all eigenvalues of A.) For instance, we can take $K = \mathbb{C}$. Then, any monic irreducible polynomial has the form $x - \lambda$ for some $\lambda \in K$. Consider the matrix for multiplication by x in the module R/qR, with $q = (x - \lambda)^k$. Instead of choosing the basis vectors x^j as above, it is nicer to work with $(x - \lambda)^j$. Then, $x(x - \lambda)^j = \lambda(x - \lambda)^j + (x - \lambda)^{j+1}$,

so we obtain instead of M_q the matrix

$$J_q = \begin{bmatrix} \lambda & 0 & 0 & \cdots & 0 & 0 \\ 1 & \lambda & 0 & & 0 & 0 \\ 0 & 1 & \lambda & & 0 & 0 \\ \vdots & & & & & \vdots \\ 0 & 0 & 0 & \cdots & 1 & \lambda \end{bmatrix}. \tag{5.17}$$

In the corresponding basis, A takes the form

$$\begin{bmatrix} J_{q_1} & 0 & \cdots & 0 \\ 0 & J_{q_2} & & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & J_{q_m} \end{bmatrix}. \tag{5.18}$$

This is called the *Jordan canonical form*. It is unique up to reordering the blocks.

In all but the very simplest cases, it makes little sense to compute canonical forms by hand. Therefore, we will not focus on how that can be done, except for a brief discussion on the case $K = \mathbb{C}$. Note that $\mathbb{C}[x]/(x-\lambda) \simeq \mathbb{C}$, so with $p = x - \lambda$ the numbers (5.10) are given by

$$\begin{aligned} d_l &= \operatorname{rank}((A-\lambda I)^l) - \operatorname{rank}((A-\lambda I)^{l+1}) \\ &= \dim \operatorname{Ker}((A-\lambda I)^{l+1}) - \dim \operatorname{Ker}((A-\lambda I)^l). \end{aligned}$$

It is easy to understand directly why the primary canonical form is determined by the solutions to $(A-\lambda I)^l v = 0$. If J is the block (5.17), then we have e.g.

$$(J-\lambda I)^2 = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & & 0 \\ 1 & 0 & 0 & & 0 \\ 0 & 1 & 0 & & 0 \\ 0 & 0 & 1 & & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}. \tag{5.19}$$

If this matrix has size $k \geq 2$, it is visibly of rank $k-2$ and thus the nullspace $\operatorname{Ker}((J-\lambda I)^2)$ has dimension 2. If $k = 1$, then $(J-\lambda I)^2 = 0$ and the nullspace has dimension 1. In general, $(J-\lambda I)^k$ has only zeroes except for a diagonal of 1:s, which is pushed to the southwest as k increases. We find that $\dim \operatorname{Ker}((J-\lambda I)^k) = \min(k, m)$, where m is the size of the block J. We obtain $\dim \operatorname{Ker}((A-\lambda I)^k)$ by adding these numbers for all Jordan blocks corresponding to the eigenvalue λ.

As an example, suppose we are given a complex 6×6-matrix A. We compute the characteristic polynomial $\det(xI - A) = (x-1)^4(x-2)^2$. We start with the eigenvalue $x = 1$ and look for eigenvectors. Suppose that the equation $(A - I)v = 0$ has a two-dimensional space of solutions. This means that there are exactly two Jordan blocks with diagonal entries 1. The sum of their dimension must be equal to 4, so the blocks have size $(3, 1)$ or $(2, 2)$. Next, we look at the equation $(A-I)^2v = 0$. In the first case, the solution space would be 3-dimensional ($\min(2,3) + \min(2,1) = 2 + 1 = 3$) and in the second case 4-dimensional. Let us say that we are in the first case. We then turn to the eigenvalue 2 and find that the space of eigenvectors is two-dimensional. There are then two blocks with diagonal entries 2, which necessarily have size 1. We conclude that the Jordan canonical form for A is

$$\begin{bmatrix} 1 & & & & & \\ 1 & 1 & & & & \\ & 1 & 1 & & & \\ & & & 1 & & \\ & & & & 2 & \\ & & & & & 2 \end{bmatrix},$$

where all missing entries are zero. The prime powers corresponding to the Jordan blocks are $(x-1)^3 = x^3 - 3x^2 + 3x - 1$, $(x-1)$, $(x-2)$ and $(x-2)$, so the primary canonical form is

$$\begin{bmatrix} & & 1 & & & \\ 1 & & -3 & & & \\ & 1 & 3 & & & \\ & & & 1 & & \\ & & & & 2 & \\ & & & & & 2 \end{bmatrix}.$$

To get the rational canonical form we arrange the prime powers in an array as explained in §5.3. We get

$$\begin{matrix} (x-1) & (x-1)^3 \\ (x-2) & (x-2) \end{matrix}$$

and can read of the invariant factors along the columns:

$$\begin{aligned} c_1 &= (x-1)(x-2) = x^2 - 3x + 2, \\ c_2 &= (x-1)^3(x-2) = x^4 - 5x^3 + 9x^2 - 7x + 2. \end{aligned}$$

Note that c_1c_2 is indeed the characteristic polynomial. This gives the rational

canonical form

$$\begin{bmatrix} & -2 & & & & \\ 1 & 3 & & & & \\ & & & & -2 & \\ & & 1 & & 7 & \\ & & & 1 & -9 & \\ & & & & 1 & 5 \end{bmatrix}.$$

As a final remark, note that since $\mathrm{Ann}(M)$ is a non-zero ideal in R, it is equal to qR for a unique monic polynomial q, called the *minimal polynomial* of A. Equivalently, q is the smallest degree monic polynomial such that $q(A) = 0$. In terms of the rational canonical form, it is clear that $q = q_m$, the largest invariant factor of M. As was remarked above, the characteristic polynomial for A is $p = q_1 \cdots q_m$. This implies the following useful result.

Theorem 5.4.2 (Cayley–Hamilton Theorem)**.** *If p is the characteristic polynomial of a square matrix A, then $p(A) = 0$.*

Exercise 5.4.1. Let $A = \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]$. Find the primary and rational canonical forms for A, over the fields $\mathbb{Q}$ and $\mathbb{R}$.

Exercise 5.4.2. A complex matrix A has characteristic polynomial $(x-1)^3(x+1)^2$. Write down all possible Jordan canonical forms for A and give the minimal polynomial in each case.

Exercise 5.4.3. Verify Cayley–Hamilton's theorem by hand for 2×2-matrices.

5.5 Systems of differential equations

As an application of the Jordan canonical form, let us have a brief look at systems of first order ordinary differential equations

$$\begin{cases} x_1' = a_{11}x_1 + \cdots + a_{1n}x_n, \\ x_2' = a_{21}x_1 + \cdots + a_{2n}x_n, \\ \cdots \\ x_n' = a_{n1}x_1 + \cdots + a_{nn}x_n, \end{cases} \tag{5.20}$$

where a_{ij} are real (or complex) numbers. Any higher order system can be reduced to this form by a standard trick. For instance, the equation

$$y'' + ay' + by = 0$$

can be written as a first order system in the variables $(x_1, x_2) = (y, y')$, namely,

$$\begin{cases} x_1' = x_2, \\ x_2' = -bx_1 - ax_2. \end{cases}$$

We can write (5.20) in the matrix form $x' = Ax$, where $A = (a_{ij})$ is the coefficient matrix and $x = (x_1, \ldots, x_n)$ is viewed as a column vector with derivative $x' = (x_1', \ldots, x_n')$. Suppose A has Jordan canonical form $J = TAT^{-1}$. Then, $y = Tx$ satisfies the simpler equation $y' = Jy$. Clearly, this system splits into one independent system for each Jordan block. Thus, the general system (5.20) can be reduced to the case when the coefficent matrix is a Jordan block.

The 1×1 Jordan blocks correspond to the equation $y' = \lambda y$, with solutions $y(t) = Ce^{\lambda t}$. For diagonalizable matrices, no other solutions appear; each x_j solving (5.20) is a linear combination of the solutions $y_j(t) = C_j e^{\lambda_j t}$ to the corresponding diagonal system.

The 2×2 Jordan blocks correspond to

$$\begin{cases} y_1' = \lambda y_1, \\ y_2' = y_1 + \lambda y_2. \end{cases}$$

Here, the first equation has solutions $y_1(t) = Ce^{\lambda t}$. Plugging this into the second equation, one finds that $y_2(t) = (Ct + D)e^{\lambda t}$. More generally, $k \times k$ Jordan blocks lead to solutions of the form $p(t)e^{\lambda t}$, where p is a polynomial of degree at most $k - 1$.

To be more precise, one can show that the system (5.20) has the general solution $x(t) = e^{tA}x(0)$, where

$$e^{tA} = \sum_{m=0}^{\infty} \frac{t^m}{m!} A^m.$$

(This series converges for any complex square matrix A.) For a $k \times k$ Jordan block J with eigenvalue λ, we see as in (5.19) that $(J - \lambda I)^k = 0$. It follows that

$$\begin{aligned} e^{tJ} = e^{\lambda t} e^{(J-\lambda I)t} &= e^{\lambda t} \sum_{m=1}^{k-1} \frac{t^m}{m!} (J - \lambda I)^m \\ &= e^{\lambda t} \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ t & 1 & 0 & & 0 \\ t^2/2 & t & 1 & & 0 \\ \vdots & & & & \vdots \\ \frac{t^{k-1}}{(k-1)!} & & & \cdots & 1 \end{bmatrix}. \end{aligned}$$

The general solution of the system $y' = Jy$ is obtained by multiplying this matrix with an arbitrary vector of initial values. Of course, this can be verified by a direct computation.

Note that the eigenvalues λ may be complex even when A is a real matrix. To get solutions in real form, exponentials coming from the eigenvalues λ and $\bar{\lambda}$ should be rewritten in terms of trigonometric functions. It is then clear that non-diagonalizability of A corresponds to resonance phenomena. Namely, oscillating solutions like $\cos(\lambda t)$ are replaced with unbounded solutions like $t\cos(\lambda t)$.

5.6 Additional exercises

Exercise 5.6.1. Suppose the matrix A has minimal polynomial $p(x) = (x-1)^2$. Find an explicit expression for the integer powers A^k as linear combinations of A and I. If you put $k = 1/2$ in this identity, do you get a solution to $B^2 = A$?

Exercise 5.6.2. Find all abelian groups of order 162. Give both the primary decomposition and the invariant factor decomposition of each group.

Exercise 5.6.3. Let

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

Find the following canonical forms of A: the Jordan canonical form, the primary canonical form over $\mathbb{Q}$ and the rational canonical form.

Exercise 5.6.4. Let

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

Find the Jordan canonical form and the minimal polynomial of A.

Exercise 5.6.5. Let $J(a)$ be the Jordan block matrix $\begin{bmatrix} a & 0 \\ 1 & a \end{bmatrix}$. Consider $A = J(a) \otimes J(b)$ as an operator on $\mathbb{C}^2 \otimes \mathbb{C}^2$. Compute the Jordan canonical form of A. (The answer will look different depending on how many of the numbers a and b are zero.)

Exercise 5.6.6. An evil giant is thinking about an abelian group G of order 144. You are allowed to ask two questions of the form: "What is the dimension of $p^l G/p^{l+1}G$, considered as a vector space over $\mathbb{Z}/p\mathbb{Z}$?", where you may choose p

and l as you like. The giant will answer thruthfully, but if you cannot guess the group after two questions, she will kill you. Is there any way that you can be sure of surviving?

Exercise 5.6.7. Show that any real 2×2-matrix is conjugate over $\mathbb{R}$ to exactly one of the matrices $\begin{bmatrix} x & 0 \\ 0 & y \end{bmatrix}$, $\begin{bmatrix} x & y \\ -y & x \end{bmatrix}$ $(y \neq 0)$, $\begin{bmatrix} x & 0 \\ 1 & x \end{bmatrix}$.

Exercise 5.6.8. Show that the two matrices

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} -1 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & -3 & 1 \end{bmatrix}$$

are conjugate over $\mathbb{Q}$.

Exercise 5.6.9. Show that a complex square matrix is similar to its transpose.

Exercise 5.6.10. Show that a complex matrix A satisfying $A^k = I$ for some k is diagonalizable. Give two proofs, one conceptual by noting that $x^k - 1$ has no multiple roots; one more computational by considering powers of Jordan blocks.

Chapter 6

Group representations

6.1 Fundamental facts

Recall from §1.5 that a representation of a group G on a vector space V is a group homomorphism $\pi : G \to \mathrm{GL}(V)$. More concretely, to each $a \in G$ we associate an invertible linear map $\pi(a)$ on V, such that $\pi(ab) = \pi(a) \circ \pi(b)$ for all $a, b \in G$. It follows that $\pi(g^{-1}) = \pi(g)^{-1}$ and $\pi(1) = \mathrm{Id}_V$. By standard abuse of terminology, we will refer to the space V as a representation. We will often write $\pi = \pi_V$.

Throughout this chapter, G is a finite group and V a finite-dimensional complex vector space. In this case, the theory of group representations is particularly elegant and complete, but we stress that it excludes many situations of great interest.

The most basic problem of representation theory is to classify all representations (up to a natural notion of equivalence). As we will see, in our setting this question has a very satisfactory answer. First of all, any representation splits uniquely as a direct sum of so called irreducible representations. These irreducible representations are enumerated by the conjugacy classes of G; in particular, there are only finitely many. Finally, using so called characters we can in principle decompose any representation explicitly as a sum of its irreducible components.

In order to talk about equivalent representations, we need the following notion.

Definition 6.1.1. *Let V and W be two representations of G. Then, a linear map $\phi : V \to W$ is called* intertwining[1] *if $\phi \circ \pi_V(g) = \pi_W(g) \circ \phi$ for all $g \in G$. If there exists an invertible intertwining map $V \to W$, then we call V and W* equivalent *and write $V \simeq W$.*

If $V \simeq W$, then the two spaces have the same dimension and can both be identified with $\mathbb{C}^n$. The identity $\pi_V(g) = \phi^{-1} \circ \pi_W(g) \circ \phi$ means that we can make this identification (that is, pick bases for V and W) so that $\pi_V(g)$ and $\pi_W(g)$ are given by the same matrix for all g. Thus, we can in effect identify π_V and π_W.

We will denote the vector space of linear maps $V \to W$ by $\mathrm{Hom}(V, W)$ and the subspace of intertwining maps by $\mathrm{Hom}_G(V, W)$. Note that $\mathrm{Hom}(V, W)$ is itself a

[1] Swedish: sammanflätande.

representation, with

$$\pi_{\mathrm{Hom}(V,W)}(g)(\phi) = \pi_W(g) \circ \phi \circ \pi_V(g^{-1}).$$

When V is a representation, we write V^G for the subspace of invariant elements, that is,

$$V^G = \{v \in V;\ \pi(g)v = v \text{ for all } g \in G\}. \tag{6.1}$$

In this notation, $\mathrm{Hom}_G(V,W) = \mathrm{Hom}(V,W)^G$.

A useful method to obtain invariant elements is by taking averages over the group action.

Lemma 6.1.2. *Let V be a representation of G, and let $v \in V$. Then,*

$$\bar{v} = \frac{1}{|G|} \sum_{g \in G} \pi(g)v \in V^G. \tag{6.2}$$

Moreover, if $v \in V^G$ then $\bar{v} = v$.

Proof. For any $h \in G$,

$$\pi(h)\bar{v} = \frac{1}{|G|} \sum_{g \in G} \pi(hg)v = \frac{1}{|G|} \sum_{g \in G} \pi(g)v = \bar{v},$$

where we replaced g by $h^{-1}g$ in the sum. For the second statement, simply note that if $v \in V^G$ then each term in (6.2) is equal to v. □

Applying the Lemma to $\mathrm{Hom}(V,W)$, we find that for any $\phi \in \mathrm{Hom}(V,W)$, the average

$$\bar{\phi} = \frac{1}{|G|} \sum_{g \in G} \pi_W(g) \circ \phi \circ \pi_V(g^{-1})$$

is an intertwining map.

If V is a representation, a subrepresentation $U \subseteq V$ is a subspace such that $\pi_V(g)(U) \subseteq U$ for all $g \in G$. Equivalently, the restriction of the operators $\pi_V(g)$ to U define a representation of G on U. It is easy to check that the kernel and image of an intertwining map are subrepresentations.

Proposition 6.1.3. *If V is a representation of G and $U \subseteq V$ a subrepresentation, then there exists another subrepresentation W such that $V = U \oplus W$.*

Proof. First take any subspace X such that $V = U \oplus X$ as vector spaces. Let $P : V \to U$ be the corresponding projection, that is, $P(u + x) = u$ for any $u \in U$ and $x \in X$. Then, $\bar{P} : V \to U$ is intertwining. Let $u \in U$ and consider

$$\bar{P}(u) = \frac{1}{|G|} \sum_{g \in G} (\pi(g) \circ P \circ \pi(g^{-1}))(u).$$

Since U is a subrepresentation, we have $\pi(g^{-1})(u) \in U$ and consequently

$$(\pi(g) \circ P \circ \pi(g^{-1})(u) = (\pi(g) \circ \pi(g^{-1}))(u) = u, \qquad g \in G.$$

Hence, $\bar{P}(u) = u$. It follows that $\bar{P}$ is a projection on U, that is, $\mathrm{Im}(\bar{P}) = U$ and $\bar{P}^2 = \bar{P}$. Now let $W = \mathrm{Ker}(\bar{P})$. Then, W is a subrepresentation and (by the last part of Lemma 3.4.1) $V \simeq U \oplus W$. □

Any representation V has the trivial subrepresentations $\{0\}$ and V. We call V *irreducible* if $V \neq \{0\}$ and it has no non-trivial subrepresentations. The set of equivalence classes of irreducible representations of G will be denoted $\mathrm{Irr}(G)$.

Corollary 6.1.4 (Maschke's Theorem)**.** *Any representation is a direct sum of irreducible subrepresentations.*

This follows from Proposition 6.1.3 by iteration, which must terminate as we only consider finite-dimensional representations. In §6.2, we will show that the decomposition into irreducible representations is unique.

Note that, although we assume that the ground field is $\mathbb{C}$, the only thing used so far is that $|G|^{-1}$ exists. In fields of finite characteristic p, such as $\mathbb{Z}_p$, Maschke's theorem does not hold in general. This makes representation theory in finite characteristic (so called *modular* representation theory) more involved than characteristic 0.

Proposition 6.1.5 (Schur's Lemma)**.** *If V and W are irreducible representations and $\phi : V \to W$ is intertwining, then $\phi = 0$ if $V \not\simeq W$ and $\phi = \lambda\,\mathrm{Id}$ if $V = W$.*

Proof. The kernel and image of ϕ are subrepresentations. By irreducibility, they are either $\{0\}$ or the whole space V and W. Thus, either $\phi = 0$ or ϕ is bijective. In the second case, $V \simeq W$ and we can identify V with W as representations. After this identification, $\phi : V \to V$. Let λ be an eigenvalue of ϕ. Then, $0 \neq \mathrm{Ker}(\phi - \lambda\,\mathrm{Id})$. As this is a subrepresentation, $\mathrm{Ker}(\phi - \lambda\,\mathrm{Id}) = V$, that is, $\phi = \lambda\,\mathrm{Id}$. □

We stress that Schur's lemma does not hold over the real numbers, as not all real matrices have a real eigenvalue.

A useful consequence of Schur's lemma is that

$$\dim_{\mathbb{C}} \mathrm{End}_G(V) = \begin{cases} 0, & V = \{0\}, \\ 1, & V \text{ irreducible}, \\ \geq 2, & \text{else} \end{cases} \tag{6.3}$$

(here we write $\mathrm{End}_G(V) = \mathrm{Hom}_G(V, V)$). For the last statement, note that if $V = U \oplus W$ as representations, then the maps $\phi(u+w) = au + bw$, $u \in U$, $w \in W$, $a, b \in \mathbb{C}$, form a two-dimensional subspace of $\mathrm{End}_G(V)$.

For the next result we recall some definitions.

Definition 6.1.6. *An* inner product *on a complex vector space V is a map $V \times V \to \mathbb{C}$, denoted $(u, v) \mapsto \langle u, v\rangle$, which is linear in u, conjugate-symmetric, that is,*

$$\langle v, u\rangle = \overline{\langle u, v\rangle}$$

and positive definite, that is,

$$\langle u, u\rangle > 0 \quad \textit{for} \quad u \neq 0.$$

If V and W are inner product spaces, any $A \in \mathrm{Hom}(V, W)$ has an adjoint $A^* \in \mathrm{Hom}(W, V)$ defined by $\langle Av, w\rangle = \langle v, A^*w\rangle$ for all $v \in V$, $w \in W$. We say that $A : V \to V$ is *unitary* if $A^{-1} = A^*$, or equivalently $\langle v, w\rangle = \langle Av, Aw\rangle$. If $V = W = \mathbb{C}^n$ with the standard inner product

$$\langle x, y\rangle = x_1\bar{y_1} + \cdots + x_n\bar{y_n},$$

then $(A^*)_{ij} = \bar{A}_{ji}$.

Proposition 6.1.7. *If V is a representation of G, then there is a basis for V where all elements of G act by unitary matrices. Equivalently, there is a G-invariant inner product on V, that is, $\langle \pi(g)u, \pi(g)v\rangle = \langle u, v\rangle$ for all $g \in G$ and $u, v \in V$.*

Proof. Start with an arbitrary inner product $\langle u, v\rangle_0$ on V. Then define

$$\langle u, v\rangle = \frac{1}{|G|}\sum_{g\in G}\langle \pi(g)u, \pi(g)v\rangle_0.$$

This is clearly an inner product. If we simultaneously replace u by $\pi(h)u$, v by $\pi(h)v$ and g by gh^{-1}, the right-hand side does not change. Thus, it is G-invariant. □

By the spectral theorem, every unitary matrix can be diagonalized, so we have the following consequence. This also follows immediately from Exercise 5.6.10, see also Exercise 6.4.4.

Corollary 6.1.8. *If V is a representation of G, then the operator $\pi_V(g)$ is diagonalizable for any $g \in G$.*

We conclude with some natural operations on representations. If V and W are representations, then $V \times W$ is a representation with

$$\pi_{V\times W}(g)(v, w) = (\pi_V(g)v, \pi_W(g)w).$$

It is easy to see that $V \times W = V' \oplus W'$, where $V' = V \times \{0\} \simeq V$ and $W' = \{0\} \times W \simeq W$ as representations. Thus, we usually write $V \oplus W$ instead of

$V \times W$. In an appropriate choice of basis, the representation $V \oplus W$ is given by block matrices,

$$\pi_{V\oplus W}(g) = \begin{bmatrix} \pi_V(g) & 0 \\ 0 & \pi_W(g) \end{bmatrix}. \tag{6.4}$$

Direct product should not be confused with tensor product. If V and W are representations, then $V \otimes W$ is a representation with

$$\pi_{V\otimes W}(g)\left(\sum_i v_i \otimes w_i\right) = \sum_i \pi_V(g)v_i \otimes \pi_W(g)w_i.$$

The space $\mathbb{C}$ is a representation with $\pi_{\mathbb{C}}(g) = \mathrm{Id}$ for all $g \in G$. It is called the *trivial representation*. We have already seen that, if V and W are representations, then $\mathrm{Hom}(V, W)$ is a representation. Choosing W as the trivial representation gives the *dual representation* $V^* = \mathrm{Hom}(V, \mathbb{C})$. Explicitly,

$$\pi_{V^*}(g)(\xi) = \xi \circ \pi_V(g^{-1}), \qquad g \in G, \quad \xi \in V^*.$$

Exercise 6.1.1. Show the equivalence of representations $\mathrm{Hom}(V, W) \simeq V^* \otimes W$.

Exercise 6.1.2. Let $U \subseteq V$ be a subrepresentation. Given an invariant inner product on V, show that

$$U^\perp = \{x \in V;\ \langle x, u\rangle = 0 \text{ for all } u \in U\}$$

is a subrepresentation. Use this to give an alternative proof of Proposition 6.1.3 as a consequence of Proposition 6.1.7.

Exercise 6.1.3. Let V be a representation equipped with two invariant scalar products $\langle\cdot,\cdot\rangle_1$ and $\langle\cdot,\cdot\rangle_2$. Show that there is an intertwining map $\Phi : V \to V$ such that $\langle\Phi(u), v\rangle_1 = \langle u, v\rangle_2$ for all u, $v \in V$. Deduce that if V is irreducible, then the invariant scalar product is unique up to multiplication by a non-zero constant.

Exercise 6.1.4. Let V and W be group representations equipped with invariant inner products and let $A \in \mathrm{Hom}_G(V, W)$. Show that $A^* \in \mathrm{Hom}_G(W, V)$.

Exercise 6.1.5. In this problem, we forget our standing assumption that groups are finite and vector spaces complex. Let K be a field and define $\phi : K \to \mathrm{GL}(2, K)$ by $\phi(x) = \left[\begin{smallmatrix} 1 & x \\ 0 & 1 \end{smallmatrix}\right]$. Show that ϕ is a representation of K (viewed as an additive group) over the field K. Show that ϕ cannot be decomposed as a sum of irreducible representations. (Choosing K as a finite field shows that Maschke's theorem does not hold for finite groups but arbitrary fields. Choosing $K = \mathbb{C}$ shows that it does not hold for complex representations of infinite groups.)

6.2 Characters

Recall that, if V is a finite-dimensional vector space, then the trace of an operator $\phi \in \mathrm{GL}(V)$ can be defined as $\mathrm{Tr}(\phi) = \sum_{k=1}^{n} a_{kk}$, where $A = (a_{ij})_{i,j=1}^{n}$ is a matrix representing ϕ. This definition makes sense since $\mathrm{Tr}(T^{-1}AT) = \mathrm{Tr}(A)$, so the trace is independent of the choice of basis. Over the complex numbers, $\mathrm{Tr}(\phi)$ can alternatively be defined as the sum of the eigenvalues of ϕ, counted with multiplicity.

The *character* of a representation $\pi : G \to \mathrm{GL}(V)$ is the map $\chi : G \to \mathbb{C}$ defined by $\chi(g) = \mathrm{Tr}(\pi(g))$. We will often write $\chi = \chi_V$. One might think that almost all information is lost when passing from a $\mathrm{GL}(V)$-valued representation to its $\mathbb{C}$-valued character. However, we will see that a representation is uniquely determined by its character. We note right away that the dimension is determined. Namely, $\chi(1) = \mathrm{Tr}(\mathrm{Id}_V) = \dim V$.

The following simple fact will be useful.

Lemma 6.2.1. *If π is a representation of G and $g \in G$, then any eigenvalue λ of $\pi(g)$ is a root of unity.*

Proof. Iterating the eigenvalue equation $\pi(g)v = \lambda v$ gives $\pi(g^n)v = \lambda^n v$. Since $g^n = 1$ for some n, it follows that $\lambda^n = 1$. □

Lemma 6.2.2. *If χ is the character of a representation π, then $\chi(g^{-1}) = \overline{\chi(g)}$ for any $g \in G$.*

Proof. If $\lambda_1, \dots, \lambda_n$ are all the eigenvalues of $\pi(g)$, then $\pi(g^{-1}) = \pi(g)^{-1}$ has eigenvalues $\lambda_1^{-1}, \dots, \lambda_n^{-1}$. It follows that

$$\chi(g) = \lambda_1 + \cdots + \lambda_n, \qquad \chi(g^{-1}) = \lambda_1^{-1} + \cdots + \lambda_n^{-1} = \bar{\lambda}_1 + \cdots + \bar{\lambda}_n, \tag{6.5}$$

where we used Lemma 6.2.1 in the last step. □

Lemma 6.2.3. *When V and W are representations, we have*

$$\chi_{V\oplus W} = \chi_V + \chi_W, \qquad \chi_{V\otimes W} = \chi_V \chi_W,$$
$$\chi_{V^*}(g) = \overline{\chi_V(g)}, \qquad \chi_{\mathrm{Hom}(V,W)}(g) = \overline{\chi_V(g)}\chi_W(g).$$

Proof. The first identity is immediate from (6.4). For the second identity, pick bases (e_i) of V and (f_i) of W and suppose that, with respect to these bases, $\pi_V(g) = (a_{ij})$, $\pi_W(g) = (b_{ij})$. Then,

$$\pi_{V\otimes W}(g)(e_i \otimes f_j) = \sum_{kl} a_{ki} b_{lj} e_k \otimes f_l.$$

Summing up the diagonal coefficients gives

$$\chi_{V\otimes W}(g) = \sum_{kl} a_{kk}b_{ll} = \chi_V(g)\chi_W(g).$$

For the last identity, let (e_i^*) be the dual basis of (e_i). It is easy to check that, in this basis, the matrix representing $\pi_{V^*}(g)$ is the transpose of the matrix representing $\pi_V(g^{-1})$. Thus, $\chi_{V^*}(g) = \chi_V(g^{-1}) = \overline{\chi_V(g)}$ by Lemma 6.2.2. The last identity follows from the previous two using Exercise 6.1.1. □

Recalling the notation (6.1), we clearly have

$$\mathrm{Id}_{V^G} = \frac{1}{|G|}\sum_{g\in G}\pi(g)$$

for any representation V. Taking the trace of both sides gives

$$\dim(V^G) = \frac{1}{|G|}\sum_{g\in G}\chi(g).$$

Applying this to the representation $\mathrm{Hom}(V, W)$, with V and W irreducible, Schur's lemma gives the orthogonality relation

$$\frac{1}{|G|}\sum_{g\in G}\overline{\chi_V(g)}\chi_W(g) = \begin{cases} 1, & V \simeq W, \\ 0, & V \not\simeq W. \end{cases}$$

Let $\mathrm{L}^2(G)$ be the $|G|$-dimensional complex vector space of functions $G \to \mathbb{C}$ equipped with the scalar product

$$\langle \phi, \psi\rangle_{\mathrm{L}^2(G)} = \frac{1}{|G|}\sum_{g\in G}\phi(g)\overline{\psi(g)}.$$

We have then proved the following result, which will be successively strengthened in Theorem 6.4.1 and Theorem 6.8.3.

Lemma 6.2.4. *Characters of irreducible representations are orthonormal in* $\mathrm{L}^2(G)$.

In particular, it follows that there are only finitely many irreducible representations (up to equivalence). Let us denote them $(V_j)_{j=1}^m$. Suppose now that V is an arbitrary representation. By Maschke's theorem, we may write

$$V \simeq \underbrace{V_1 \oplus \cdots \oplus V_1}_{k_1} \oplus \cdots \oplus \underbrace{V_m \oplus \cdots \oplus V_m}_{k_m}; \tag{6.6}$$

that is, V_j appears in V with *multiplicity* k_j. It follows that

$$\chi_V = k_1\chi_{V_1} + \cdots + k_m\chi_{V_m}.$$

We can then compute the multiplicities using Lemma 6.2.4.

Corollary 6.2.5. *Any representation V can be split into irreducible components as in* (6.6)*, where the multiplicities are given by*

$$k_j = \langle \chi_V, \chi_{V_j} \rangle_{\mathrm{L}^2(G)}. \tag{6.7}$$

Moreover,

$$\|\chi_V\|^2_{\mathrm{L}^2(G)} = \langle \chi_V, \chi_V \rangle_{\mathrm{L}^2(G)} = k_1^2 + \cdots + k_m^2. \tag{6.8}$$

This has several important consequences. First, it follows that the representation π_V is uniquely determined by χ_V. Second, it follows that the decomposition (6.6) is unique. Finally, we obtain a very useful irreducibility criterion.

Corollary 6.2.6. *A representation V is irreducible if and only if $\|\chi_V\|_{\mathrm{L}^2(G)} = 1$.*

To give an example, consider the action of the group S_3 on $V = \mathbb{C}^3$ by permuting the coordinates, that is,

$$\pi(\sigma)(x_1, x_2, x_3) = (x_{\sigma^{-1}(1)}, x_{\sigma^{-1}(2)}, x_{\sigma^{-1}(3)}).$$

It is often called the *defining representation*. Equivalently, in terms of the standard basis vectors, $\pi(\sigma)(e_j) = e_{\sigma(j)}$. Thus, the group elements act by the 3×3-permutation matrices

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix},$$
$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix},$$

which are obtained by permuting the columns of the identity matrix. The traces of these elements are, respectively, 3, 1, 1, 1, 0, 0. Thus, $\|\chi_V\|^2 = (3^2 + 1^2 + 1^2 + 1^2)/6 = 2$. Clearly, the only positive integer solution to $k_1^2 + \cdots + k_m^2 = 2$ is $m = 2$, $k_1 = k_2 = 1$. Thus, V must be the sum of two irreducible representations, appearing without multiplicity. Since $\dim(V) = 3$, the only possibility is that $V = V_0 \oplus V_2$ with $\dim(V_0) = 1$ and $\dim(V_2) = 2$.[2] It is easy to see this directly; the spaces are $V_0 = \{(x, x, x);\ x \in \mathbb{C}\}$ and $V_2 = \{(x_1, x_2, x_3) \in \mathbb{C}^3;\ x_1 + x_2 + x_3 = 0\}$. (Note that, in agreement with Exercise 6.1.2, $V_2 = V_0^{\perp}$ with respect to the standard inner product on $\mathbb{C}^3$.) On V_0, all elements of S_3 act as the identity, that is, it is the trivial representation. The space V_2 is often called the *standard representation*. It also exists over the real numbers, that is, S_3 acts on $V_2^{\mathbb{R}} = \{(x_1, x_2, x_3) \in$

[2] We save the notation V_1 for a later purpose.

$\mathbb{R}^3$; $x_1+x_2+x_3=0\}$ by permuting the coordinates. We can embed an equilateral triangle in $V_2^{\mathbb{R}}$ with corners at, say, $(2,-1,-1)$, $(-1,2,-1)$ and $(-1,-1,2)$, so that S_3 acts on the triangle. This takes us back to the example at the very beginning of the present notes!

Exercise 6.2.1. Show that if U is a one-dimensional representation and V an irreducible representation, then $U \otimes V$ is irreducible.

Exercise 6.2.2. Let G be a group such that any element is conjugate to its inverse. (Note that S_n is an example as a permutation and its inverse have the same cycle structure.) Show that $V \simeq V^*$ for any representation V of G.

Exercise 6.2.3. If V and W are any representations, show that $\dim \mathrm{Hom}_G(V,W) = \langle \chi_V, \chi_W \rangle_{\mathrm{L}^2(G)}$. If the two representations decompose into irreducibles as $V = \bigoplus_j k_j V_j$ and $W = \bigoplus_j l_j V_j$, show that $\dim \mathrm{Hom}_G(V,W) = \sum_j k_j l_j$.

Exercise 6.2.4. Let V be a representation of a group G and W a representation of a group H. Show that $V \otimes W$ is a representation of $G \times H$ with $\pi_{V\otimes W}(g,h) = \pi_V(g) \otimes \pi_W(h)$. Show that the map $(V,W) \mapsto V \otimes W$ gives a bijection $\mathrm{Irr}(G) \times \mathrm{Irr}(H) \to \mathrm{Irr}(G \times H)$.

Exercise 6.2.5. If V is a one-dimensional representation, show that $|\chi_V(g)| = 1$ for all g. If V is any representation, show that $|\chi_V(g)| \leq \dim(V)$.

6.3 The regular representation

How can we know that we have found all irreducible representations? It turns out that there is a "master representation" that contains them all, namely, the regular representation that we already introduced in §1.5.

For later purposes, it will be useful to first discuss some more general representations. Let X be a G-set, that is, a set equipped with a homomorphism (or group action) $\phi : G \to \mathrm{S}_X$, which we write as $(\phi(g))(x) = g(x)$. Let V be the complex vector space generated by X. It is the space of formal sums $\sum_{x\in X} a_x x$, $a_x \in \mathbb{C}$, or, equivalently (writing $a_x = a(x)$), the space of complex-valued functions on X. The group action extends linearly to a representation $\pi : G \to \mathrm{End}(V)$. That is, $\pi(g)\left(\sum_x a_x x\right) = \sum_x a_x g(x)$. Equivalently, g acts on functions by $(\pi(g)(f))(h) = f(gh^{-1})$. A representation of this form is called a *permutation representation*.

As an example, consider the action of the symmetric group S_n on $X = \{1,\dots,n\}$. In this case, V can be identified with $\mathbb{C}^n$. The corresponding representation

$$\pi(\sigma)(x_1,\dots,x_n) = (x_{\sigma^{-1}(1)},\dots,x_{\sigma^{-1}(n)})$$

is the *defining representation* of S_n.

In the basis $(x)_{x\in X}$ for V, G acts by the matrices $(\pi(g))_{xy} = \delta_{x,g(y)}$. In particular, the character $\mathrm{Tr}(\pi(g))$ is the number of fix-points $\{x \in X;\, g(x) = x\}$. As this is real-valued, it follows from Lemma 6.2.3 that $V \simeq V^*$ as representations. In §7.3, we will need the following explicit expression for an equivalence. Recall that if V is a finite dimensional vector space with basis $(e_i)_{i=1}^n$, then there is a dual basis $(e_i^*)_{i=1}^n$ for V^* defined by $e_i^*(e_j) = \delta_{ij}$.

Lemma 6.3.1. *If X is a finite G-set, then the map $x \mapsto x^*$ defines an equivalence of representations $V = \mathbb{C}[X] \to V^*$.*

Proof. In general, that $\phi : V \to V^*$ is intertwining means that

$$\pi_{V^*}(g) \circ \phi = \phi \circ \pi_V(g), \qquad g \in G.$$

Applying this to an element $v \in V$ and using that $\pi_{V^*}(g)(\xi) = \xi \circ \pi_V(g^{-1})$ gives

$$\phi(v) \circ \pi_V(g^{-1}) = \phi(\pi_V(g)v), \qquad g \in G,\ v \in V.$$

In the case at hand, it suffices to check this when $v = x \in X$, that is

$$x^* \circ \pi_V(g^{-1}) = (\pi_v(g)x)^*.$$

Here, the left-hand side is the functional $y \mapsto \delta_{x,g^{-1}(y)}$ and the right-hand side the functional $y \mapsto \delta_{g(x),y}$, which are clearly equal. □

In the special case $X = G$, the representation V is called the *regular representation*. Explictly, it is the space $\mathbb{C}[G]$ of formal sums $\sum_{h\in G} a_h h$, equipped with the group action

$$\pi(g)\left(\sum_h a_h h\right) = \sum_h a_h gh.$$

(As a vector space, $V = \mathrm{L}^2(G)$, but we prefer to use the second notation only when it is considered as an inner product space.)

We will denote the character of the regular representation by χ_R. Note that $\chi_R(g)$ is the number of fix-points $\{h \in G;\, gh = h\}$, which is equal to $|G|$ if $g = 1$ and 0 else. If V is an arbitrary representation, it follows that

$$\langle \chi_R, \chi_V \rangle = \frac{1}{|G|}\sum_{g\in G} \chi_R(g)\overline{\chi_V(g)} = \overline{\chi_V(1)} = \dim V.$$

Comparing this with (6.7), we can draw the following conclusion.

Theorem 6.3.2. *Any irreducible representation V appears in the regular representation with multiplicity* $\dim V$.

Since the regular representation has dimension $|G|$ and the sum of $\dim V$ copies of V has dimension $(\dim V)^2$, we have the following important consequence.

Corollary 6.3.3. *We have*

$$|G| = \sum_{V \in \mathrm{Irr}(G)} (\dim V)^2. \tag{6.9}$$

Since the dimension of a representation is often referred to as its *degree*, (6.9) is called the *degree equation*.

Let us return to the example of S_3. At the end of §6.2, we found two irreducible representations: the trivial representation V_0 and the two-dimensional representation V_2. Thus, (6.9) starts as $6 = 1^2 + 2^2 + \cdots$. We see that we are missing just one term, corresponding to another one-dimensional representation. This is the *alternating representation* $\pi(\sigma) = \mathrm{sgn}(\sigma)\,\mathrm{Id}_{\mathbb{C}}$. We will denote it V_1.

Exercise 6.3.1. A permutation matrix is a matrix containing exactly one 1 in each row and column, all other entries being 0. Equivalently, it can be obtained from the identity matrix by permuting the columns. Show that a representation V is a permutation representation if and only if there is a basis for V in which every group element acts by a permutation matrix.

6.4 The character table

Although it follows from Lemma 6.2.4 that a group G can have at most $|G|$ irreducible representations, we still do not know the exact number. This will be deduced from a stronger form of the lemma. To this end, we first note that by the matrix identity $\mathrm{Tr}(ABA^{-1}) = \mathrm{Tr}(B)$, we have $\chi(hgh^{-1}) = \chi(g)$ for any character χ and elements g, $h \in G$. A function $f : G \to \mathbb{C}$ that satisfies $f(hgh^{-1}) = f(g)$ is called a *class function*. Equivalently, a class function is a function that is constant on the conjugacy classes $\{hgh^{-1};\, h \in G\}$.

Theorem 6.4.1. *The irreducible characters form an orthonormal basis for the space of class functions in* $\mathrm{L}^2(G)$*. In particular, the number of irreducible characters is the same as the number of conjugacy classes.*

Proof. We have already proved that the characters are orthonormal (Lemma 6.2.4) and we just observed that they are class functions. All that remains is to prove that they span the space of all class functions. To this end, we let α be a class function orthogonal to all characters and prove that $\alpha = 0$. For such α and

any representation V, let $\phi = \sum_g \overline{\alpha(g)}\pi_V(g) \in \mathrm{End}(V)$. We claim that ϕ is intertwining. Indeed,

$$\begin{aligned}\pi_V(h) \circ \phi \circ \pi_V(h^{-1}) &= \sum_{g\in G} \overline{\alpha(g)}\pi_V(hgh^{-1}) = \sum_{g\in G} \overline{\alpha(h^{-1}gh)}\pi_V(g) \\ &= \sum_{g\in G} \overline{\alpha(g)}\pi_V(g) = \phi.\end{aligned}$$

In the case when V is irreducible, Schur's lemma gives $\phi = \lambda \,\mathrm{Id}$ for some $\lambda \in \mathbb{C}$. But since

$$\mathrm{Tr}(\phi) = \sum_g \overline{\alpha(g)}\chi_V(g) = |G|\langle \chi_V, \alpha\rangle = 0,$$

we must have $\lambda = 0$ and hence $\phi = 0$ for any irreducible representation. By Maschke's theorem, $\phi = 0$ in general. In particular, for the regular representation, $0 = \phi(h) = \sum_g \overline{\alpha(g)}gh$ for all h, which implies $\alpha = 0$. □

It follows that the representation theory of G is encoded by the square table $(\chi(C))_{\chi,C}$, where χ runs over the irreducible characters and C over the conjugacy classes. It is called the *character table*.

Corollary 6.4.2. *The rows and columns of the character table satisfy the orthogonality relations*

$$\sum_C \frac{|C|}{|G|}\,\chi(C)\overline{\psi(C)} = \delta_{\chi,\psi}, \tag{6.10a}$$

$$\sum_\chi \chi(C)\overline{\chi(D)} = \frac{|G|}{|C|}\,\delta_{CD}. \tag{6.10b}$$

The first of these relations is a reformulation of Lemma 6.2.4, using that the characters are constant on the conjugacy classes. It can be written as the unitarity relation $AA^* = \mathrm{Id}$, where $A_{\chi,C} = \sqrt{|C|/|G|}\chi(C)$. It implies $A^*A = \mathrm{Id}$, which is the second relation.

Although the irreducible representations and the conjugacy classes of G are equinumerous, there is in general no natural correspondence betwee these two objects.

Corollary 6.4.3. *If G is an abelian group with n elements, then G has exactly n inequivalent irreducible representations, which are all one-dimensional. On the other hand, if G is a non-abelian group with n elements, then G has strictly fewer than n inequivalent irreducible representations, and not all of them are one-dimensional.*

This follows immediately from Theorem 6.4.1 and the degree equation (6.9), as a group is abelian if and only if each element is in its own conjugacy class.

Exercise 6.4.1. If V is irreducible, show that

$$\sum_{g\in G} \overline{\chi_V(g)}\pi_V(g) = \frac{|G|}{\dim V}\ \mathrm{Id}_V\,.$$

Exercise 6.4.2. Let g and h be two elements in a group such that $\chi(g) = \chi(h)$ for all irreducible characters χ. Show that g and h belong to the same conjugacy class.

Exercise 6.4.3. By Corollary 6.4.3, any irreducible representation of an abelian group is one-dimensional. Give a more direct proof of this using Schur's lemma.

Exercise 6.4.4. Give an alternative proof of Corollary 6.1.8 by considering the decomposition of V into irreducible representations of the subgroup generated by g.

Exercise 6.4.5. Let 1_C denote the function that is 1 at a conjugacy class C and 0 else. Show that $\sqrt{|G|/|C|}\,1_C$ form an orthonormal basis for the class functions, where C runs over the conjugacy classes. Show that the change of base matrix between this basis and the orthonormal basis consisting of characters can be identified with the normalized character table.[3]

6.5 Examples

We will give some examples of character tables. We first consider the cyclic group $\mathbb{Z}_n$, then very briefly general finite abelian groups, then S_3 and finally S_4.

By Corollary 6.4.3, any irreducible representation of $\mathbb{Z}_n$ is one-dimensional, that is, it is a map $\pi : \mathbb{Z}_n \to \mathbb{C}$, such that $\pi(x+y) = \pi(x)\pi(y)$. It is easy to guess that $\pi(x) = e^{\lambda x}$ for some λ. Since $\pi(x+n) = \pi(x)$ we must have $\lambda = 2\mathrm{i}\pi\xi/n$, $\xi \in \mathbb{Z}$. Taking $\xi \in \mathbb{Z}_n$ gives inequivalent representations π_ξ. By Corollary 6.4.3, these are all the irreducible representations. The corresponding characters are $\chi_\xi(x) = \pi_\xi(x) = e^{2\mathrm{i}\pi x\xi/n}$, so the character table can be identified with the matrix $(e^{2\mathrm{i}\pi x\xi/n})_{0\le x,\xi\le n-1}$.

For $\mathbb{Z}_n$, there is a visible duality between characters and group elements. To make this more precise, note that $\chi_\xi\chi_\eta = \chi_{\xi+\eta}$, which reflects the equivalence of representations $\pi_\xi \otimes \pi_\eta \simeq \pi_{\xi+\eta}$. Thus, the characters form an abelian group under multiplication, and this group is ismomorphic to $\mathbb{Z}_n$. More generally, for any finite group G, the characters of one-dimensional representations form a group under multiplication. If G is abelian, this group is denoted $\hat{G}$. By Theorem 5.1.1, G is

[3]The functions 1_C are called class sums and will be important in §6.9.

then a direct product of cyclic groups. Using also Exercise 6.2.4, it follows that $\hat{G} \simeq G$ for any finite abelian group.

Next, we consider the group S_3. As we saw at the end of §6.3, it has exactly three irreducible representations: the trivial representation V_0, the alternating representation V_1 and the standard representation V_2. Let us write $\chi_j = \chi_{V_j}$. In agreement with Theorem 6.4.1, S_3 has three conjugacy classes, exemplified by the identity 1, the transposition (12) and the 3-cycle (123). The character χ_0 of the trivial representation is the constant function 1. The character χ_1 is simply the sign, which assumes the values 1, -1, 1 on the three respective elements. Note also that $\chi_2(1) = \dim(V_2) = 2$. Thus, the character table of S_3 has the form

	1	(12)	(123)
χ_0	1	1	1
χ_1	1	-1	1
χ_2	2		

The missing entries can be computed very easily using that each column must be orthogonal to the first, see (6.10b). This gives the end result

	1	(12)	(123)
χ_0	1	1	1
χ_1	1	-1	1
χ_2	2	0	-1

(6.11)

Finally, consider the group S_4. It has five conjugacy classes, exemplified by the elements 1, (12), (123), (1234) and (12)(34). The number of elements in these classes are, respectively, 1, 6, 8, 6 and 3. (This follows from Lemma 1.6.1 but is also easy to figure out directly.) We know two one-dimensional representations of S_4, the trivial representation and the alternating representation $\pi(\sigma)x = \mathrm{sgn}(\sigma)x$, $x \in \mathbb{C}$. Let us call them V_0 and V_1, respectively. Computing the sign of the five representative elements, we find that the character table starts as

	1	(12)	(123)	(1234)	(12)(34)
χ_0	1	1	1	1	1
χ_1	1	-1	1	-1	1

Next, we consider the defining representation on $V = \mathbb{C}^4$, that is,

$$\pi(\sigma)(x_1, x_2, x_3, x_4) = (x_{\sigma^{-1}(1)}, x_{\sigma^{-1}(2)}, x_{\sigma^{-1}(3)}, x_{\sigma^{-1}(4)}).$$

The matrix for $\pi(\sigma)$ in the standard basis is given by acting with σ on the columns

of the identity matrix. For instance,

$$\pi(1\ 2\ 3)(x_1,x_2,x_3,x_4) = (x_3,x_1,x_2,x_4) = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

In particular, the trace $\chi(\sigma)$ is the number of fix-points of σ. For our five elements, the respective number of fix-points is 4, 2, 1, 0, 0. Clearly, V is not irreducible, since $\{(x,x,x,x);\ x \in \mathbb{C}\}$ forms an invariant subspace equivalent to V_0. If we write $V \simeq V_0 \oplus V_2$, then $\chi_2 = \chi_V - \chi_0$ takes the values 3, 1, 0, -1, -1. Let us apply Corollary 6.2.6 to prove that V_2 is irreducible. Indeed,

$$\|\chi_2\|^2 = \frac{1}{24}\sum_C |C| \cdot |\chi_2(C)|^2 = \frac{1}{24}\left(1 \cdot 9 + 6 \cdot 1 + 8 \cdot 0 + 6 \cdot 1 + 3 \cdot 1\right) = 1.$$

Explicitly, $V_2 = V_0^\perp$ is the subspace $x_1 + \cdots + x_4 = 0$ of V. We can obtain yet another irreducible representation as $V_3 = V_1 \otimes V_2$, cf. Exercise 6.2.1, with character $\chi_3 = \chi_1\chi_2$. It can be realized explicitly on the same space as V_2, but with the group action

$$\pi(\sigma)(x_1,x_2,x_3,x_4) = \operatorname{sgn}(\sigma)(x_{\sigma^{-1}(1)},\ldots,x_{\sigma^{-1}(4)}).$$

So far we have the character table

	1	(12)	(123)	(1234)	(12)(34)
χ_0	1	1	1	1	1
χ_1	1	-1	1	-1	1
χ_2	3	1	0	-1	-1
χ_3	3	-1	0	1	-1

We know that we are missing exactly one irreducible representation. If its dimension is d, then (6.9) gives $24 = 1^2 + 1^2 + 3^2 + 3^2 + d^2$, so $d = 2$. Thus, the full character table is of the form

	1	(12)	(123)	(1234)	(12)(34)
χ_0	1	1	1	1	1
χ_1	1	-1	1	-1	1
χ_2	3	1	0	-1	-1
χ_3	3	-1	0	1	-1
χ_4	2				

The missing entries are now easy to fill in from the fact that each column is orthogonal to the first one (and to each other). We obtain the final result

	1	(12)	(123)	(1234)	(12)(34)
χ_0	1	1	1	1	1
χ_1	1	-1	1	-1	1
χ_2	3	1	0	-1	-1
χ_3	3	-1	0	1	-1
χ_4	2	0	-1	0	2

(6.12)

Note that this does not tell us what the representation V_4 is. We will return to that question in the next section.

Exercise 6.5.1. Deduce from the character table of S_3 that $V_1 \otimes V_2 \simeq V_2$. Then give an explicit equivalence between the two representations.

Exercise 6.5.2. Let V be the standard representation of S_3. Find the explicit decomposition of $V^{\otimes n}$ as a sum of irreducible representations.

Exercise 6.5.3. Define the finite Fourier transform of $f : \mathbb{Z}_n \to \mathbb{C}$ by

$$\hat{f}(\xi) = \frac{1}{n} \sum_{x \in \mathbb{Z}_n} f(x) e^{-2\pi \mathrm{i} x\xi/n}, \qquad \xi \in \mathbb{Z}_n.$$

Deduce from Theorem 6.4.1 that $\sqrt{n}\|\hat{f}\|_{\mathrm{L}^2(\mathbb{Z}_n)} = \|f\|_{\mathrm{L}^2(\mathbb{Z}_n)}$. Choosing $f(x) = 1$ for $0 \le x \le k-1$ and $f(x) = 0$ for $k \le x \le n-1$, show that

$$\sum_{x=1}^{n-1} \left(\frac{\sin(\pi k x/n)}{\sin(\pi x/n)} \right)^2 = k(n-k), \qquad 0 \le k \le n.$$

6.6 Interpreting the character table

Although a group is not uniquely determined by its character table (see Exercices 6.6.3–6.6.4), a lot of information about it is. To give some examples, we will show how to determine all normal subgroups and identify the center and the commutator subgroup.

When χ is a character, we will write $\mathrm{Ker}(\chi) = \{g \in G;\, \chi(g) = \chi(1)\}$. The reason for this notation is the following fact.

Lemma 6.6.1. *If the representation π has character χ, then* $\mathrm{Ker}(\pi) = \mathrm{Ker}(\chi)$.

Proof. Let $\lambda_1, \ldots, \lambda_m$ be the eigenvalues of $\pi(g)$ (counted with multiplicity). By Corollary 6.1.8, $g \in \mathrm{Ker}(\pi)$ if and only if $\lambda_j = 1$ for all j. Moreover, $g \in \mathrm{Ker}(\chi)$ if and only if $\lambda_1 + \cdots + \lambda_m = m$. The equivalence of these conditions follows from the fact that $|\lambda_j| = 1$ for all j, see Lemma 6.2.1. □

Let N be a normal subgroup of G and consider a representation of G/N. By generalities on normal subgroups, there is a bijection between representations ρ of G/N and representations π of G such that $N \subseteq \mathrm{Ker}(\pi)$. This correspondence is simply given by $\rho(gN) = \pi(g)$. As a subspace is closed under ρ if and only if it is closed under π, the representation ρ is irreducible if and only if π is. We conclude that the irreducible characters of G/N are given by

$$\mathrm{Irr}(G/N) = \{\chi \in \mathrm{Irr}(G);\ N \subseteq \mathrm{Ker}(\chi)\}. \tag{6.13}$$

If $g \in \mathrm{Ker}(\chi)$ for all χ, then choosing D as the conjugacy class of g and $C = \{1\}$ in (6.10b) we find that $g = 1$. Applying this fact to the group G/N gives

$$N = \bigcap_{\chi \in \mathrm{Irr}(G),\, N \subseteq \mathrm{Ker}(\chi)} \mathrm{Ker}(\chi). \tag{6.14}$$

Thus, any normal subgroup is an intersection of kernels of irreducible characters. Since the converse is clear, all normal subgroups can be found from the character table.

Proposition 6.6.2. *The normal subgroups of G are precisely the intersections of kernels of irreducible characters.*

Let us apply Proposition 6.6.2 to the character table (6.12) of S_4. Obviously, the trivial representation χ_0 has S_4 as kernel. The representations χ_2 and χ_3 have trivial kernel (that is, they are faithful). The kernel of χ_1, that is, $\{\sigma \in \mathrm{S}_4;\ \mathrm{sgn}(\sigma) = 1\}$ is known as the *alternating group* A_4. The kernel of χ_4 is the subgroup $N = \{\mathrm{id}, (12)(34), (13)(24), (14)(23)\}$. Since $N \subseteq \mathrm{A}_4$, we do not find any more normal subgroups by taking intersections of the kernels. Thus, S_4 has only two non-trivial normal subgroups, A_4 and N. Note that S_4/N is a group of order 6. As it is nonabelian (for instance, $(12)(13)(12)(13) = (123) \notin N$, so $(12)(13) \neq (13)(12)$ in S_4/N), it follows from Exercise 1.7.21 that $\mathrm{S}_4/N \simeq \mathrm{S}_3$.[4] Thus, χ_4 can also be viewed as an irreducible representation of S_3. As it has dimension 2, it is the standard representation. This answers our question about what the representation with character χ_4 "is": it is the standard representation of S_3, disguised as a representation of S_4.

Let us now consider the commutator subgroup (or derived group) G'. By definition, it is the subgroup generated by all elements of the form $aba^{-1}b^{-1}$. Since G is abelian if and only if G' is trivial, $|G|/|G'|$ measures how commutative G is.

[4]The existence of N is quite exceptional. If $n \neq 4$, then the only normal subgroups of S_n are A_n and the trivial subgroups. A geometric way of understanding S_3 as a quotient of S_4 is as follows. A cube has four long diagonals and three axes connecting the midpoints of opposite faces. The rotations of the cube can be identified with the group S_4 acting on the diagonals (see Exercise 1.7.17). Each rotation also permutes the three axes, which gives a surjective homomorphism $\mathrm{S}_4 \to \mathrm{S}_3$. The kernel of this homomorphism (that is, the group of rotations fixing each pair of opposite faces) is N.

Proposition 6.6.3. *The subgroup G' is the intersection of the kernels of all one-dimensional representations of G. Moreover, the number of one-dimensional representations of G is $|G|/|G'|$.*

Proof. By (6.13),

$$\mathrm{Irr}(G/G') = \{\chi \in \mathrm{Irr}(G);\ G' \subseteq \mathrm{Ker}(\chi)\}$$

and by (6.14),

$$G' = \bigcap_{\chi \in \mathrm{Irr}(G),\, G' \subseteq \mathrm{Ker}(\chi)} \mathrm{Ker}(\chi).$$

Thus, it suffices to show that a representation of G is one-dimensional if and only if it can be projected to G/G'. Since G/G' is abelian, any representation of G/G' is one-dimensional. Conversely, if π is one-dimensional, then all the 1×1-matrices $\pi(a)$ commute. In particular, $\pi(aba^{-1}b^{-1}) = \pi(a)\pi(b)\pi(a)^{-1}\pi(b)^{-1} = 1$, so π projects to G/G'. □

Next, we consider the center. Recall that $\mathrm{Z}(G)$ is the set of elements in G that commute with all other elements; it is an abelian normal subgroup. Since G is abelian if and only if $\mathrm{Z}(G) = G$, $|G|/|\mathrm{Z}(G)|$ measures how non-commutative G is.

When χ is a character, we define

$$\mathrm{Z}(\chi) = \{g \in G;\ |\chi(g)| = \chi(1)\}.$$

Again using that the eigenvalues of $\pi(g)$ are roots of unity, this can only happen when they are all equal, so we can also write

$$\mathrm{Z}(\chi) = \{g \in G;\ \pi(g) \in \mathbb{C}\,\mathrm{Id}\}, \qquad \chi = \mathrm{Tr}(\pi).$$

The reason for the notation is the following fact.

Lemma 6.6.4. *The center of G is $\mathrm{Z}(G) = \bigcap_{\chi \in \mathrm{Irr}(G)} \mathrm{Z}(\chi)$.*

Proof. Let π be an irreducible representation on V and χ its character. If $g \in \mathrm{Z}(G)$ then $\pi(g)\pi(h) = \pi(h)\pi(g)$ for all $h \in G$. Thus, $\pi(g) \in \mathrm{End}_G(V)$. By Schur's lemma, $g \in \mathrm{Z}(\chi)$. Conversely, if $\pi(g) = \lambda\,\mathrm{Id}$, then for any $h \in G$ we have $\pi(g)\pi(h)\pi(g)^{-1}\pi(h)^{-1} = \mathrm{Id}$, which implies $ghg^{-1}h^{-1} \in \mathrm{Ker}(\pi)$. If this holds for each irreducible representation π, then (6.14) with $N = \{1\}$ gives $ghg^{-1}h^{-1} = 1$, so $g \in \mathrm{Z}(G)$. □

As an example, suppose G has the character table

$$\begin{array}{c|cccccc} & C_1 & C_2 & C_3 & C_4 & C_5 & C_6 \\ \chi_1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \chi_2 & 1 & -1 & 1 & -1 & \mathrm{i} & -\mathrm{i} \\ \chi_3 & 1 & 1 & 1 & 1 & -1 & -1 \\ \chi_4 & 1 & -1 & 1 & -1 & -\mathrm{i} & \mathrm{i} \\ \chi_5 & 2 & 2 & -1 & -1 & 0 & 0 \\ \chi_6 & 2 & -2 & -1 & 1 & 0 & 0 \end{array} .$$

We follow the convention that the first column corresponds to the conjugacy class $C_1 = \{1\}$ and the first row to the trivial representation. This can be seen immediately since, by (6.10), they are the unique row and column with only positive entries. The entries of the first column are then the dimensions of the corresponding representations. In this case, $\chi_1, \ldots, \chi_4$ are one-dimensional and χ_5, χ_6 two-dimensional. The order of the group is thus $4 \cdot 1^2 + 2 \cdot 2^2 = 12$. By (6.10b), the order of a conjugacy class can be read off the character table as

$$|C| = \frac{|G|}{\sum_{\chi \in \mathrm{Irr}(G)} |\chi(C)|^2}.$$

(Here and below, we use $\mathrm{Irr}(G)$ to denote the irreducible characters as well as the irreducible representations; as a representation is uniquely determined by a character this should not lead to confusion.) In the case at hand, we find that $|C_1| = |C_2| = 1$, $|C_3| = |C_4| = 2$, $|C_5| = |C_6| = 3$. We may check again that the total number of elements is $2 \cdot 1 + 2 \cdot 2 + 2 \cdot 3 = 12$. Writing for short $C_{i_1 \cdots i_k} = C_{i_1} \cup \cdots \cup C_{i_k}$, the kernels of the characters are $\mathrm{Ker}(\chi_1) = C_{123456} = G$, $\mathrm{Ker}(\chi_2) = \mathrm{Ker}(\chi_4) = C_{13}$, $\mathrm{Ker}(\chi_3) = C_{1234}$, $\mathrm{Ker}(\chi_5) = C_{12}$, $\mathrm{Ker}(\chi_6) = C_1$. Taking intersections of these does not lead to any further subgroups; for instance, $C_{13} \cap C_{12} = C_1$. Thus, G has exactly five normal subgroups. Moreover $G' = \bigcap_{j=1}^{4} \mathrm{Ker}(\chi_j) = C_{13}$. To compute the center, we see at a glance that $Z(\chi) = G$ for the one-dimensional representations and $Z(\chi) = C_{12}$ for the two-dimensional ones. Thus, $Z(G) = C_{12}$. In conclusion, the lattice of normal subgroups looks like

$$C_1 = \{1\} \subseteq \left\{ \begin{array}{c} C_{12} = Z(G) \\ C_{13} = G' \end{array} \right\} \subseteq C_{1234} \subseteq C_{123456} = G.$$

We mention that there is exactly one group with this character table, called the dicyclic group of order 12.

Exercise 6.6.1. Let G be a group with the following character table. Find the order of G and the number of elements in each conjugacy class. Describe the lattice

of normal subgroups and identify the commutator subgroup and the center.

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}
χ_1	1	1	1	1	1	1	1	1	1	1
χ_2	1	1	−1	−1	1	1	−1	1	1	−1
χ_3	1	1	1	1	1	−1	−1	−1	−1	−1
χ_4	1	1	−1	−1	1	−1	1	−1	−1	1
χ_5	2	−1	0	0	2	2	0	−1	2	0
χ_6	2	−1	0	0	2	−2	0	1	−2	0
χ_7	3	0	−1	1	−1	3	1	0	−1	−1
χ_8	3	0	1	−1	−1	3	−1	0	−1	1
χ_9	3	0	−1	1	−1	−3	−1	0	1	1
χ_{10}	3	0	1	−1	−1	−3	1	0	1	−1

Exercise 6.6.2. Prove that, for any finite group G, the one-dimensional characters form a group isomorphic to G/G'.

Exercise 6.6.3. Let G be the group of symmetries of the square (the dihedral group of order 8). Show that G has the character table

1	1	1	1	1
1	1	1	−1	−1
1	1	−1	1	−1
1	1	−1	−1	1
2	−2	0	0	0

Exercise 6.6.4. Let G be the quaternion group (see Exercise 1.7.3). Show that G has the same character table as the dihedral group of order 8. Also show that the two groups are not isomorphic.

6.7 Group algebra and Fourier transform

So far, we have only considered $\mathbb{C}[G]$ as a vector space. However, it has a natural multiplication, which makes it an associative algebra. In the description as formal sums, this multiplication is simply obtained by extending the group multiplication:

$$\sum_g a_g g \sum_g b_h h = \sum_{g,h} a_g b_h gh. \tag{6.15}$$

In the description as functions, it is a convolution:

$$(\phi\psi)(g) = \sum_{h\in G} \phi(h)\psi(h^{-1}g). \tag{6.16}$$

We call $\mathbb{C}[G]$ equipped with this product the *group algebra* of G.

As we have seen, left multiplication $\pi(g)h = gh$ gives a representation of G on $\mathbb{C}[G]$ called the *regular representation*. It decomposes as

$$\mathbb{C}[G] \simeq \bigoplus_{V \in \mathrm{Irr}(G)} (\dim V)\, V. \tag{6.17}$$

This does not say anything about the algebra structure on $\mathbb{C}[G]$. However, we can easily obtain a refinement of (6.17) that does. To this end, we consider $\mathbb{C}[G]$ as a representation of $G \times G$ by the rule $\pi(g,h)k = gkh^{-1}$. Note that if V is any representation of G, then $\mathrm{End}(V)$ is a representation of $G \times G$ under $\pi_{\mathrm{End}(V)}(g,h)\phi = \pi_V(g) \circ \phi \circ \pi_V(h^{-1})$. Moreover, $\mathrm{End}(V)$ is an algebra (under composition of maps). This gives meaning to the following result.

Theorem 6.7.1. *As a representation of $G \times G$ and as an associative algebra,*

$$\mathbb{C}[G] \simeq \bigoplus_{V \in \mathrm{Irr}(G)} \mathrm{End}(V). \tag{6.18}$$

Proof. A representation $\pi_V : G \to \mathrm{GL}(V)$ extends by linearity to a map $\pi_V : \mathbb{C}[G] \to \mathrm{End}(V)$. It is trivial to check that π_V commutes with the action of $G \times G$ and preserves multiplication. Thus, it is enough to check that $\mathcal{F} = \bigoplus_V \pi_V$ is bijective. If $\mathcal{F}(x) = 0$, then $\pi_V(x) = 0$ for any irreducible representation V and hence for any representation V. But then, in the regular representation, $x = \pi(x)1 = 0$. This shows that $\mathcal{F}$ is injective. The surjectivity follows by comparing dimensions; indeed, by (6.17), $\dim(\mathbb{C}[G]) = \sum_V \dim(V)^2 = \sum_V \dim \mathrm{End}(V)$. □

One can prove that if $V \in \mathrm{Irr}(G)$ then $\mathrm{End}(V) \in \mathrm{Irr}(G \times G)$, so (6.18) is in fact a decomposition into irreducible components (see Exercise 6.7.3).

The isomorphism $\mathcal{F} : \mathbb{C}[G] \to \bigoplus_V \mathrm{End}(V)$ defined in the proof of Theorem 6.7.1 is called the *Fourier transform* on G. To see why, consider the case $G = \mathbb{Z}_n$. As all irreducible representations V are one-dimensional, we can identify $\mathrm{End}(V)$ with $\mathbb{C}$. We can then view $\mathcal{F}$ as a map from functions $G \to \mathbb{C}$ to functions $\mathrm{Irr}(G) = \hat{G} \to \mathbb{C}$. The value of $\mathcal{F}(f)$ on the character $\chi_\xi(x) = e^{2\pi i x\xi/n}$ is then $\sum_{g\in G} f(g)\chi_\xi(g) = \sum_{x=0}^{n-1} f(x)e^{2\pi i x\xi/n}$. This is the *finite Fourier transform*, see Exercise 6.5.3. For general finite groups, the Fourier transform takes a scalar-valued function on G to an operator-valued function on $\mathrm{Irr}(G)$.

In classical Fourier analysis, one of the most fundamental problems is how to recreate a function from its Fourier transform. The analogous result for finite groups is as follows.

Theorem 6.7.2 (Fourier inversion formula)**.** *If $A \in \mathrm{End}(V)$, then*

$$\mathcal{F}^{-1}(A)(g) = \frac{\dim V}{|G|}\, \mathrm{Tr}_V\left(A\pi_V(g^{-1})\right).$$

Here, we consider $\mathbb{C}[G]$ as a space of functions on G.

Proof. An equivalent formulation is that if $A = \sum_V A_V \in \bigoplus_V \mathrm{End}(V)$, then

$$\mathcal{F}^{-1}(A)(g) = \sum_{V \in \mathrm{Irr}(G)} \frac{\dim V}{|G|} \operatorname{Tr}_V \left(A_V \pi_V(g^{-1}) \right).$$

By linearity, it is enough to verify this when $A = \mathcal{F}(h)$ for some $h \in G$. On the left, $\mathcal{F}^{-1}(A) = h$ should be identified with the function $g \mapsto \delta_{g,h}$. On the right, $A_V = \pi_V(h)$. Thus, we must prove that

$$\delta_{g,h} = \sum_{V \in \mathrm{Irr}(G)} \frac{\dim V}{|G|} \operatorname{Tr}_V \left(\pi_V(h) \pi_V(g^{-1}) \right) = \sum_{V \in \mathrm{Irr}(G)} \frac{\dim V}{|G|} \chi_V \left(hg^{-1} \right). \quad (6.19)$$

This is the special case $hg^{-1} \in C$ and $D = \{1\}$ of the orthogonality relation (6.10b). □

Exercise 6.7.1. Prove that the center of the group algebra is the space of class functions.

Exercise 6.7.2. Prove that if G and H are finite abelian groups of the same order, then $\mathbb{C}[G] \simeq \mathbb{C}[H]$ as associative algebras. Give an explicit isomorphism between $\mathbb{C}[\mathbb{Z}_2 \times \mathbb{Z}_2]$ and $\mathbb{C}[\mathbb{Z}_4]$.[5]

Exercise 6.7.3. Prove that $\chi_{\mathrm{End}(V)}(g,h) = \chi_V(g)\chi_V(h^{-1})$. Deduce that $\mathrm{End}(V) \in \mathrm{Irr}(G \times G)$ if $V \in \mathrm{Irr}(G)$.

Exercise 6.7.4. In general, a *representation* of an associative algebra A over a commutative ring R is an algebra homomorphism $A \to \mathrm{End}_R(V)$ for some R-module V. Show that there is a one-to-one correspondence between representations of a finite group and representations of the group algebra.

Exercise 6.7.5. We have seen that, when G is embedded into $G \times G$ as $g \mapsto (g, 1)$, then each irreducible representation V appears in $\mathbb{C}[G]$ with multiplicity $\dim V$. Show that, if G is embedded as $g \mapsto (g, g)$, then the multiplicity is $\sum_C \chi_V(C)$, where the sum runs over all conjugacy classes. Deduce that the row sums in the character table are always non-negative integers.

[5] Although we do not discuss it explicitly in the present notes, $\mathbb{C}[G]$ has the additional structure of a Hopf algebra which, roughly speaking, means that there is a rule for defining tensor products of representations of $\mathbb{C}[G]$ (and also notions of dual and trivial representations). If $\mathbb{C}[G] \simeq \mathbb{C}[H]$ as Hopf algebras, then $G \simeq H$ as groups.

6.8 Peter–Weyl Theorem

One important class of theorems in classical Fourier analysis, associated with the names Parseval and Plancherel, state that various versions of the Fourier transform are unitary. We will show such a result in the setting of finite groups. As we will see, it is an easy consequence of the unitarity of the character table (6.10).

To talk about unitarity, we must introduce inner products on $\mathbb{C}[G]$ and $\mathrm{End}(V)$. The group algebra will be identified with the inner product space $\mathrm{L}^2(G)$. On $\mathrm{End}(V)$, we will use a rescaled version of the following product.

Lemma 6.8.1. *If V is a representation of G, then $\langle A, B\rangle = \mathrm{Tr}(AB^*)$ is a $G \times G$-invariant inner product on $\mathrm{End}(V)$, where $*$ denotes adjoint with respect to a G-invariant inner product on V.*

Proof. If we choose an orthogonal basis $(e_k)_{k=1}^n$ for V and let $A = (a_{kl})$, $B = (b_{kl})$ be the corresponding matrices, then it is easy to check that

$$\langle A, B\rangle = \sum_{k,l=1}^{n} a_{kl}\bar{b}_{kl}. \tag{6.20}$$

In particular, this implies that the inner product is positive definite. The invariance follows from an easy computation:

$$\begin{aligned}\big\langle \pi_{\mathrm{End}(V)}(g,h)A, \pi_{\mathrm{End}(V)}(g,h)B\big\rangle &= \mathrm{Tr}\left(\pi_V(g)A\pi_V(h)^{-1}(\pi_V(g)B\pi_V(h^{-1}))^*\right)\\ &= \mathrm{Tr}\left(\pi_V(g)A\pi_V(h)^{-1}\pi_V(h)B^*\pi_V(g^{-1})\right) = \mathrm{Tr}\left(\pi_V(g)AB^*\pi_V(g^{-1})\right)\\ &= \mathrm{Tr}(AB^*) = \langle A, B\rangle,\end{aligned}$$

where we used that $\pi_V(g)^* = \pi_V(g^{-1})$. □

If $V \in \mathrm{Irr}(G)$, it follows from Exercise 6.1.3 and Exercise 6.7.3 that the $G \times G$-invariant inner product on $\mathrm{End}(V)$ is unique up to rescaling. We will in fact define

$$\langle A, B\rangle = \lambda_V \,\mathrm{Tr}_V(AB^*), \qquad A,\, B \in \mathrm{End}(V), \tag{6.21}$$

where $\lambda_V = \dim V/|G|^2$. We then have a natural inner product on $\bigoplus_V \mathrm{End}(V)$, given by

$$\left\langle \sum_V A_V, \sum_V B_V \right\rangle = \sum_V \lambda_V \,\mathrm{Tr}_V(A_V B_V^*), \qquad A_V,\, B_V \in \mathrm{End}(V).$$

Theorem 6.8.2. *With respect to the inner products defined above, the Fourier transform $\mathcal{F} : \mathrm{L}^2(G) \to \bigoplus_{V\in\mathrm{Irr}(G)} \mathrm{End}(V)$ is unitary, that is, $\mathcal{F}^{-1} = \mathcal{F}^*$.*

With respect to the direct sum composition $\bigoplus_V \operatorname{End}(V)$, $\mathcal{F}$ takes the form of a block matrix

$$\mathcal{F} = \begin{bmatrix} \pi_{V_0} \\ \pi_{V_1} \\ \vdots \end{bmatrix}, \qquad \operatorname{Irr}(G) = \{V_0, V_1, \dots\}.$$

Thus, the equivalent identities $\mathcal{F}\mathcal{F}^* = \operatorname{Id}_{\bigoplus_V \operatorname{End}(V)}$, $\mathcal{F}^*\mathcal{F} = \operatorname{Id}_{\mathrm{L}^2(G)}$ can be written as

$$\pi_W \circ \pi_V^* = \delta_{V,W} \operatorname{Id}_{\operatorname{End}(V)}, \tag{6.22a}$$

$$\sum_{V \in \operatorname{Irr}(G)} \pi_V^* \circ \pi_V = \operatorname{Id}_{\mathrm{L}^2(G)}. \tag{6.22b}$$

Note that it follows from Schur's Lemma and Exercise 6.7.3 that (6.22a) holds up to a multiplicative constant. Although one could prove Theorem 6.8.2 using this observation, we will simply observe that (6.22b) has already appeared in slight disgusise in §6.7. Namely, as the group elements generate $\mathrm{L}^2(G)$, it suffices to compare $\langle Xg, h\rangle_{\mathrm{L}^2(G)}$, where X are the two sides of (6.22b) and g, $h \in G$. On the right, we get $\delta_{g,h}/|G|$. On the left, we get

$$\sum_{V \in \operatorname{Irr}(G)} \langle \pi_V(g), \pi_V(h)\rangle_{\operatorname{End}(V)} = \sum_{V \in \operatorname{Irr}(G)} \frac{\dim(V)}{|G|^2} \operatorname{Tr}\left(\pi_V(g)\pi_V(h)^*\right).$$

As $\pi_V(h)^* = \pi_V(h^{-1})$ for an invariant scalar product, we are reduced to the identity (6.19).

In Theorem 6.4.1, we saw that the characters of irreducible representations form an orthogonal basis for the subspace of $\mathrm{L}^2(G)$ consisting of class functions. As a consequence of Theorem 6.8.2, we can find an extension to an orthogonal basis for the whole space. To this end, assume that we have chosen an invariant inner product and an orthogonal basis on each $V \in \operatorname{Irr}(G)$. Identifying elements of $\operatorname{End}(V)$ with matrices, let $E_{kl}^{(V)}$ be the matrix with a 1 in position (k, l) and a 0 everywhere else. It is then clear from (6.20) that $(E_{kl}^{(V)})_{k,l=1}^n$ form an orthogonal basis for $\operatorname{End}(V)$. By Theorem 6.8.2, $\mathcal{F}^{-1}(E_{k,l}^{(V)})$ form an orthogonal basis for $\mathrm{L}^2(G)$, where V varies over $\operatorname{Irr}(G)$ and $1 \le k, l \le \dim(V)$. Applying Theorem 6.7.2 we find that, as a function on G,

$$\mathcal{F}^{-1}\left(E_{k,l}^{(V)}\right)(g) = \frac{\dim V}{|G|} \operatorname{Tr}\left(E_{kl}^{(V)} \pi_V(g^{-1})\right) = \frac{\dim V}{|G|} \pi_V(g^{-1})_{lk}. \tag{6.23}$$

Here, we used the elementary identity $\operatorname{Tr}(E_{kl}A) = A_{lk}$, where A_{lk} is a matrix element of A. Since the Fourier transform is unitary, it follows that the matrix elements $\pi^V(g)_{kl}$ form an orthogonal basis of $\mathrm{L}^2(G)$. In view of the rescaling (6.21), $\|E_{kl}^{(V)}\|^2 = \dim V/|G|^2$. As $\mathcal{F}$ preserves the norm, it follows that $\|\pi_V(g)\|^2 = 1/\dim V$. Thus, we arrive at the following result.

Theorem 6.8.3 (Peter–Weyl Theorem for finite groups)**.** *The matrix elements* $\pi_V(g)_{kl}$ *of all irreducible representations form an orthogonal basis for* $\mathrm{L}^2(G)$*. Explicitly, the orthogonality relation is*

$$\frac{1}{|G|}\sum_{g\in G}\pi_V(g)_{ij}\,\overline{\pi_W(g)_{kl}} = \frac{\delta_{ik}\delta_{jl}\delta_{VW}}{\dim V}. \tag{6.24}$$

The relations (6.24) are called Schur's orthogonality relations. Theorem 6.8.3 can be generalized to compact groups; for instance, the case $G = \mathbb{R}/\mathbb{Z}$ is related to classical Fourier series.

6.9 Frobenius divisibility

To go further into the representation theory of finite groups, one often needs some algebraic number theory and Galois theory. Just to give a glimpse of what can be done, we will prove the following theorem of Frobenius.

Theorem 6.9.1. *If* V *is an irreducible representation of a group* G*, then* $\dim V$ *divides* $|G|$*.*

With a little bit of extra effort one can show that $\dim V$ divides $|G|/|\mathrm{Z}(G)|$.

To prove Theorem 6.9.1, we need the notion of an *algebraic integer*, which is defined as a solution x to an equation of the form

$$x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = 0, \qquad a_0, \ldots, a_{n-1} \in \mathbb{Z}. \tag{6.25}$$

For instance, i and $\sqrt{2}$ are algebraic integers, but $1/2$ is not. Denoting the algebraic integers by $\mathbb{A}$, it is very easy to see that $\mathbb{A}\cap\mathbb{Q} = \mathbb{Z}$. Thus, to prove Theorem 6.9.1 it is enough to prove that $|G|/\dim V \in \mathbb{A}$.

Lemma 6.9.2. *The following conditions are equivalent:*

(A) α *is an algebraic integer.*

(B) *There exists a ring* $R \subseteq \mathbb{C}$ *such that* $\alpha \in R$ *and* R *is finitely generated as an abelian group.*

Proof. If α solves (6.25), then $\mathbb{Z}[\alpha]$ is generated by $1, \alpha, \ldots, \alpha^{n-1}$. Thus, we can take $R = \mathbb{Z}[\alpha]$.

Conversely, if R is a ring as in (B) and $\beta_1, \ldots, \beta_n$ are generators, then we can write $\alpha\beta_j = \sum_k a_{jk}\beta_k$ for some $a_{jk} \in \mathbb{Z}$. This means that the column vector $(\beta_1, \ldots, \beta_n)$ is an eigenvector of the matrix $A = (a_{jk})$ with eigenvalue α. It follows that $\det(\alpha\,\mathrm{Id} - A) = 0$, which is an equation for α of the form (6.25). □

Corollary 6.9.3. *The algebraic integers form a ring.*

Proof. Let α and β be algebraic integers. If α solves a monic equation of degree m and β a monic equation of degree n, then $\mathbb{Z}[\alpha,\beta]$ is generated as an abelian group by the monomials $\alpha^k\beta^l$ with $k < m$ and $l < n$. Since $\alpha+\beta \in \mathbb{Z}[\alpha,\beta]$ and $\alpha\beta \in \mathbb{Z}[\alpha,\beta]$ we conclude from Lemma 6.9.2 that $\alpha+\beta$ and $\alpha\beta$ are algebraic integers. □

Recall the group algebra $\mathbb{C}[G]$ introduced in §6.7. We write its elements as $\alpha = \sum_{g\in G}\alpha(g)g$, which may be viewed either as a formal linear combination of elements in G or as a function $\alpha : G \to \mathbb{C}$. Let $\pi_V : G \to \mathrm{GL}(V)$ be an irreducible representation. We may then extend π_V linearly to an algebra homomorphism $\pi_V : \mathbb{C}[G] \to \mathrm{End}(V)$.

Lemma 6.9.4. *Let R be the subspace of $\mathbb{C}[G]$ consisting of integer-valued class functions. Then, R is a subring and finitely generated as an abelian group.*

Proof. It is clear that the space $\mathbb{Z}[G]$ consisting of all integer-valued functions on G form a subring of $\mathbb{C}[G]$. We will show that R is the center of $\mathbb{Z}[G]$, which implies that R is a ring. Since $\mathbb{Z}[G]$ is generated as a ring by the elements of G, it is enough to investigate which formal sums $x = \sum_g \alpha(g)g$ commute with all group elements. We have $xh = hx$ if and only if $x = \sum_g \alpha(g)hgh^{-1}$. Replacing g by $h^{-1}gh$ we see that $\alpha(h^{-1}gh) = \alpha(g)$ for all $h \in G$, that is, α is a class function.

For the second statement, note that any element in R is a $\mathbb{Z}$-linear combination of the elements $1_C = \sum_{g\in C} g$, which take the value 1 on a conjugacy class C and 0 elsewhere. Hence, the elements 1_C generate R as a group. □

If $\alpha \in \mathbb{C}[G]$ is a class function, then by the proof of Theorem 6.4.1

$$\pi_V(\alpha) = \sum_{g\in G}\alpha(g)\pi_V(g) = \lambda\,\mathrm{Id}, \tag{6.26}$$

for some $\lambda = \lambda(\alpha) \in \mathbb{C}$. Consider the restriction of λ to R. Since π_V is a ring homomorphism, so is the map $\lambda : R \to \mathbb{C}$. Hence, $\lambda(R)$ is a subring of $\mathbb{C}$ that is finitely generated as an abelian group. It then follows from Lemma 6.9.2 that $\lambda(R) \subseteq \mathbb{A}$.

Let us now consider the number $\lambda(\bar{\chi}_V)$ (which need not be in $\lambda(R)$). Taking traces in (6.26) gives

$$\lambda(\bar{\chi}_V)\dim(V) = \sum_{g\in G}\overline{\chi_V(g)}\chi_V(g) = |G|,$$

by Lemma 6.2.4. On the other hand, $\bar{\chi}_V = \sum_C \bar{\chi}_V(C)1_C$, the sum being over conjugacy classes. It follows that

$$\frac{|G|}{\dim(V)} = \lambda(\bar{\chi}_V) = \sum_C \bar{\chi}_V(C)\lambda(1_C).$$

By Lemma 6.2.1, $\bar{\chi}_V(C)$ is a sum of roots of unity. As any root of unity is an algebraic integer, $\lambda(1_C) \in \lambda(R)$ are algebraic integers and $\mathbb{A}$ is a ring, it follows that $|G|/\dim V \in \mathbb{A}$. As we explained above, this proves Theorem 6.9.1.

Exercise 6.9.1. Let G be a group of order p^2, where p is a prime. Use Theorem 6.9.1 to prove that G is abelian (thus, by Theorem 5.1.1, $G \simeq \mathbb{Z}_p^2$ or $G \simeq \mathbb{Z}_{p^2}$).

Exercise 6.9.2. Let G be a group of order pq, where $p < q$ are primes. Using Proposition 6.6.3 and Theorem 6.9.1, show that if $p \nmid q - 1$, then G is abelian (thus, by Theorem 5.1.1, $G \simeq \mathbb{Z}_p \times \mathbb{Z}_q$).[6]

Exercise 6.9.3. Show that if an entry in the character table of a finite group is rational, then it is an integer.

Exercise 6.9.4. Show that, in the notation used above,

$$\lambda(1_C) = \frac{\chi_V(C)|C|}{\dim(V)}.$$

Deduce that if $\chi_V(C) \in \mathbb{Q}$ (so by the previous exercise $\chi_V(C) \in \mathbb{Z}$), then $\chi_V(C)|C|$ is divisible by $\dim(V)$.

Exercise 6.9.5. Using the method of proof of Lemma 6.9.2, find an equation of the form (6.25) for $x = \sqrt{2} + \sqrt{3}$.

Exercise 6.9.6. Using (6.16), give a more direct proof of the fact that the integer-valued class functions form a ring.

6.10 Additional exercises

Exercise 6.10.1. Let $\pi : G \to \mathrm{S}_X$ be an action of a finite group on a finite set X. Let V be the free vector space generated by X. Show that π extends to a group action $G \to \mathrm{End}(V)$. Show that the map $\phi(x) = x^*$, $x \in X$, defines an equivalence of representations $\phi : V \to V^*$.

[6] If $p \mid q - 1$ there is also a non-abelian group of order pq.

Exercise 6.10.2. A finite group G has six conjugacy classes, which we denote $C_1, \ldots, C_6$. We know two irreducible characters χ_1 and χ_2 of the group, given by

	C_1	C_2	C_3	C_4	C_5	C_6
χ_1	1	-1	1	-1	i	$-$i
χ_2	2	2	-1	-1	0	0

Compute the whole character table of G.

Exercise 6.10.3. If χ is an irreducible character, show that $\mathrm{Z}(\chi)/\operatorname{Ker}(\chi)$ is a cyclic group. Deduce that if G has a faithful irreducible representation, then $\mathrm{Z}(G)$ is cyclic.

Exercise 6.10.4. Show that

$$\sum_{g\in G} \phi(g)\psi(g^{-1}) = \frac{1}{|G|} \sum_{V\in \mathrm{Irr}(G)} (\dim V)\chi_V(\phi\psi), \qquad \phi,\, \psi \in \mathbb{C}[G]$$

(here, the character χ_V is extended linearly to $\mathbb{C}[G]$).

Exercise 6.10.5. Let W be any representation. Suppose the decomposition into irreducibles is

$$W = k_1V_1 \oplus \cdots \oplus k_mV_m.$$

Prove that the projection from W to the subspace equivalent to k_1V_1 (an *isotypic component*) is given by

$$\frac{\dim V_1}{|G|} \sum_{g\in G} \overline{\chi_{V_1}(g)}\pi_W(g).$$

Exercise 6.10.6. In the context of Problem 6.10.5, show that W is contained in the eigenspace of the class sum 1_C with eigenvalue $|C|\chi_{V_1}(C)/\dim(V_1)$. Deduce that any isotypic component W can be obtained as the intersection of eigenspaces of class sums.

Exercise 6.10.7. Suppose V and W are irreducible representations. Show that

$$\frac{\dim V}{|G|} \sum_{g\in G} \chi_V(gh)\overline{\chi_W(g)} = \begin{cases} \chi_V(h), & V = W, \\ 0, & V \not\cong W. \end{cases}$$

Exercise 6.10.8. Let C and D be two conjugacy classes of G. For $e \in G$, let $n(C, D, e)$ be the number of pairs $(c, d) \in C \times D$ such that $cd = e$. Show that $n(C, D, e)$ depends only on the conjugacy class E of e, so it makes sense to write it as $n(C, D, E)$. Show that the elements 1_C introduced in §6.9 satisfy

$$1_C 1_D = \sum_E n(C, D, E)1_E.$$

Finally, show that the numbers $\lambda(1_C)$ satisfy

$$\det_{D,E}\big(\lambda(1_C) - n(C,D,E)\big) = 0,$$

where the determinant is indexed by conjugacy classes of G. (This gives an explicit equation of the form (6.25) for the algebraic integers $\lambda(1_C)$.)

Exercise 6.10.9. Prove that any irreducible representation V of a group G is equivalent to a left ideal $\mathbb{C}[G]e$, where e is an idempotent, that is, $e^2 = e$.
Hint: Choose a basis for V so that $\mathrm{End}(V)$ becomes a matrix algebra. Identify V with the left ideal $\mathrm{End}(V)E$, where E is (say) the matrix with 1 in the upper left corner and 0 elsewhere. Then apply the equivalence (6.18) to get a corresponding statement in the group algebra.

Chapter 7

Representations of the symmetric group

7.1 Statement of results

In this chapter, we describe all irreducible representations of the group S_n and compute the character table. We will state and explain the main results before going into the proofs.

As we have seen in §1.6, there is a bijection between conjugacy classes of S_n and partitions of n, that is, sequences $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$ of positive integers with $\sum_j \lambda_j = n$. We will also use partitions to label the irreducible representations. It will be convenient to picture partitions as *Young diagrams*. To explain this, the following example should be sufficient:

$$\lambda = (4,2,1) \qquad \begin{array}{|c|c|c|c|}\hline \multicolumn{1}{|c|}{} & \multicolumn{1}{|c|}{} & \multicolumn{1}{|c|}{} & \multicolumn{1}{|c|}{} \\ \cline{1-2} \multicolumn{1}{|c|}{} & \multicolumn{1}{|c|}{} \\ \cline{1-1} \multicolumn{1}{|c|}{} \\ \cline{1-1} \end{array} \,. \tag{7.1}$$

As a starting point, we consider the action of S_n on words with n letters. A word is an ordered sequence of elements from an alphabet, which for our purposes can be any set of cardinality at least n; we will usually work with the positive integers. Explicitly, this action is[1]

$$\sigma(x_1 \cdots x_n) = x_{\sigma^{-1}(1)} \cdots x_{\sigma^{-1}(n)}. \tag{7.2}$$

When λ is a partition of n, let u_λ be the word

$$u_\lambda = \underbrace{1 \cdots 1}_{\lambda_1} \underbrace{2 \cdots 2}_{\lambda_2} \cdots \underbrace{m \cdots m}_{\lambda_m}. \tag{7.3}$$

The group S_n acts on the set of all permutations of u_λ. As was discussed in §6.3, this gives rise to a permutation representation U_λ on the vector space consisting of

[1]The inverses are necessary to get a left action. A permutation σ acts by moving the i-th letter to the $\sigma(i)$-th position, hence the i-th letter in $\sigma(x)$ is the $\sigma^{-1}(i)$-th letter in x.

formal complex linear combinations of these words. For instance, if $\lambda = (2,1)$, this is a vector space with basis 211, 121, 112. It is easy to see that it is equivalent to the defining representation of S_3 on $\mathbb{C}^3$. As this example shows, the representation U_λ is typically not irreducible.

When λ is a partition, its *transpose* λ' is the partition obtained by reflecting the Young diagram in the main diagonal. For instance, the partition $\lambda = (4,2,1)$ in (7.1) has transpose

$$\lambda = (3,2,1,1) \qquad \begin{array}{|c|c|c|}\hline & & \\ \hline & \\ \cline{1-2} \\ \cline{1-1} \\ \cline{1-1} \end{array} \quad .$$

Recall that the sign of a permutation gives a homomorphism $\mathrm{sgn} : \mathrm{S}_n \to \mathbb{C}$. The tensor product $\mathrm{sgn} \otimes U_\lambda$ can be identified with U_λ as a vector space, but the group acts on the basis vectors by

$$\sigma(x_1 \cdots x_n) = \mathrm{sgn}(\sigma) x_{\sigma^{-1}(1)} \cdots x_{\sigma^{-1}(n)}.$$

We will show that there is a unique irreducible representation V_λ that appears as a subrepresentation of both U_λ and $\mathrm{sgn} \otimes U_{\lambda'}$. Moreover, any irreducible representation appears in this way.

Theorem 7.1.1. *The representations V_λ, indexed by partitions λ of n, form a complete set of irreducible representations of S_n.*

The description of V_λ that we just gave is rather implicit. We will discuss several ways to realize V_λ more explicitly. One useful realization uses the *Young symmetrizer* c_λ. To describe it, label the boxes of the Young diagram of λ with the symbols $1, \ldots, n$, as in the following example (this is called the *canonical tableau*):

$$\begin{array}{|c|c|c|c|}\hline 1 & 2 & 3 & 4 \\ \hline 5 & 6 \\ \cline{1-2} 7 \\ \cline{1-1} \end{array} \quad . \tag{7.4}$$

Let P_λ denote the subgroup of S_n consisting of permutations fixing the symbols in each row and Q_λ the subgroup fixing the symbols in each column (note that P_λ is the stabilizer of the word u_λ). In our example, $P_\lambda \simeq S_4 \times S_2$ consists of permutations of $\{1, \ldots, 7\}$ fixing $\{1,2,3,4\}$, $\{5,6\}$ and $\{7\}$ and $Q_\lambda \simeq S_3 \times S_2$ fixes $\{1,5,7\}$, $\{2,6\}$, $\{3\}$ and $\{4\}$. Now let

$$c_\lambda = \sum_{p \in P_\lambda} p \sum_{q \in Q_\lambda} \mathrm{sgn}(q) q \in \mathbb{C}[\mathrm{S}_n]. \tag{7.5}$$

Then, the representation V_λ is equivalent to the principal left ideal $\mathbb{C}[\mathrm{S}_n] c_\lambda$ in the group algebra.

The following result describes the character table of S_n.

Theorem 7.1.2. *Let λ be a partition of n with m parts, let χ_λ be the character of the irreducible representation V_λ and C the conjugacy class with cycle structure $(c_1, \ldots, c_n)$. Then, $\chi_\lambda(C)$ is the coefficient of*

$$x_1^{\lambda_1+m-1} x_2^{\lambda_2+m-2} \cdots x_m^{\lambda_m}$$

in the polynomial

$$\prod_{1\le i<j\le m} (x_i - x_j) \prod_{k=1}^{n} (x_1^k + \cdots + x_m^k)^{c_k}.$$

One can use Theorem 7.1.2 to compute the dimension of V_λ. There is an attractive way to express the result in terms of *hook lengths*. The hook length of a box in a Young diagram is the number of boxes in the hook consisting of the box itself and all boxes that are straight below it or straight to the right of it.

Corollary 7.1.3. *We have $\dim V_\lambda = n!/H$, where H is the product of the hook lengths of all the boxes in the Young diagram for λ.*

As an example, in the following diagram each box is labelled by its hook length. The corresponding irreducible representation of S_7 has dimension $7!/6\cdot4\cdot3\cdot2 = 35$.

6	4	2	1
3	1		
1			

(7.6)

The Young symmetrizers c_λ also allow us to answer a question raised in §4.3. Let V be a finite-dimensional complex vector space and let S_n act on $V^{\otimes n}$ by

$$\pi(\sigma)(v_1 \otimes \cdots \otimes v_n) = v_{\sigma^{-1}(1)} \otimes \cdots \otimes v_{\sigma^{-1}(n)}.$$

We want to understand how $V^{\otimes n}$ decomposes under this action. We have constructed two invariant subspaces, equivalent to $V^{\odot n}$ and $V^{\wedge n}$. As S_n acts trivially on $V^{\odot n}$, it is isomorphic to a multiple of the trivial representation. Similarly, $V^{\wedge n}$ is a multiple of the alternating representation. When $n \ge 3$ and $\dim(V) \ge 2$ there are additional components corresponding to higher-dimensional representations of S_n. They can be constructed as $V^\lambda = \pi(c_\lambda)V^{\otimes n}$. The cases $\lambda = (n)$ and $\lambda = (1, \ldots, 1)$ correspond to $V^{\odot n}$ and $V^{\wedge n}$, respectively.

To understand the situation, it is useful to consider the representation ρ of $\mathrm{GL}(V)$ on $V^{\otimes n}$ by

$$\rho(A)(v_1 \otimes \cdots \otimes v_n) = Av_1 \otimes \cdots \otimes Av_n. \tag{7.7}$$

It commutes with the action of S_n, which implies that V^λ is a representation of $\mathrm{GL}(V)$. The following result is proved in §7.4.

Theorem 7.1.4. *As representations of* $\mathrm{S}_n \times \mathrm{GL}(V)$*,*

$$V^{\otimes n} \simeq \bigoplus_{\lambda} V_\lambda \otimes V^\lambda, \tag{7.8}$$

where the sum runs over partitions of n*.*

It follows that

$$V^{\otimes n} \simeq \bigoplus_{\lambda} \dim(V^\lambda) V_\lambda \simeq \bigoplus_{\lambda} \dim(V_\lambda) V^\lambda \tag{7.9}$$

as a representation of S_n and $\mathrm{GL}(V)$, respectively. The symmetric role played by the groups S_n and $\mathrm{GL}(V)$ is known as *Schur–Weyl duality*.

Note that it can happen that $V^\lambda = \{0\}$, so that the corresponding term is absent from (7.8). For instance, $V^{(1,\dots,1)} = V^{\wedge n} = 0$ when $\dim V < n$.

Theorem 7.1.5. *The space* $V^\lambda = \{0\}$ *if and only if the number of parts in* λ *is larger than* $\dim(V)$*. Otherwise,* V^λ *is an irreducible representation of* $\mathrm{GL}(V)$*.*

Studying the representations V^λ takes us into the representation theory of Lie groups[2]. Just as for finite groups, characters are an indispensable tool in this theory. We will compute the character of V^λ. By this, we mean an explicit expression for $\chi^\lambda(A) = \mathrm{Tr}_{V^\lambda}(\rho(A))$ in terms of the eigenvalues of $A \in \mathrm{GL}(V)$.

Our expression for the trace is given in terms of certain symmetric polynomials known as *Schur polynomials*. They are given by

$$S_\lambda(x_1, \dots, x_m) = \frac{\det_{1 \le i,j \le m}(x_i^{\lambda_j + m - j})}{\prod_{1 \le i < j \le m}(x_i - x_j)}. \tag{7.10}$$

Here, λ is a partition with at most m parts. If the number of parts k is strictly less than m we should interpret $\lambda_{k+1} = \cdots = \lambda_m = 0$ in (7.10).

Theorem 7.1.6. *If* $\dim(V) = m$*,* λ *is a partition with at most* m *parts and* $A \in \mathrm{GL}(V)$ *has eigenvalues* $x_1, \dots, x_m$ *(counted with multiplicity), then* $\chi^\lambda(A) = S_\lambda(x_1, \dots, x_m)$*.*

As an application of Theorem 7.1.6, we can compute the dimension of V^λ.

Corollary 7.1.7. *If* V *is a vector space of dimension* m*, then*

$$\dim V^\lambda = \prod_{i,j} \frac{m - i + j}{h(i,j)}.$$

Here, the product is over all boxes in the Young diagram of λ*, where the box in row* i *and column* j *has coordinates* (i,j)*, and* $h(i,j)$ *denotes the hook length of that box.*

[2]A Lie group is a group that is also a smooth manifold, such that the group operations are smooth functions.

As an example, in the following Young diagram, we choose $m = 3$ and label each box by $3 - i + j$.

3	4	5	6
2	3		
1			

.

Comparing with (7.6), the corresponding representation V^λ has dimension

$$\frac{6 \cdot 5 \cdot 4 \cdot 3^2 \cdot 2}{6 \cdot 4 \cdot 3 \cdot 2} = 15.$$

Note that if λ has more than m parts, then some boxes will be labelled 0, so Corollary 7.1.7 agrees with the first part of Theorem 7.1.5.

Although any irreducible representation of S_n is equivalent to some V_λ, the group $\mathrm{GL}(V)$ has other representations than V^λ. We will describe without proof how to construct the remaining representations. We first note that, if $k \in \mathbb{Z}$, then $\det^k(g) = \det(g)^k$ defines a one-dimensional representation of $\mathrm{GL}(V)$. If $\lambda_1 \geq \cdots \geq \lambda_m$ is a sequence of integers (some or all of which may be negative), we define

$$W^{(\lambda_1,\ldots,\lambda_m)} = \det{}^{-k} \otimes V^{(\lambda_1+k,\ldots,\lambda_m+k)},$$

where k is large enough so that $\lambda_m + k \geq 0$. We claim that the result is independent of k. Moreover, any finite-dimensional continuous[3] irreducible representation of $\mathrm{GL}(V)$ is equivalent to one of the representations $W^{(\lambda)}$.

Exercise 7.1.1. Use the hook length formula to compute the dimension of all irreducible representations of S_5. Show that the result agrees with the degree equation (6.9).

Exercise 7.1.2. Compute the dimensions of all the spaces involved in (7.8) when $\dim(V) = 2$ and $n = 5$.

Exercise 7.1.3. Use Theorem 7.1.2 to verify that $V_{(n-1,1)}$ is the standard representation.

7.2 Irreducible representations

When λ is a partition of n, we will denote by X_λ the set of all permutations of the word u_λ defined in (7.3), and by U_λ the corresponding permutation representation of S_n, realized on the complex vector space with basis X_λ. We want to show that

$$\dim \mathrm{Hom}_{\mathrm{S}_n}(\mathrm{sgn} \otimes U_{\lambda'}, U_\lambda) = 1. \tag{7.11}$$

[3]As $\mathrm{GL}(V)$ is a topological group, it is natural to require that a representation on W is continuous as a map $\mathrm{GL}(V) \to \mathrm{GL}(W)$.

It then follows from Schur's lemma (see also Exercise 6.2.3) that there is a unique irrreducible representation that appears as a component of both U_λ and $\mathrm{sgn} \otimes U_{\lambda'}$. We will denote this representation by V_λ.

By Exercise 6.1.1 and Exercise 6.2.2, the space of intertwiners (7.11) is equivalent to the space of invariants in $\mathrm{sgn}^* \otimes U_{\lambda'}^* \otimes U_\lambda \simeq \mathrm{sgn} \otimes U_\lambda \otimes U_{\lambda'}$. More generally, we will consider the invariants in $W_{\lambda\mu} = \mathrm{sgn} \otimes U_\lambda \otimes U_\mu$, where λ and μ are two arbitrary partitions of n. As a vector space, $W_{\lambda\mu}$ is spanned by pairs of words $(x, y) \in X_\lambda \times X_\mu$. Tensoring with sgn means that S_n acts on the basis elements by $\pi(\sigma)(x, y) = \mathrm{sgn}(\sigma)(\sigma(x), \sigma(y))$.

Suppose $(x, y) \in X_\lambda \times X_\mu$ is such that the pairs $(\sigma(x), \sigma(y))$ are distinct for distinct $\sigma \in \mathrm{S}_n$. By definition, the set $\{(\sigma(x), \sigma(y));\, \sigma \in \mathrm{S}_n\}$ is then a *free orbit* . Equivalently, if we write x and y on top each other, then the words obtained by reading down columns are all distinct. For instance if $\lambda = (2, 2)$ and $\mu = (2, 1, 1)$ we can take $x = 1122$, $y = 1231$. Since the columns in

$$\begin{matrix} 1 & 1 & 2 & 2 \\ 1 & 2 & 3 & 1 \end{matrix} \tag{7.12}$$

are all distinct, (x, y) is in a free orbit. On the other hand, if $x = 1122$ and $y = 1123$ we get

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

In this case, $(x, y) = (\sigma(x), \sigma(y))$ for the permutation $\sigma = (12)$, so the orbit of (x, y) is not free.

Lemma 7.2.1. *Let λ and μ be partitions of n. Then, the dimension of the space of invariant vectors $W_{\lambda\mu}^{\mathrm{S}_n}$ equals the number of free orbits in $X_\lambda \times X_\mu$.*

We stress that this statement involves two distinct group actions: the notion of free orbits refers to the diagonal action $\sigma(x, y) = (\sigma(x), \sigma(y))$ of S_n on $X_\lambda \times X_\mu$, whereas the notion of invariant vectors refers to the action $\pi(\sigma)(x, y) = \mathrm{sgn}(\sigma)(\sigma(x), \sigma(y))$ on the vector space spanned by $X_\lambda \times X_\mu$.

Proof. By Lemma 6.1.2, we can project to the space of invariants by averaging over the group action. That is, the space of invariants is spanned by the elements

$$P(x, y) = \frac{1}{n!} \sum_{\sigma \in \mathrm{S}_n} \mathrm{sgn}(\sigma)(\sigma(x), \sigma(y)) \tag{7.13}$$

for $(x, y) \in X_\lambda \times X_\mu$. The intertwining property of P is

$$P(\tau(x), \tau(y)) = \mathrm{sgn}(\tau) P(x, y). \tag{7.14}$$

If the orbit of (x, y) is not free, then there is a transposition τ with $(\tau(x), \tau(y)) = (x, y)$, and (7.14) gives $P(x, y) = 0$. If the orbit of (x, y) is free, then the basis elements $(\sigma(x), \sigma(y))$ on the right-hand side of (7.13) are mutually distinct, so $P(x, y) \neq 0$. In this case, (7.14) shows that P maps the orbit into a one-dimensional subspace. Moroever, if (x_0, y_0) is in a free orbit distinct from that of (x, y), then the basis elements $(\sigma(x), \sigma(y))$ and $(\tau(x_0), \tau(y_0))$ are mutually distinct for all σ, $\tau \in \mathrm{S}_n$. This shows that the projections of the free orbits are linearly independent one-dimensional subspaces, so we can form a basis for $W_{\lambda\mu}^{\mathrm{S}_n}$ consisting of one element for each free orbit. □

Many fundamental notions related to the representation theory of S_n are conveniently understood in terms of *tableaux*[4], that is, labelled Young diagrams. When λ and μ are two partitions of n, a tableau of *shape* λ and *weight* μ is a labelling of the boxes in the Young diagram of λ by the letters from u_μ. A tableau is *row-strict* if the labels are strictly increasing from left to right in each row.

Lemma 7.2.2. *There is a bijection between free orbits in $X_\lambda \times X_\mu$ and row-strict tableaux of shape λ and weight μ. Given an element (x, y) in a free orbit, the corresponding tableau is obtained by writing x above y and then inserting the letters from y that appear under the letter k into the k-th row, in increasing order.*

To prove this one must verify that the tableau described does not depend on the choice of the element (x, y), and that each row-strict tableaux of shape λ and weight μ can be obtained in this way from a unique free orbit. Both these statements should be obvious.

As an example, let $\lambda = (3, 2)$ and $\mu = (2, 1, 1, 1)$ and consider the pair $x = 12121$, $y = 12413$. Since all the columns in

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

are distinct, (x, y) generates a free orbit. It corresponds to the tableau

$$\begin{array}{|c|c|c|} \hline 1 & 3 & 4 \\ \hline 1 & 2 \\ \cline{1-2} \end{array} \quad .$$

We will need the *dominance order* of partitions. To this end, we first extend any partition $\lambda_1 \geq \cdots \geq \lambda_m$ to an infinite sequence by defining $\lambda_{m+1} = \lambda_{m+2} = \cdots = 0$. With this convention, we write $\mu \leq \lambda$ if

$$\mu_1 + \cdots + \mu_j \leq \lambda_1 + \cdots + \lambda_j, \qquad j = 1, 2, 3, \ldots.$$

[4] *Tableau* is the singular form and *tableaux* plural.

Clearly, if μ has m rows it is enough to postulate this for $j \leq m$. This is a partial but not a total order. For instance, $\lambda = (3,3)$ and $\mu = (4,1,1)$ are not comparable in the dominance order, that is, neither $\lambda \leq \mu$ nor $\mu \leq \lambda$ holds.

Before we proceed, we say a few words on the *lexicographic order* on partitions, which will appear in §7.6. This is the same order that we use to compare words in a lexicon; to order ABACUS and ABANDON, we observe that the first three letters agree and then compare the fourth letter. For instance, the partitions of 5 are ordered as

$$(5) \underset{\text{lex}}{>} (4,1) \underset{\text{lex}}{>} (3,1,1) \underset{\text{lex}}{>} (2,2,1) \underset{\text{lex}}{>} (2,1,1,1) \underset{\text{lex}}{>} (1,1,1,1,1).$$

(In general, we may have to add zeroes at the end so that we can write $(3,1,1) >_{\text{lex}} (3,1,0) = (3,1)$, but when comparing partitions of the same number that is never necessary.) The lexicographic order is total, that is, for two partitions λ and μ we either have $\lambda >_{\text{lex}} \mu$, $\lambda = \mu$ or $\lambda <_{\text{lex}} \mu$. Moreover, it refines the dominance order in the sense that $\lambda \geq \mu$ implies $\lambda \geq_{\text{lex}} \mu$.

We now give a criterion for the existence of free orbits.

Lemma 7.2.3. *Let λ and μ be two partitions of n. If there exists a free orbit in $X_\lambda \times X_\mu$ then $\mu \leq \lambda'$. Moreover, if $\mu = \lambda'$ there is a unique free orbit.*

One can also prove that if $\mu \leq \lambda'$ and $\mu \neq \lambda'$ then there is more than one free orbit.

Proof. Given a free orbit, consider the corresponding row-strict tableau of shape λ and weight μ. Since the labels are strictly increasing along rows, any letter j can only appear in the j leftmost columns. Thus, the $\mu_1 + \cdots + \mu_j$ copies of the letters $1, \ldots, j$ must all appear in the $\lambda'_1 + \cdots + \lambda'_j$ boxes in the j leftmost columns. This gives

$$\mu_1 + \cdots + \mu_j \leq \lambda'_1 + \cdots + \lambda'_j \tag{7.15}$$

for each j, that is, $\mu \leq \lambda'$.

In the special case when $\mu_j = \lambda'_j$ for all j, the same argument shows that there is a unique free orbit, corresponding to the labelling where j only appears in the j-th column. □

Explicitly, when $\mu = \lambda'$, the unique free orbit is generated by

$$\begin{array}{cccccccccccc} 1 & 1 & \cdots & 1 & 2 & 2 & \cdots & 2 & \cdots & m & \cdots & m \\ 1 & 2 & \cdots & \lambda_1 & 1 & 2 & \cdots & \lambda_2 & \cdots & 1 & \cdots & \lambda_m \end{array} . \tag{7.16}$$

Combining Lemma 7.2.1 and Lemma 7.2.3 proves (7.11).

We are now ready to prove our first main result, that is, that the representations V_λ form a complete set of irreducible representations of S_n.

Proof of Theorem 7.1.1. As we know that the irreducible representations of S_n and the partitions of n are equinumerous, it suffices to show that $V_\lambda \not\simeq V_\mu$ for $\lambda \neq \mu$. Suppose that $V_\lambda \simeq V_\mu$. Then, V_λ is a subrepresentation of both U_λ and $\mathrm{sgn} \otimes U_{\mu'}$. This gives $\dim W_{\lambda\mu'}^{\mathrm{S}_n} = \dim \mathrm{Hom}_{\mathrm{S}_n}(U_\lambda, \mathrm{sgn} \otimes U_{\mu'}) > 0$. By Lemma 7.2.3 and Lemma 7.2.1, it follows that $\lambda \leq \mu$. By symmetry, $\mu \leq \lambda$ and we may conclude that $\lambda = \mu$. □

By a variation of the same argument we have the following fact.

Proposition 7.2.4. *The decomposition of U_λ into irreducible representations takes the form*

$$U_\lambda = \sum_{\mu \geq \lambda} K_{\mu\lambda} V_\mu,$$

where $K_{\mu\lambda}$ are non-negative integers with $K_{\lambda\lambda} = 1$.

The numbers $K_{\mu\lambda}$ are called Kostka numbers. We mention without proof that $K_{\mu\lambda}$ is in fact positive for all $\mu \geq \lambda$. Moreover, one can interpret $K_{\mu\lambda}$ combinatorially as the number of *semi-standard tableaux* of shape μ and weight λ. That is, they count the number of ways to insert the letters of u_λ into the Young diagram for μ in such a way that the symbols increase weakly from left to right and strictly from top to bottom. For instance, the multiplicity of $V_{(3,2)}$ in $U_{(2,1,1,1)}$ is 3, the corresponding semi-standard tableaux being

$$\begin{array}{|c|c|c|}\hline 1&1&2\\\hline 3&4\\\cline{1-2}\end{array}\quad,\quad \begin{array}{|c|c|c|}\hline 1&1&3\\\hline 2&4\\\cline{1-2}\end{array}\quad,\quad \begin{array}{|c|c|c|}\hline 1&1&4\\\hline 2&3\\\cline{1-2}\end{array}\quad.$$

Proof of Proposition 7.2.4. Suppose that V_μ appears in the decomposition of U_λ. Since it also appears in $\sigma \otimes U_{\mu'}$, it follows that there is a non-zero intertwiner $\sigma \otimes U_{\mu'} \to U_\lambda$. As we have seen above, there is then a free orbit in $X_{\mu'} \times X_\lambda$, which, by Lemma 7.2.3, can only happen if $\lambda \leq \mu$. Moreover, since there is a unique free orbit in $X_{\lambda'} \times X_\lambda$, we see in the same way that V_λ appears in U_λ with multiplicity 1. □

Exercise 7.2.1. Show that if $\lambda = (1, \ldots, 1)$, then U_λ is equivalent to the regular representation of S_n. Show that if $\lambda = (n-1, 1)$ it is equivalent to the defining representation and if $\lambda = (n)$ to the trivial representation.

Exercise 7.2.2. Show that if $\lambda = (1, \ldots, 1)$, then V_λ is equivalent to the alternating representation, if $\lambda = (n-1, 1)$ to the standard representation and if $\lambda = (n)$ to the trivial representation.

Exercise 7.2.3. Show that $V_\lambda \otimes \mathrm{sgn} \simeq V_{\lambda'}$ as representations.

Exercise 7.2.4. Show that $\lambda \leq \mu$ if and only if $\mu' \leq \lambda'$.

Exercise 7.2.5. Let x and y be two words in English (or some other language that you know) with the same number of letters, and X and Y the vector spaces spanned by permutations of these words. Give examples where (a) there is no free orbit in $X \times Y$; (b) there is more than one free orbit in $X \times Y$; (c) there is a unique free orbit in $X \times Y$.[5]

Exercise 7.2.6. Draw the lattice consisting of all partitions of 6 in the dominance order. Also explain how it looks in the lexicographic order.

7.3 Explicit realizations of representations

Our description of the irreducible representations V_λ is rather implicit. We will give several more concrete descriptions. Recall from §7.1 the subgroups P_λ and Q_λ that fix, respectively, the rows and columns in the canonical tableau on λ. We will need the elements $a_\lambda = \sum_{p \in P_\lambda} p$, $b_\lambda = \sum_{q \in Q} \operatorname{sgn}(q) q$ and $c_\lambda = a_\lambda b_\lambda$ of the group algebra.

We first discuss different realizations of U_λ. Our original construction started from the word u_λ, which has stabilizer P_λ. We wrote X_λ for the orbit of u_λ and then realized U_λ as the space $\mathbb{C}[X_\lambda]$ of formal linear combinations of elements in the orbit (or, equivalently, complex-valued functions on the orbit).

As we discussed in §1.5, if a group G acts on a set X and $x \in X$ has stabilizer G_x, then the orbit $G(x)$ is equivalent as a G-set to G/G_x. In particular, any two such orbits are equivalent. In the case at hand, this shows that if X is any set equipped with an action of S_n and $x \in X$ is any element with stabilizer P_λ, then the orbit $\mathrm{S}_n(x)$ is equivalent to X_λ. Consequently, the permutation representation $\mathbb{C}[\mathrm{S}_n(x)]$ is equivalent to U_λ.

Besides our original construction using words, there are several other natural choices for X and x. Perhaps the most canonical choice is $X = \mathrm{S}_n/P_\lambda$ and $x = P_\lambda$. Then, we obtain a realization of U_λ as the space $\mathbb{C}[\mathrm{S}_n/P_\lambda]$.

Second, we can take $X = \mathbb{C}[\mathrm{S}_n]$ and $x = a_\lambda$. It is very easy to see that the stabilizer of a_λ is indeed P_λ. Thus, U_λ is equivalent to $\mathbb{C}[\mathrm{S}_n a_\lambda] \simeq \mathbb{C}[S_n]a_\lambda$, which is a left ideal in the group algebra.

Third, we can take X to be all monomials in the variables $x_1, \ldots, x_n$. Let $x = p_\lambda = \prod_{j=1}^n x_j^{c_j - 1}$, where c_j is the row number of the box containing j in the canonical tableau. For instance, if $\lambda = (4, 2, 1)$ we have (cf. (7.4))

$$p_\lambda = x_1^0 x_2^0 x_3^0 x_4^0 \cdot x_5^1 x_6^1 \cdot x_7^2 = x_5 x_6 x_7^2. \tag{7.17}$$

[5]Our presentation in this chapter is heavily influenced by J. D. Wiltshire-Gordon, A. Woo and M. Zajaczkowska, Specht polytopes and Specht matroids, in Combinatorial algebraic geometry, Fields Institute, Toronto 2017, pp. 201–228. These authors use the impressive example x =SASSAFRAS*andy*=TENNESSEE to illustrate case (c). Sassafras is a genus of trees native to North America and China.

Again, the stabilizer of p_λ is P_λ, so its orbit spans a representation equivalent to U_λ.

We summarize the preceding discussion as follows.

Proposition 7.3.1. *The representation $U_\lambda = \mathbb{C}[\mathrm{S}_n(u_\lambda)]$ is equivalent to $\mathbb{C}[\mathrm{S}_n/P_\lambda]$, to $\mathbb{C}[\mathrm{S}_n]a_\lambda$ and to $\mathbb{C}[\mathrm{S}_n(p_\lambda)]$, An explicit equivalence between these four realizations of U_λ is given on basis elements by $\sigma(u_\lambda) \mapsto \sigma P_\lambda \mapsto \sigma a_\lambda \mapsto \sigma(p_\lambda)$, $\sigma \in \mathrm{S}_n$.*

Let us now consider the irreducible representation V_λ, viewed as a subrepresentation of U_λ.

Proposition 7.3.2. *Let X be an S_n-set and $x \in X$ an element with stabilizer P_λ, so that $\mathbb{C}[\mathrm{S}_n(x)] \simeq U_\lambda$. Then, the subspace $\pi(\mathbb{C}[\mathrm{S}_n]b_\lambda)x$ is equivalent to V_λ.*

Before proving Proposition 7.3.2, let us see how it works for $\lambda = (3,1)$. We use the realization of U_λ based on words. Then, U_λ is spanned by the words 2111, 1211, 1121 and 1112, which can be identified with the standard basis vectors $e_1, \ldots, e_4$ in the defining representation of S_4 on $\mathbb{C}^4$. As $Q_\lambda = \{\mathrm{id}, (14)\}$, we have

$$\pi(b_\lambda)u_\lambda = u_\lambda - (14)(u_\lambda) = 1112 - 2111 = e_4 - e_1. \tag{7.18}$$

It follows that $\{\pi(\sigma b_\lambda)u_\lambda;\ \sigma \in \mathrm{S}_4\} = \{e_i - e_j;\ 1 \leq i \neq j \leq 4\}$, which clearly spans the subspace $\{x \in \mathbb{C}^4;\ x_1 + x_2 + x_3 + x_4 = 0\}$. In agreement with Exercise 7.1.3 and Exercise 7.2.2, we find that $V_{(3,1)}$ is the standard representation.

Proof of Proposition 7.3.2. As we have seen, it is no restriction to work with the specific realization where X consists of words of length n in the letters $\{1, 2, 3 \ldots, n\}$ and $x = u_\lambda$. We know from (7.11) that the space of intertwiners $\mathrm{sgn} \otimes U_{\lambda'} \to U_\lambda$ is one-dimensional. By definition, the image of any non-zero intertwiner can be identified with V_λ. By the proof of (7.11), to construct such an intertwiner we may start with the pair of words (x, y), where $x = u_\lambda$ and

$$y = 12\cdots\lambda_1 12\cdots\lambda_2\cdots 12\cdots\lambda_m.$$

Then (x, y) is in the unique free orbit in $X_\lambda \times X_{\lambda'}$. By (7.13),

$$\sum_{\sigma\in\mathrm{S}_n} \mathrm{sgn}(\sigma)\, \sigma(x) \otimes \sigma(y) \tag{7.19}$$

is an invariant element of $\mathrm{sgn} \otimes U_\lambda \otimes U_{\lambda'}$.

To identify (7.19) with an intertwiner, we need an isomorphism $\mathrm{sgn} \otimes U_\lambda \otimes U_{\lambda'} \to \mathrm{Hom}(\mathrm{sgn} \otimes U_{\lambda'}, U_\lambda)$. Since U_λ is a permutation representation, we may apply Lemma 6.3.1[6] and get the intertwiner

$$\sum_{\sigma\in\mathrm{S}_n} \mathrm{sgn}(\sigma)\sigma(x)\sigma(y)^*,$$

[6]This Lemma was not included in the previous version of these notes.

which takes an element $\tau(y)$ to

$$\sum_{\sigma\in \mathrm{S}_n} \mathrm{sgn}(\sigma)\sigma(x)\delta_{\tau(y),\sigma(y)}.$$

Since $\sigma(y) = \tau(y)$ if and only if $\sigma \in \tau Q_\lambda$, this can be written

$$\sum_{q\in Q_\lambda} \mathrm{sgn}(\tau q)(\tau q)(x) = \mathrm{sgn}(\tau)\pi(\tau b_\lambda)(x). \tag{7.20}$$

Thus, the image of the intertwiner is indeed $\pi(\mathbb{C}[\mathrm{S}_n]b_\lambda)x$. □

Proposition 7.3.1 gives in particular the following realizations of V_λ as left ideals in the group algebra.

Corollary 7.3.3. *We have the equivalence of representations* $V_\lambda \simeq \mathbb{C}[\mathrm{S}_n]b_\lambda a_\lambda \simeq \mathbb{C}[\mathrm{S}_n]a_\lambda b_\lambda$.

Proof. The first equivalence is a special case of Proposition 7.3.2. For the second equivalence, we note that V_λ and $\mathrm{sgn}\otimes V_{\lambda'}$ can both be described as the unique irreducible representation contained in both U_λ and $\mathrm{sgn}\otimes U_{\lambda'}$. Hence, $V_\lambda \simeq \mathrm{sgn}\otimes V_{\lambda'}$ (cf. Exercise 7.2.3). Moreover, there is an element $\sigma \in \mathrm{S}_n$ such that $\sigma P_{\lambda'}\sigma^{-1} = Q_\lambda$ and $\sigma Q_{\lambda'}\sigma^{-1} = P_\lambda$. The map $g \mapsto \sigma g\sigma^{-1}$ extends to an automorphism of the group algebra. Hence,

$$V_{\lambda'} \simeq \mathbb{C}[\mathrm{S}_n]b_{\lambda'}a_{\lambda'} = \mathbb{C}[\mathrm{S}_n]\sum_{q\in Q_{\lambda'}} \mathrm{sgn}(q)q \sum_{p\in P_{\lambda'}} p \simeq \mathbb{C}[\mathrm{S}_n]\sum_{p\in P_\lambda} \mathrm{sgn}(p)p \sum_{q\in Q_\lambda} q. \tag{7.21}$$

We now use that the map $\phi(\sigma) = \mathrm{sgn}(\sigma)\sigma$ extends to an equivalence of representations $\phi : \mathbb{C}[\mathrm{S}_n] \to \mathrm{sgn}\otimes\mathbb{C}[\mathrm{S}_n]$. Applying ϕ to (7.21) gives

$$V_\lambda \simeq \mathrm{sgn}\otimes V_{\lambda'} \simeq \mathbb{C}[\mathrm{S}_n]\sum_{p\in P_\lambda} p \sum_{q\in Q_\lambda} \mathrm{sgn}(q)q = \mathbb{C}[\mathrm{S}_n]a_\lambda b_\lambda.$$

□

Note that if we instead apply ϕ to

$$U_{\lambda'} \simeq \mathbb{C}[\mathrm{S}_n]a_{\lambda'} \simeq \mathbb{C}[\mathrm{S}_n]\sum_{q\in Q_\lambda} q$$

we get

$$\mathrm{sgn}\otimes U_{\lambda'} \simeq \mathbb{C}[\mathrm{S}_n]b_\lambda. \tag{7.22}$$

The left ideals in Corollary 7.3.3 are of a special type, namely, ideals generated by idempotents. This is not surprising as we saw in Exercise 6.10.9 that *any* irreducible representation of a finite group is generated by an idempotent in the group algebra. In the case at hand, note first that a_λ is proportional to an idempotent. Indeed, if we write $a_\lambda^2 = \sum_{p_1,p_2\in P_\lambda} p_1p_2$ and replace p_2 by $p_1^{-1}p_2$ we get $a_\lambda^2 = |P_\lambda|a_\lambda$.

We will need the following fact, which we formulate in a more general setting.

Lemma 7.3.4. *If R is a ring, $e \in R$ an idempotent and M an R-module, then*

$$\mathrm{Hom}_R(Re, M) \simeq eM, \tag{7.23}$$

as additive groups. Here, $x \in eM$ corresponds to the homomorphism $y \mapsto yx$.

Proof. Let $\phi \in \mathrm{Hom}_R(Re, M)$ and let $x = \phi(e)$. Then $x = \phi(e^2) = e\phi(e) = ex$, so $x \in eM$. Moreover, for any $y \in Re$ we have $y = ye$ and hence $\phi(y) = y\phi(e) = yx$. Conversely, if $x \in eM$, then $\phi(y) = yx$ is clearly an R-module homomorphism. □

In particular, when $R = \mathbb{C}[G]$ we find that

$$\mathrm{Hom}_G\left(\mathbb{C}[G]e, \mathbb{C}[G]f\right) \simeq e\mathbb{C}[G]f. \tag{7.24}$$

This holds also as an isomorphism of complex vector spaces (more generally, if R is an associative algebra over a commutative ring S and M a representation of R, that is, an S-module equipped with an algebra homorphism $R \to \mathrm{End}_S(M)$, then (7.23) holds as an isomorphism of S-modules). We can then prove the following fact.

Proposition 7.3.5. *The element $c_\lambda = a_\lambda b_\lambda$ is proportional to an idempotent.*

Proof. By (7.22) and (7.24),

$$\mathrm{Hom}_{\mathrm{S}_n}(U_\lambda, \mathrm{sgn} \otimes U_{\lambda'}) \simeq \mathrm{Hom}_{\mathrm{S}_n}(\mathbb{C}[\mathrm{S}_n]a_\lambda, \mathbb{C}[\mathrm{S}_n]b_\lambda) \simeq a_\lambda\mathbb{C}[\mathrm{S}_n]b_\lambda. \tag{7.25}$$

As we know that this space is one-dimensional and contains c_λ, it must equal $\mathbb{C}c_\lambda$. That is,

$$a_\lambda x b_\lambda \in \mathbb{C}a_\lambda b_\lambda, \qquad x \in \mathbb{C}[\mathrm{S}_n]. \tag{7.26}$$

In particular, if $x = b_\lambda a_\lambda$ we find that $c_\lambda^2 \in \mathbb{C}c_\lambda$. It remains to show that $c_\lambda^2 \neq 0$.

To complete the proof, consider the matrix A representing c_λ in the regular representation. Explicitly, if $c_\lambda = \sum_{g \in \mathrm{S}_n} a_g g$, then $A = (a_{hg^{-1}})_{g,h \in G}$. In particular, $\mathrm{Tr}(A) = n! a_{\mathrm{id}}$. We claim that $a_{\mathrm{id}} = 1$. Indeed, the only contribution to this coefficient comes from the terms in (7.5) with $p = q^{-1} \in P_\lambda \cap Q_\lambda = \{\mathrm{id}\}$. Thus, $\mathrm{Tr}(A) = n!$. But if $c_\lambda^2 = 0$, then $A^2 = 0$. It would then follow that all eigenvalues of A were zero and, in particular, $\mathrm{Tr}(A) = 0$. □

So far, we have not used the final realization mentioned in Proposition 7.3.1. This gives a very concrete realization of the representations V_λ, due to Specht[7]. Recall that we put $p_\lambda = \prod_{j=1}^n x_j^{c_j - 1}$, where c_j is the row number of the box containing j in the canonical tableau. The orbit of p_λ under permutation of the

[7]The representations V_λ are often called *Specht modules*.

variables spans a representation equivalent to U_λ. By Proposition 7.3.2, V_λ is the subrepresentation generated by

$$\pi(b_\lambda)p_\lambda = \sum_{q\in Q_\lambda} \operatorname{sgn}(q) \prod_{j=1}^{n} x_{q(j)}^{c_j-1}.$$

We can factor $q = q_1 q_2 \dots q_{\lambda_1}$, where q_j acts by permuting the elements in column j of the canonical tableau. In the example (7.17), we get

$$\sum_{q_1\in \mathrm{S}_{\{1,5,7\}}} \operatorname{sgn}(q_1) x^0_{q_1(1)} x^1_{q_1(5)} x^2_{q_1(7)} \sum_{q_2\in \mathrm{S}_{\{2,6\}}} \operatorname{sgn}(q_2) x^0_{q_2(2)} x^1_{q_2(6)};$$

the factors corresponding to q_3 and q_4 are trivial. This can be written as the product of determinants

$$\begin{vmatrix} 1 & x_1 & x_1^2 \\ 1 & x_5 & x_5^2 \\ 1 & x_7 & x_7^2 \end{vmatrix} \cdot \begin{vmatrix} 1 & x_2 \\ 1 & x_6 \end{vmatrix}.$$

By the Vandermonde determinant evaluation (see Lemma 7.5.3 below), this can be evaluated as

$$(x_5 - x_1)(x_7 - x_1)(x_7 - x_5) \cdot (x_6 - x_2).$$

For a general partition, one similarly obtains the *Specht polynomial* $\prod(x_j - x_i)$, the product running over all pairs $i < j$ such that i and j are in the same column of the canonical tableau. It follows that V_λ can be realized as the span of all polynomials obtained by permuting the variables in the Specht polynomial.

As a simple example, when $\lambda = (n-1, 1)$ the Specht polynomial is $x_1 - x_n$. Permuting the variables we obtain the polynomials $x_i - x_j$, $i \neq j$. These clearly span the space $\{a_1x_1 + \dots + a_nx_n; a_1 + \dots + a_n = 0\}$. Identifying the polynomial $a_1x_1 + \dots + a_nx_n$ with the vector $(a_1, \dots, a_n)$, we again see that $V_{(n-1,1)}$ is the standard representation.

Proposition 7.3.2 describes V_λ explicitly as the span of the elements $\pi(\sigma b_\lambda)x$, $\sigma \in \mathrm{S}_n$. This spanning set is typically redundant. It would be nice to have a smaller set of permutations that generate a basis for V_λ. This can be achieved and the answer is again given in terms of tableaux. When λ is a partition of n, a *standard tableau* is a labelling of the boxes of the Young diagram for n with the numbers $1, 2, \dots, n$, such that the labels increase strictly from left to right and from top to bottom. For instance, the standard tableaux of shape $(3, 1)$ are

$$\begin{array}{|c|c|c|}\hline 1 & 2 & 3 \\ \hline 4 \\ \cline{1-1} \end{array}, \qquad \begin{array}{|c|c|c|}\hline 1 & 2 & 4 \\ \hline 3 \\ \cline{1-1} \end{array}, \qquad \begin{array}{|c|c|c|}\hline 1 & 3 & 4 \\ \hline 2 \\ \cline{1-1} \end{array}. \tag{7.27}$$

Theorem 7.3.6. *Let, as in* Proposition 7.3.2*, V_λ be realized as $\pi(\mathbb{C}[\mathrm{S}_n]b_\lambda)x$. Then, a basis for V_λ is given by $\pi(\sigma b_\lambda)x$, where σ runs over permutations such that, if one replaces the number j in the canonical tableau everywhere with $\sigma(j)$, one obtains a standard tableau.*

The proof of Theorem 7.3.6 is not included in these notes.

As an example, the tableaux (7.27) correspond to the permutations

$$(\sigma(1),\sigma(2),\sigma(3),\sigma(4)) = (1,2,3,4), \quad (1,2,4,3), \quad (1,3,4,2),$$

that is, to $\sigma = \mathrm{id}$, $\sigma = (34)$ and $\sigma = (234)$. By (7.18), $\pi(b_\lambda)u_\lambda$ can be identified with the vector $e_4 - e_1 = (-1,0,0,1)$ in $V_{(3,1)} = \{x \in \mathbb{C}^4;\, x_1 + \cdots + x_4 = 0\}$. Corollary 7.3.6 asserts that a basis for $V_{(3,1)}$ is given by $e_{\sigma(4)} - e_{\sigma(1)}$ with σ as above, that is, by $e_4 - e_1$, $e_3 - e_1$ and $e_2 - e_1$.

Exercise 7.3.1. Prove that $\phi(x) = xa_\lambda$ defines an equivalence of representations $\phi : \mathbb{C}[\mathrm{S}_n]a_\lambda b_\lambda \to \mathbb{C}[\mathrm{S}_n]b_\lambda a_\lambda$, with inverse $\phi^{-1}(x) = Cxb_\lambda$ for some constant C.

Exercise 7.3.2. Prove that $c_\lambda x c_\mu = 0$ if $\lambda \neq \mu$ and $x \in \mathbb{C}[\mathrm{S}_n]$.

Exercise 7.3.3. Prove that $a_\lambda \mathbb{C}[\mathrm{S}_n]b_\mu = 0$, $\lambda \not\leq \mu$.

7.4 Schur–Weyl duality

In this section we will prove Theorem 7.1.4 and start working towards a proof of Theorem 7.1.5. We recall the general setup. Let V be a complex vector space and consider the representations of S_n and $\mathrm{GL}(V)$ on $V^{\otimes n}$ by

$$\pi(\sigma)(v_1 \otimes \cdots \otimes v_n) = v_{\sigma^{-1}(1)} \otimes \cdots \otimes v_{\sigma^{-1}(n)},$$
$$\rho(A)(v_1 \otimes \cdots \otimes v_n) = Av_1 \otimes \cdots \otimes Av_n.$$

We will use the same symbols to denote the linear extension $\pi : \mathbb{C}[\mathrm{S}_n] \to \mathrm{End}(V^{\otimes n})$ and the natural extension $\rho : \mathrm{End}(V) \to \mathrm{End}(V^{\otimes n})$ defined by the same identity. (Note that ρ is not linear, so it is not a representation of the algebra $\mathrm{End}(V)$.) For λ a partition of n, we define $V^\lambda = \pi(c_\lambda)V^{\otimes n}$. Since $\pi(\sigma)$ and $\rho(A)$ always commute, V^λ a representation of $\mathrm{GL}(V)$. Note that $V^{(n)} = V^{\odot n}$ and $V^{(1,\dots,1)} = V^{\wedge n}$, considered as subspaced of $V^{\otimes n}$.

Theorem 7.1.4 asserts that

$$V^{\otimes n} \simeq \bigoplus_\lambda V_\lambda \otimes V^\lambda \tag{7.28}$$

as a representation of $\mathrm{S}_n \times \mathrm{GL}(V)$. This is now easy to prove. Indeed, since $V_\lambda \simeq \pi[\mathrm{S}_n]c_\lambda$ it follows from Lemma 7.3.4 that $\mathrm{Hom}_{\mathrm{S}_n}(V_\lambda, V^{\otimes n}) \simeq V^\lambda$, where

$x \in V^\lambda$ corresponds to the homomorphism $y \mapsto \pi(y)x$ from V_λ to $V^{\otimes n}$. This shows that the multiplicity of V_λ in $V^{\otimes n}$ is $\dim(V^\lambda)$ (see Problem 6.2.4), so (7.28) holds as representations of S_n. Moreover, as the embedding $y \otimes x \mapsto \pi(y)x$ is intertwining also for the $\mathrm{GL}(V)$ action, the proof of Theorem 7.1.4 is complete.

The proof of Theorem 7.1.5 is more involved. Below, we will prove that V^λ is either zero or irreducible. To determine exactly when it is zero has to wait until the end of §7.6.

We know that $\mathrm{GL}(V)$ commutes with the action of S_n, that is, $\rho(\mathrm{GL}(V)) \subseteq \mathrm{End}_{\mathrm{S}_n}(V^{\otimes n})$. We will show that $\rho(\mathrm{GL}(V))$ in fact spans $\mathrm{End}_{\mathrm{S}_n}(V^{\otimes n})$. Thus, if a subspace is closed under the action of $\mathrm{GL}(V)$ it must be closed under $\mathrm{End}_{\mathrm{S}_n}(V^{\otimes n})$. We will then use (7.28) to show that V^λ has no non-trivial invariant subspaces.

Note that, by the natural isomorphism $\mathrm{End}(V^{\otimes n}) \simeq \mathrm{End}(V)^{\otimes n}$ we have

$$\mathrm{End}_{\mathrm{S}_n}(V^{\otimes n}) = \mathrm{End}(V^{\otimes n})^{\mathrm{S}_n} \simeq (\mathrm{End}(V)^{\otimes n})^{\mathrm{S}_n} \simeq \mathrm{End}(V)^{\odot n}. \tag{7.29}$$

Thus, we want to prove that the elements $\rho(A) = A^{\otimes n}$, $A \in \mathrm{GL}(V)$, span $\mathrm{End}(V)^{\odot n}$. This will follow from the following fact.

Lemma 7.4.1. *If V is a finite-dimensional complex vector space, then $V^{\odot n}$ is spanned by $v \otimes \cdots \otimes v$, $v \in V$.*

Proof. We find it easiest to understand this using the identification of symmetric tensors with polynomials. Namely, if $(e_j)_{j=1}^m$ is a basis for V, then $e_{j_1} \odot \cdots \odot e_{j_m} \mapsto x_{j_1} \cdots x_{j_m}$ gives an isomorphism from $V^{\odot n}$ to the space of homogeneous polynomials of degree n in $x_1, \ldots, x_m$. Lemma 7.4.1 asserts that these polynomials are spanned by all powers $f = (a_1x_1 + \cdots + a_mx_m)^n$. Let U be the span of all such powers. Considering $f = f(a_1)$ as a function of a_1, we have $h^{-1}(f(a_1 + h) - f(a_1)) \in U$. By continuity, $\partial f/\partial a_1 \in U$. Repeating this argument, any derivative $\partial^{l_1+\cdots+l_m} f/\partial a_1^{l_1} \cdots \partial a_m^{l_m} \in U$. But if $l_1 + \cdots + l_m = n$, this is a non-zero multiple of $x_1^{l_1} \cdots x_m^{l_m}$, which clearly span the homogeneous polynomials. □

Corollary 7.4.2. *The space $\mathrm{End}_{\mathrm{S}_n}(V^{\otimes n})$ is spanned by $\rho(\mathrm{GL}(V))$.*

Proof. By (7.29) and Lemma 7.4.1, $\mathrm{End}_{\mathrm{S}_n}(V^{\otimes n})$ is spanned by $\rho(\mathrm{End}(V))$. But $\mathrm{GL}(V)$ is dense in $\mathrm{End}(V)$ (if $\det(A) = 0$ there is an arbitrarily small perturbation of A with non-zero determinant), so $\rho(\mathrm{End}(V))$ and $\rho(\mathrm{GL}(V)$ span the same subspace. □

Finally, we need the following fact.

Lemma 7.4.3. *Any $A \in \mathrm{End}(V^\lambda)$ can be extended to an element of $\mathrm{End}_{\mathrm{S}_n}(V^{\otimes n})$.*

Proof. Applying the isomorphism (7.28), one may define the extension of A to be $\mathrm{id} \otimes A$ on the term corresponding to λ and zero else. □

More generally, it is easy to see that

$$\mathrm{End}_{\mathrm{S}_n}(V^{\otimes n}) = \bigoplus_{\lambda,\mu} \mathrm{Hom}_{\mathrm{S}_n}(V_\lambda \otimes V^\lambda, V_\mu \otimes V^\mu) \simeq \bigoplus_{\lambda} \mathrm{End}(V^\lambda).$$

This can be compared with (6.18), which tells us that

$$\mathbb{C}[\mathrm{S}_n] \simeq \bigoplus_{\lambda} \mathrm{End}(V_\lambda).$$

We can now prove the following weak version of Theorem 7.1.5.

Corollary 7.4.4. *As a representation of* $\mathrm{GL}(V)$, V^λ *is either zero or irreducible.*

Proof. Suppose that $U \subseteq V^\lambda$ is closed under $\rho(\mathrm{GL}(V))$. By Corollary 7.4.2, U is closed under $\mathrm{End}_{\mathrm{S}_n}(V^{\otimes n})$ and hence, by Lemma 7.4.3, under $\mathrm{End}(V^\lambda)$. But there is no non-trivial invariant subspace under $\mathrm{End}(V^\lambda)$, since any non-zero vector is mapped to any other non-zero vector by some linear transformation. □

As an example, consider the decomposition of $V^{\otimes 3}$. The irreducible representations of S_3 are the trivial representation $V_{(3)}$, the alternating representation $V_{(1,1,1)}$ and the standard representation $V_{(2,1)}$. We already now that the first two components correspond to $V^{\odot 3}$ and $V^{\wedge 3}$, respectively. We will now describe the third component. The description in terms of the Young symmetrizer $c_{(2,1)}$ is not so illuminating. A more concrete way is to exploit the action of the class sums $1_C = \sum_{g \in C} g$. By Exercise 6.10.6 or Exercise 6.9.4, a class sum acts in an irreducible representation V by

$$\pi_V(1_C) = \frac{|C|\chi_V(C)}{\dim V}\ \mathrm{Id}_V\,.$$

(If you have not done these exercises, note that the left-hand-side is an intertwining operator on V, so by Schur's lemma it is a multiple of the identity. To compute which multiple, take the trace.) In the case $C = \{(123), (132)\}$, we see from (6.11) that $\pi_{(3)}(1_C) = \pi_{(1,1,1)}(1_C) = 2\,\mathrm{Id}$, but $\pi_{(2,1)}(1_C) = -\,\mathrm{Id}$. Thus, in any representation V of S_3, the only possible eigenvalues of $\pi_V(1_C)$ are 2 and -1, and the isotypic component corresponding to $V_{(2,1)}$ is the eigenspace with eigenvalue -1. In the case at hand, this means that the subspace $V_{(2,1)} \otimes V^{(2,1)}$ in (7.28) is the nullspace of $\pi_{V^{\otimes 3}}(\mathrm{id} + (123) + (132))$. If we think of $V^{\otimes 3}$ as the space of trilinear forms $V^* \times V^* \times V^* \times \mathbb{C}$, this means that any trilinear form f can be written uniquely as the sum of a symmetric form, an antisymmetric form and a form g that is cyclically antisymmetric in the sense that

$$g(u, v, w) + g(v, w, u) + g(w, u, v) = 0.$$

The explicit projection from f to g can be obtained from Exercise 6.10.5 as

$$g(u,v,w) = \frac{2f(u,v,w) - f(v,w,u) - f(w,u,v)}{3}.$$

Exercise 7.4.1. Using the character table (6.12), describe any space $V_\lambda \otimes V^\lambda \subseteq V^{\otimes 4}$ as an eigenspace of some class sum 1_C.

7.5 Two determinant evaluations

In order to prove Theorem 7.1.2, we will need two well-known determinant evaluations: the Vandermonde determinant and the Cauchy determinant. To prove them, the following elementary result is useful. Note that it also implies that the Schur polynomials defined in (7.10) are actual polynomials.

Lemma 7.5.1. *If $f = f(x_1, \ldots, x_n)$ is an antisymmetric polynomial (over $\mathbb{C}$, say), then f is divisible by $\prod_{1 \le i < j \le n}(x_i - x_j)$.*

Proof. Considering f as a function of x_1, it vanishes at the points $x_2, \ldots, x_n$. Thus, it is divisible by $g = (x_1 - x_2)\cdots(x_1 - x_n)$. Considering now f/g as a polynomial in x_2, it vanishes at $x_3, \ldots, x_n$. Iterating this argument, we arrive at the desired conclusion. □

For the determinant evaluations, we only need the following special case.

Corollary 7.5.2. *If $f = f(x_1, \ldots, x_n)$ is an antisymmetric polynomial (over $\mathbb{C}$, say) of degree at most $n-1$ in each variable, then f is a constant multiple of $\prod_{1 \le i < j \le n}(x_i - x_j)$.*

Another way to see this is the following. The space of polynomials with the stated properties can be identified with $V^{\wedge n}$, where V is the space of polynomials in one variable of degree at most $n-1$. But since $\dim(V) = n$, we have $\dim(V^{\wedge n}) = 1$, so there is only one such polynomial up to multiplication by constants.

Lemma 7.5.3 (Vandermonde determinant)**.** *We have*

$$\det_{1 \le i,j \le n}(x_i^{n-j}) = \prod_{1 \le i < j \le n}(x_i - x_j). \tag{7.30}$$

Proof. By Corollary 7.5.2, the two sides of (7.30) differ by a multiplicative constant. To compute the constant, we compare the coefficient of $x_1^{n-1}x_2^{n-2}\cdots x_{n-1}$ on both sides. On the left, this monomial appears by multiplying the entries on the main diagonal; that is, in (1.8) it corresponds to the term with $\sigma = \mathrm{id}$. Thus, the coefficient is 1. On the right, it appears when expanding the product and always choosing x_i in $x_i - x_j$. Thus, the coefficient is again 1. □

Lemma 7.5.4 (Cauchy determinant)**.** *We have*

$$\det_{1\le j,k\le n}\left(\frac{1}{1-x_jy_k}\right)=\frac{\prod_{1\le j<k\le n}(x_j-x_k)(y_j-y_k)}{\prod_{j,k=1}^{n}(1-x_jy_k)}. \tag{7.31}$$

Proof. Clearly, $\prod_{j,k}(1-x_jy_k)\det((1-x_jy_k)^{-1})$ is a polynomial in the variables x_j and y_j. Moreover, it has degree at most $n-1$ in each of these variables, and is separately antisymmetric in x_j and y_j. Thus, by Corollary 7.5.2, the two sides of (7.31) agree up to a multiplicative constant. To compute the constant, we multiply (7.31) by $\prod_{j,k}(1-x_jy_k)$ and let $x_n=1/y_n$. On the left, we only get a contribution from terms involving the lower right matrix element, that is, from terms with $\sigma(n)=n$ in (1.8). If we let D_n denote the left-hand side of (7.31), these terms combine to $D_{n-1}/(1-x_ny_n)$. Thus, we have

$$\begin{aligned}\prod_{j,k=1}^{n}(1-x_jy_k)\cdot D_n\Big|_{x_n=1/y_n}&=\frac{\prod_{j,k=1}^{n}(1-x_jy_k)}{1-x_ny_n}\Big|_{x_n=1/y_n}\cdot D_{n-1}\\&=\prod_{j=1}^{n-1}(1-x_jy_n)(1-y_j/y_n)\prod_{j,k=1}^{n-1}(1-x_jy_k)\cdot D_{n-1}.\end{aligned}$$

On the right, we have

$$\prod_{1\le j<k\le n}(x_j-x_k)(y_j-y_k)\Big|_{x_n=1/y_n}=\prod_{j=1}^{n-1}(x_j-1/y_n)(y_j-y_n)\prod_{1\le j<k\le n-1}(x_j-x_k)(y_j-y_k).$$

Noting that $(1-x_jy_n)(1-y_j/y_n)=(x_j-1/y_n)(y_j-y_n)$, the result now follows by induction on n. □

Exercise 7.5.1. Prove that if p_j is a polynomial of degree j with leading coefficient a_j, then

$$\det_{1\le i,j\le n}(p_{n-j}(x_i))=a_0a_1\cdots a_{n-1}\prod_{1\le i<j\le n}(x_i-x_j).$$

Exercise 7.5.2. Find an alternative proof of the Cauchy determinant by expansion along a row and induction on n. You will need the partial fraction identity

$$\frac{\prod_{j=1}^{n-1}(x-a_j)}{\prod_{j=1}^{n}(x-b_j)}=\sum_{k=1}^{n}\frac{1}{x-b_k}\frac{\prod_{j=1}^{n-1}(b_k-b_j)}{\prod_{j=1,\,j\ne k}^{n}(b_k-b_j)},$$

that you may have learned as a tool for integrating rational functions.

7.6 Characters

Throughout this section, V is a complex vector space of dimension m. We will compute the characters $\chi_\lambda(g) = \mathrm{Tr}_{V_\lambda}(g)$ and $\chi^\lambda(A) = \mathrm{Tr}_{V^\lambda}(A)$, where $g \in \mathrm{S}_n$, $A \in \mathrm{GL}(V)$.

By Theorem 7.1.4,

$$\mathrm{Tr}_{V^{\otimes n}}(\rho(A)\pi(\sigma)) = \sum_\lambda \chi_\lambda(\sigma)\chi^\lambda(A), \tag{7.32}$$

where the sum runs over partitions of n. The left-hand side can be computed directly as follows.

Lemma 7.6.1. *Suppose that $A \in \mathrm{GL}(V)$ has eigenvalues $x_1, \ldots, x_m$ and let $\sigma \in \mathrm{S}_n$ have cycle structure $(c_1, \ldots, c_n)$. Then,*

$$\mathrm{Tr}_{V^{\otimes n}}(\rho(A)\pi(\sigma)) = \prod_{k=1}^{n}(x_1^k + \cdots + x_m^k)^{c_k}. \tag{7.33}$$

Proof. Choosing a basis $(e_j)_{j=1}^m$ in V, we write $Ae_j = \sum_k A_j^k e_k$. Then,

$$\rho(A)\pi(\sigma)(e_{j_1} \otimes \cdots \otimes e_{j_n}) = \sum_{k_1,\ldots,k_n=1}^{m} A^{k_1}_{j_{\sigma^{-1}(1)}} \ldots A^{k_n}_{j_{\sigma^{-1}(n)}} e_{k_1} \otimes \cdots \otimes e_{k_n}.$$

It follows that

$$\mathrm{Tr}_{V^{\otimes n}}(\rho(A)\pi(\sigma)) = \sum_{k_1,\ldots,k_n=1}^{m} A^{k_{\sigma(1)}}_{k_1} \cdots A^{k_{\sigma(n)}}_{k_n}. \tag{7.34}$$

To explain why this can be written as indicated, we give an example. If $\sigma = (12)(345) \in \mathrm{S}_5$, then the right-hand side of (7.34) factors as

$$\sum_{k_1,k_2} A^{k_2}_{k_1} A^{k_1}_{k_2} \sum_{k_3,k_4,k_5} A^{k_4}_{k_3} A^{k_5}_{k_4} A^{k_3}_{k_5}.$$

Since

$$\mathrm{Tr}(A^j) = \sum_{k_1,\ldots,k_j} A^{k_2}_{k_1} A^{k_3}_{k_2} \cdots A^{k_j}_{k_{j-1}} A^{k_1}_{k_j},$$

this can be written as $\mathrm{Tr}(A^2)\,\mathrm{Tr}(A^3)$. For general σ, we obtain in the same way $\prod_{k=1}^n \mathrm{Tr}(A^k)^{c_k}$. Since A^k has eigenvalues $x_1^k, \ldots, x_m^k$, the result follows. □

Fixing a basis for V and generic scalars $x_1, \ldots, x_m$, let D be the diagonal matrix with x_j in position (j, j) and 0 else. Then,

$$\rho(D)(e_{k_1} \otimes \cdots \otimes e_{k_n}) = x_{k_1} \cdots x_{k_n}(e_{k_1} \otimes \cdots \otimes e_{k_n}). \tag{7.35}$$

This gives an eigenbasis for $\rho(D)$. The eigenvalues are the numbers $x^\mu = x_1^{\mu_1} \cdots x_m^{\mu_m}$, where $\mu_j \geq 0$ and $\mu_1 + \cdots + \mu_m = n$. (Note that we are not assuming that $\mu_j \geq \mu_{j+1}$, so μ is a *composition* of n rather than a partition.) Let us write W_μ for the corresponding eigenspace, so that

$$V^{\otimes n} = \bigoplus_\mu W_\mu. \tag{7.36}$$

It then follows from (7.35) that the vector

$$e_\mu = \underbrace{e_1 \otimes \cdots \otimes e_1}_{\mu_1} \otimes \cdots \otimes \underbrace{e_n \otimes \cdots \otimes e_n}_{\mu_n}$$

generates W_μ as a representation of S_n, that is, $W_\mu = \pi(\mathbb{C}[\mathrm{S}_n])e_\mu$.

Clearly, the stabilizer P of e_μ in S_n is isomorphic to $S_{\mu_1} \times \cdots \times \mathrm{S}_{\mu_m}$. Let $[\mu]$ denote the partition obtained from μ by ordering the non-zero components of μ in decreasing order. Then, $P \simeq P_{[\mu]}$, so it follows from Proposition 7.3.2 that $W_\mu \simeq U_{[\mu]}$.

Let us now consider the trace of $\rho(D)\pi(\sigma)$, where $\sigma \in \mathrm{S}_n$. On the one hand, it is given explicitly by the right-hand side of (7.33). On the other hand, by (7.36) it can be expressed as $\sum_\mu \psi_{[\mu]}(\sigma)x^\mu$, where ψ_λ is the character of U_λ. This proves the following result.

Lemma 7.6.2. *If C is the conjugacy class with cycle structure $(c_1, \ldots, c_n)$, then*

$$\prod_{k=1}^{n} (x_1^k + \cdots + x_m^k)^{c_k} = \sum_{\substack{\mu_1,\ldots,\mu_m \geq 0,\\ \mu_1+\cdots+\mu_m = n}} \psi_{[\mu]}(C)x^\mu. \tag{7.37}$$

To prove Theorem 7.1.2, we need to multiply (7.37) by $\prod_{1\leq i<j\leq m}(x_i - x_j)$. The expansion of that factor into monomials is the Vandermonde determinant. It will be convenient to introduce the notation

$$\rho = (m-1, m-2, \ldots, 0)$$

and more generally

$$\sigma(\rho) = (m - \sigma(1), \ldots, m - \sigma(m)), \quad \sigma \in \mathrm{S}_m.$$

Then, (7.30) can be written compactly as

$$\prod_{1\leq i<j\leq m} (x_i - x_j) = \sum_{\sigma\in \mathrm{S}_m} \mathrm{sgn}(\sigma) x^{\sigma(\rho)}.$$

Consequently,

$$\prod_{1\le i<j\le m}(x_i-x_j)\prod_{k=1}^{n}(x_1^k+\cdots+x_m^k)^{c_k}=\sum_{\substack{\mu_1,\ldots,\mu_m\ge 0\\ \mu_1+\cdots+\mu_m=n}}\sum_{\sigma\in \mathrm{S}_m}\mathrm{sgn}(\sigma)\psi_{[\mu]}(C)x^{\mu+\sigma(\rho)}. \quad (7.38)$$

When λ is a partition, let $\phi_\lambda=\phi_\lambda(C)$ be the coefficient of $x^{\lambda+\rho}$ in (7.38). That is,

$$\phi_\lambda=\sum_{\sigma\in \mathrm{S}'_m}\mathrm{sgn}(\sigma)\psi_{[\lambda+\rho-\sigma(\rho)]}, \quad (7.39)$$

where the prime means that we exclude terms for which $\lambda+\rho-\sigma(\rho)$ contains negative entries. We will prove Theorem 7.1.2, which says that $\chi_\lambda=\phi_\lambda$.

To give an example, if $m=3$, then $\rho=(2,1,0)$ and $\sigma(\rho)$ assumes the values

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

where we first applied the even permutations and then the odd ones. The corresponding values of $\rho-\sigma(\rho)$ are

$$(0,0,0),\ (1,1,-2),\ (2,-1,-1),\ (0,1,-1),\ (2,0,-2),\ (1,-1,0).$$

If $\lambda=(2,2,1)$, then $\lambda+\rho-\sigma(\rho)$ equals

$$(2,2,1),\ (3,3,-1),\ (4,1,0),\ (2,3,0),\ (4,2,-1),\ (3,1,1).$$

The prime in (7.39) means that the second and fifth value are excluded from the summation. Thus,

$$\phi_{(2,2,1)}=\psi_{(2,2,1)}+\psi_{(4,1)}-\psi_{(3,2)}-\psi_{(3,1,1)}.$$

Note that all the partitions appearing on the right exceed $(2,2,1)$ in the lexicographic order. This is a general phenomenon.

Lemma 7.6.3. *All partitions appearing in* (7.39) *satisfy* $[\lambda+\rho-\sigma(\rho)]\ge_{\mathrm{lex}}\lambda$*, with equality if and only if* $\sigma=\mathrm{id}$.

Proof. Note that ρ is obtained from $\sigma(\rho)$ by arranging the elements in decreasing order. This increases the lexicographic order, so $\lambda+\rho-\sigma(\rho)\ge_{\mathrm{lex}}\lambda$. Rearranging the elements of $\lambda+\rho-\sigma(\rho)$ decreasingly again increases the lexicographic order. Deleting the zeroes at the end may decrease the order, but since λ has non-negative parts it does not change the conclusion that $[\lambda+\rho-\sigma(\rho)]\ge_{\mathrm{lex}}\lambda$. Finally, if $\sigma\ne\mathrm{id}$ then $\rho>_{\mathrm{lex}}\sigma(\rho)$ which gives strict inequality. □

We are now ready to complete our computation of the irreducible characters of S_n.

Proof of Theorem 7.1.2. Recall that we must prove that $\chi_\lambda = \phi_\lambda$. By Lemma 7.6.3,

$$\phi_\lambda = \psi_\lambda + \sum_{\mu >_{\mathrm{lex}} \lambda} L_{\mu\lambda} \psi_\mu,$$

for some integers $L_{\lambda\mu}$. Moreover, by Proposition 7.2.4,

$$\psi_\lambda = \chi_\lambda + \sum_{\mu > \lambda} K_{\mu\lambda} \chi_\mu.$$

Since the lexicographic order is a refinement of the dominance order, it follows that

$$\phi_\lambda = \chi_\lambda + \sum_{\mu > \lambda} M_{\mu\lambda} \chi_\mu$$

for some integers $M_{\lambda\mu}$. By orthonormality of characters,

$$\|\phi_\lambda\|^2 = 1 + \sum_{\mu > \lambda} M_{\mu\lambda}^2.$$

Thus, to prove that $\phi_\lambda = \chi_\lambda$, we only have to prove that $\|\phi_\lambda\| = 1$.

By Lemma 1.6.1,

$$\|\phi_\lambda\|^2 = \frac{1}{n!} \sum_C |C| \, \phi_\lambda(C)^2 = \sum_{c_1 + 2c_2 + \cdots + nc_n = n} \frac{\phi_\lambda(C)^2}{\prod_{j=1}^n j^{c_j} c_j!}.$$

Recall that $\phi_\lambda(C)$ is the coefficient of $x^{\lambda+\rho}$ in (7.38). Thus, $\|\phi_\lambda\|^2$ is the coefficient of $x^{\lambda+\rho} y^{\lambda+\rho}$ in

$$\prod_{1 \le i < j \le m} (x_i - x_j)(y_i - y_j) \times \sum_{c_1 + 2c_2 + \cdots + nc_n = n} \prod_{j=1}^n \frac{(x_1^j + \cdots + x_m^j)^{c_j} (y_1^j + \cdots + y_m^j)^{c_j}}{j^{c_j} c_j!}. \quad (7.40)$$

The right-hand side of (7.40) resembles the Taylor series for the exponential function, but the restriction on the summation variables prevents us from summing it explicitly. However, we can ignore this restriction without changing the coefficient that we are interested in. In fact, we claim that $\|\phi_\lambda\|^2$ is the coefficient of $x^{\lambda+\rho} y^{\lambda+\rho}$ in

$$\prod_{1 \le i < j \le m} (x_i - x_j)(y_i - y_j) \sum_{c_1, c_2, \ldots \ge 0} \prod_{j=1}^\infty \frac{(x_1^j + \cdots + x_m^j)^{c_j} (y_1^j + \cdots + y_m^j)^{c_j}}{j^{c_j} c_j!}. \quad (7.41)$$

The reason is that, when λ is a partition of n, then $x^{\lambda+\rho}$ is homogeneous of degree $m-1+n$. On the other hand, each term in (7.41) is homogeneous in x of degree $m-1+c_1+2c_2+\cdots$. Thus, to obtain terms of the form $x^{\lambda+\rho}$ we must have $c_1+2c_2+\cdots+nc_n=n$ and $c_{n+1}=c_{n+2}=\cdots=0$. In particular, this shows that (7.41) makes sense as a formal power series. It also makes sense as a convergent series in the region $\max_{i,j}|x_iy_j|<1$. Indeed, using the standard Taylor expansions

$$e^x=\sum_{k=0}^{\infty}\frac{x^k}{k},\qquad \log\left(\frac{1}{1-x}\right)=\sum_{k=1}^{\infty}\frac{x^k}{k}$$

we can write (7.41) as

$$\prod_{1\le i<j\le m}(x_i-x_j)(y_i-y_j)\prod_{k=1}^{\infty}\exp\left(\frac{(x_1^k+\cdots+x_m^k)(y_1^k+\cdots+y_m^k)}{k}\right)$$
$$=\prod_{1\le i<j\le m}(x_i-x_j)(y_i-y_j)\prod_{i,j=1}^{m}\exp\left(\sum_{k=1}^{\infty}\frac{x_i^ky_j^k}{k}\right)=\frac{\prod_{1\le i<j\le m}(x_i-x_j)(y_i-y_j)}{\prod_{i,j=1}^{m}1-x_iy_j}.$$

We recognize this as the Cauchy determinant evaluation (7.31). Thus, (7.41) equals

$$\det_{1\le i,j\le m}\left(\frac{1}{1-x_iy_j}\right)=\sum_{\sigma\in S_m}\operatorname{sgn}(\sigma)\prod_{j=1}^{m}\frac{1}{1-x_jy_{\sigma(j)}}$$
$$=\sum_{\sigma\in S_m}\sum_{k_1,\ldots,k_m\ge 0}\operatorname{sgn}(\sigma)x_1^{k_1}\cdots x_m^{k_m}y_{\sigma(1)}^{k_1}\cdots y_{\sigma(m)}^{k_m},$$

where we used the geometric series

$$\frac{1}{1-x}=\sum_{k=0}^{\infty}x^k.$$

Note that the entries in the vector $\lambda+\rho$ are mutually distinct. Thus, only the permutation σ contributes to the coefficient of $x^{\lambda+\rho}y^{\lambda+\rho}$, and we find that this coefficient is 1. □

We now turn to the problem of computing the character χ^λ of V^λ. By (7.32) and (7.33),

$$\prod_{k=1}^{n}(x_1^k+\cdots+x_m^k)^{c_k}=\sum_{\lambda}\chi_\lambda(C)\chi^\lambda(A),$$

where C is the conjugacy class with cycle structure $(c_1,\ldots,c_n)$ and $A\in\mathrm{GL}(V)$ has eigenvalues $x_1,\ldots,x_m$. We also know that $\chi_\lambda(C)$ is the coefficient of $x^{\lambda+\rho}$ in

$$\prod_{1\le i<j\le m}(x_i-x_j)\prod_{k=1}^{n}(x_1^k+\cdots+x_m^k)^{c_k}. \tag{7.42}$$

These two expansions can be related by the following result.

Lemma 7.6.4. *Let P be an antisymmetric polynomial in m variables. Write*

$$P(x_1, \dots, x_m) = \sum_{k_1,\dots,k_m \geq 0} C_{k_1,\dots,k_m} x_1^{k_1} \cdots x_m^{k_m}.$$

Then,

$$\frac{P(x_1, \dots, x_m)}{\prod_{1 \leq i < j \leq m}(x_i - x_j)} = \sum_{\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \geq 0} C_{\lambda+\rho}\, S_\lambda(x_1, \dots, x_m), \tag{7.43}$$

where S_λ is the Schur polynomial defined in (7.10).

Proof. That P is antisymmetric means that the coefficients $C_{k_1,\dots,k_m}$ are antisymmetric in $(k_1, \dots, k_m)$. In particular, $C_{k_1,\dots,k_m} = 0$ when $k_i = k_j$ for some $i \neq j$. If that is not the case, we can write $k = \sigma(\lambda + \rho)$ for a unique sequence $\lambda_1 \geq \cdots \geq \lambda_m \geq 0$ and a unique $\sigma \in S_m$. By antisymmetry, we then have $C_{k_1,\dots,k_m} = \operatorname{sgn}(\sigma) C_{\lambda+\rho}$. It follows that

$$P(x_1, \dots, x_m) = \sum_{\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \geq 0} C_{\lambda+\rho} \sum_{\sigma \in \mathrm{S}_n} \operatorname{sgn}(\sigma) x_{\sigma(1)}^{\lambda_1+\rho_1} \cdots x_{\sigma(m)}^{\lambda_m+\rho_m}.$$

Writing the inner sum as a determinant completes the proof. □

Applying Lemma 7.6.4 to the case when P is given by (7.42) gives

$$\sum_\lambda \chi_\lambda(C)\chi^\lambda(A) = \sum_\lambda \chi_\lambda(C) S_\lambda(x_1, \dots, x_n). \tag{7.44}$$

On the left, the sum runs *a priori* over all partitions of n, whereas the sum on the right is over partitions with at most m parts. This can be viewed as a linear relation between the columns of the character table of S_n. However, as the columns are orthogonal, they are in particular linearly independent, so we must have $\chi^\lambda(A) = 0$ if λ has more than m parts and $\chi^\lambda(A) = S_\lambda(x_1, \dots, x_n)$ else. As $\dim V^\lambda = \chi^\lambda(\mathrm{Id})$, we have in the first case $V^\lambda = \{0\}$. As S_λ is not the zero polynomials (both the rows and columns in (7.10) are visibly linearly independent for generic x_j), we have in the second case $V^\lambda \neq \{0\}$, so by Corollary 7.4.4 V^λ is irreducible. This completes the proof of Theorem 7.1.5 and proves Theorem 7.1.6.

Exercise 7.6.1. Give a more direct proof of Lemma 7.6.2 (without using Schur–Weyl duality) as follows.

(a) Prove that, when H is a subgroup of G, the character of $\mathbb{C}[G/H]$ is given by $\chi(C) = |G| \cdot |C \cap H| / |H| \cdot |C|$.

(b) Let $C \subseteq \mathrm{S}_n$ be the conjugacy class with cycle structure $(c_1, \ldots, c_n)$ and $P = \mathrm{S}_{\lambda_1} \times \cdots \times S_{\lambda_m} \subseteq \mathrm{S}_n$. Prove that

$$|C \cap P| = \sum_{(c)} \prod_{j=1}^{m} \frac{\lambda_j!}{\prod_{k=1}^{n} k^{c_{jk}} c_{jk}!},$$

where the sum is over all non-negative integer $(m \times n)$-matrices (c_{jk}) with row sums $\lambda_1, \ldots, \lambda_m$ and column sums $c_1, \ldots, c_n$.

(c) Prove that

$$\psi_\lambda(C) = \sum_{(c)} \prod_{k=1}^{n} \frac{c_k!}{\prod_{j=1}^{m} c_{jk}!},$$

with the sum as in part (b).

(d) Use the multinomial theorem

$$(x_1 + \cdots + x_m)^k = \sum_{k_1 + \cdots + k_m = k} \frac{k!}{k_1! \cdots k_m!} x_1^{k_1} \cdots x_m^{k_m} \tag{7.45}$$

to complete the proof of Lemma 7.6.2.

7.7 Dimensions

Having obtained expressions for the characters, it is not hard to compute the dimension of V_λ and V^λ.

Proof of Corollary 7.1.3. By Theorem 7.1.2, $\dim(V_\lambda) = \chi_\lambda(1)$ is the coefficient of $x^{\lambda+\rho}$ in

$$(x_1 + \cdots + x_m)^n \prod_{1 \le i < j \le m} (x_i - x_j). \tag{7.46}$$

By the multinomial theorem (7.45) and the Vandermonde determinant evaluation (7.30), (7.46) equals

$$\sum_{\lambda_1 + \cdots + \lambda_m = n} \sum_{\sigma \in \mathrm{S}_n} \frac{n!}{\lambda_1! \cdots \lambda_m!} \operatorname{sgn}(\sigma) x^{\lambda + \sigma(\rho)},$$

so

$$\dim(V_\lambda) = n! \sum_{\sigma \in \mathrm{S}_n} \frac{\operatorname{sgn}(\sigma)}{\prod_{j=1}^{m} (\lambda_j + \sigma(j) - j)!} = n! \det_{1 \le j,k \le m} \left(\frac{1}{(\lambda_j + k - j)!} \right),$$

where matrix entries with $\lambda_j + k - j < 0$ should be interpreted as zero. We can write

$$\frac{1}{(\lambda_j + k - j)!} = \frac{\prod_{l=1}^{m-k}(\lambda_j + k - j + l)}{(\lambda_j + m - j)!}.$$

Note that the numerator on the right has the form $p_{m-k}(\lambda_j - j)$, where p_{m-k} is a monic polynomial of degree $m - k$. Thus, by a variation of the Vandermonde determinant evaluation (see Exercise 7.5.1),

$$\begin{aligned}\dim(V_\lambda) &= \frac{n!}{\prod_{j=1}^{m}(\lambda_j + m - j)!} \det_{1\le j,k\le m}(p_{m-k}(\lambda_j - j)) \\ &= \frac{n!\prod_{1\le i<j\le m}(\lambda_i - \lambda_j + j - i)}{\prod_{j=1}^{m}(\lambda_j + m - j)!}. \end{aligned} \tag{7.47}$$

This gives a perfectly useful formula for $\dim(V_\lambda)$. To see that it is equivalent to Corollary 7.1.3, we need to prove that

$$H(\lambda) = \frac{\prod_{j=1}^{m}(\lambda_j + m - j)!}{\prod_{1\le i<j\le m}(\lambda_i - \lambda_j + j - i)}, \tag{7.48}$$

where $H(\lambda)$ is the product of the hook lengths of all boxes in the Young diagram for λ. We prove this by induction on $n = \sum_j \lambda_j$. Let B be the right-most box in the lowest row in the Young-diagram for λ, and let μ be the partition obtained by deleting B. Most boxes in the Young diagrams for λ and μ have the same hook length. The only exceptions are the boxes immediately above B, which have hook length

$$\lambda_1 - \lambda_m + m,\ \lambda_2 - \lambda_m + m - 1, \ldots, \lambda_{m-1} - \lambda_m + 2$$

in λ and the boxes to the left of B, which have hook lengths

$$\lambda_m,\ \lambda_m - 1, \ldots, 2$$

in λ. All these boxes have their hook lengths decreased by 1 when B is removed. The box B itself has hook length 1, so it does not affect $H(\lambda)$. It follows that

$$\frac{H(\lambda)}{H(\mu)} = \lambda_m \prod_{j=1}^{m-1} \frac{\lambda_j - \lambda_m + m - j + 1}{\lambda_j - \lambda_m + m - j}.$$

Using that $\mu = (\lambda_1, \ldots, \lambda_{m-1}, \lambda_m - 1)$ (or $\mu = (\lambda_1, \ldots, \lambda_{m-1})$ in the case $\lambda_m = 1$) it is easy to check that the same recursion is satisfied by the right-hand side of (7.48). Thus, (7.48) follows by induction. □

We can also use Theorem 7.1.6 to compute the dimension of V^λ. An easy way is to specialize the variables of the Schur polynomial as a geometric progression, $x_j = t^{m-j}$. At this point, the Schur polynomial is evaluated by (7.30) as

$$S_\lambda(t^{m-1}, \dots, t, 1) = \frac{\det_{1\le i,j\le m}(t^{(m-i)(\lambda_j+m-j)})}{\prod_{1\le i<j\le m} t^{m-i} - t^{m-j}} = \prod_{1\le i<j\le m} \frac{t^{\lambda_i+m-i} - t^{\lambda_j+m-j}}{t^{m-i} - t^{m-j}}.$$

For instance by L'Hôpital's rule, we have

$$\lim_{t\to 1} \frac{t^a - t^b}{t^c - t^d} = \frac{a-b}{c-d}.$$

Thus,

$$\dim(V^\lambda) = S_\lambda(1, 1, \dots, 1) = \prod_{1\le i<j\le m} \frac{\lambda_i - \lambda_j + j - i}{j-i}.$$

If λ has less than m parts, it should be extended to $\mathbb{Z}^m_{\ge 0}$ by appending zeroes at the end. We leave to the reader to check that this is equivalent to Corollary 7.1.7.

7.8 Additional exercises

Exercise 7.8.1. Let $\sigma = (1\ 2\ \cdots\ n) \in \mathrm{S}_n$. Compute $\chi_\lambda(\sigma)$ for all λ.

Exercise 7.8.2. If σ is a product of k disjoint cycles, counting also the fix-points, and λ is a partition with $\lambda_{k+1} \ge k+1$, show that $\chi_\lambda(\sigma) = 0$.

Chapter 8

Representations of SU(2)

8.1 Compact groups

The $n \times n$ unitary matrices with determinant 1 form the *special unitary group* $\mathrm{SU}(n)$. In this chapter we take a brief look at the representation theory of SU(2), with some focus on the relation to orthogonal polynomials. The group SU(2) is one of the most important examples for applications in physics, as it describes the notion of quantum mechanical spin. It is also one of the simplest example of a compact group, whose representation theory is in general very similar to that of finite groups.

It is easy to see that $A \in \mathrm{SU}(2)$ if and only if

$$A = \begin{bmatrix} \alpha & \beta \\ -\bar{\beta} & \bar{\alpha} \end{bmatrix}, \tag{8.1}$$

where α and β are complex numbers with

$$|\alpha|^2 + |\beta|^2 = 1.$$

Thus, SU(2) can be identified with a sphere of real dimension 3. As the group operations are smooth functions on the sphere, SU(2) is a compact Lie group.

Most of the general theory for finite groups from §6 extends to compact groups. The reason is that there is an extension of our main technique, namely, averaging over the group. More precisely, we define a *compact group* G to be a compact Hausdorff topological space, equipped with a group structure such that the group operations are continuous[1]. Let $\mathcal{C}(G)$ be the vector space of continuous functions $G \to \mathbb{C}$. There is then a unique linear functional $I : \mathcal{C}(G) \to \mathbb{C}$ satisfying the following properties:

(A) I is left-invariant, that is, $I(L_g f) = I(f)$, where L_g is the shift operator $(L_g f)(h) = f(gh)$, g, $h \in G$.

[1]The reader who is not familiar with topology can think of a compact matrix group such as the orthogonal or unitary matrices.

(B) I is continuous, in the sense that if $f_n \to f$ uniformly then $I(f_n) \to I(f)$.

(C) I is normalized as $I(1) = 1$.

The functional I is called the *Haar measure*.[2] It has the additional properties:

(D) I is right-invariant, that is, $I(R_g f) = I(f)$, where $(R_g f)(h) = f(hg)$.

(E) I is positive, that is, if $f \geq 0$ then $I(f) \geq 0$, with equality only for $f = 0$.

We will write the Haar measure as

$$I(f) = \int_G f(g)\, dg.$$

We will not prove the existence and uniqueness in general, but we will give direct proofs for the cases that we need.

The Haar measure on a finite group (equipped with the discrete topology) is

$$I(f) = \frac{1}{|G|} \sum_{g \in G} f(g).$$

As another example, the Haar measure on $\mathbb{R}/\mathbb{Z}$ is given by

$$I(f) = \int_0^1 f(x)\, dx.$$

It is also easy to describe the Haar measure on $\mathrm{SU}(2)$, using the identification with the unit sphere S^3 in $\mathbb{R}^4$.

Lemma 8.1.1. *The Haar measure on* $\mathrm{SU}(2)$ *is given by integration with respect to the normalized Euclidean volume measure*

$$I(f) = \frac{1}{2\pi^2} \int_{\mathrm{S}^3} f\, dS.$$

Proof. Any left shift $A \mapsto BA$, A, $B \in \mathrm{SU}(2)$, corresponds to a real linear transformation of the parameters (α, β) in (8.1). These transformations must preserve S^3. We assume that it is known that any linear transformation that preserves the sphere is orthogonal and preserves Euclidean measure. This proves property (A). Property (B) is obvious and (C) holds since S^3 has volume $2\pi^2$. Note also that (D) is proved in the same way as (A), and (E) is obvious. □

[2]There is an equivalent definition as a measure on an appropriate class of subsets of G, but we will have no reason to discuss that.

A representation of a compact group G is a continuous homomorphism $\pi : G \to \mathrm{GL}(V)$. We will restrict to the case when V is finite-dimensional and complex. Note that the Haar measure can be extended to vector-valued functions on G. Explicitly, if $f = (f_1, \ldots, f_n) : G \to \mathbb{C}^n$ is continuous then $I(f) = (I(f_1), \ldots, I(f_n))$. If $v \in V$, we can then define

$$\bar{v} = \int_G \pi(g)v \, dg \in V^G. \tag{8.2}$$

Using this averaging operator, the proofs of Maschke's theorem and Schur's lemma extend *verbatim* to the case of compact groups. It is also easy to extend Proposition 6.1.7, which tells us that any representation can be viewed as a continuous homomorphism $G \to \mathrm{SU}(n)$ for some n.

We define an inner product on $\mathcal{C}(G)$ by

$$\langle \phi, \psi \rangle = \int_G \phi(g)\overline{\psi(g)} \, dg.$$

Again with the same proof as for finite groups, Lemma 6.2.4 and Corollary 6.2.5 hold for compact groups. One difference is that the number of inequivalent irreducible representations need not be finite.

Finally, we mention that there are extensions of Theorem 6.7.1 and Theorem 6.8.3. These are most naturally formulated in terms of the Hilbert space completion $\mathrm{L}^2(G)$ of $\mathcal{C}(G)$ with respect to the inner product defined above. We can write

$$\mathrm{L}^2(G) \simeq \bigoplus_{V \in \mathrm{Irr}(G)} \mathrm{End}(V),$$

where the sum on the right is the Hilbert space completion of the algebraic direct sum. Moreover, the matrix elements of all irreducible representations form a complete orthogonal system in $\mathrm{L}^2(G)$. As an example, for $G = \mathbb{R}/\mathbb{Z}$ this is the system $(e^{2\pi i n x})_{n=-\infty}^{\infty}$, which underlies the theory of Fourier series.

Exercise 8.1.1. Show that $\mathrm{SU}(2)$ acts on $\mathbb{C} \cup \{\infty\}$ by

$$\phi\left(\begin{bmatrix} \alpha & \beta \\ -\bar{\beta} & \bar{\alpha} \end{bmatrix}\right)(z) = \frac{\alpha z + \beta}{-\bar{\beta} z + \bar{\alpha}}.$$

Show that $\mathbb{C} \cup \{\infty\}$ can be identified with the sphere $\mathrm{S}^2 \subseteq \mathbb{R}^3$ in such a way that $\phi : \mathrm{SU}(2) \to \mathrm{SO}(3)$. (Here, $\mathrm{SO}(3)$ denotes the three-dimensional orthogonal real matrices with determinant 1, that is, the rotations of $\mathbb{R}^3$ fixing the origin.) Finally, show that ϕ gives an isomorphism $\mathrm{SU}(2)/\{\pm \mathrm{Id}\} \simeq \mathrm{SO}(3)$.

Exercise 8.1.2. Let G be the group of matrices $A_{ab} = \left[\begin{smallmatrix} a & b \\ 0 & 1 \end{smallmatrix}\right]$, where $a > 0$ and $b \in \mathbb{R}$. (It can be viewed as a group of affine transformations $x \mapsto ax + b$ of $\mathbb{R}$.) Define the functionals

$$I_1(f) = \iint f(A_{ab}) \, \frac{da\, db}{a^2}, \qquad I_2(f) = \iint f(A_{ab}) \, \frac{da\, db}{a}$$

on, say, continuous and compactly supported functions on G. Show that I_1 is left invariant and I_2 right invariant. (This shows that, in the more general setting of locally compact groups, conditions (A) and (D) are not equivalent.)

8.2 Representations

We already know some representations of SU(2). If $V = \mathbb{C}^2$ then, by Theorem 7.1.5, $V^{(n)} \simeq V^{\odot n}$ is an irreducible representation of $\mathrm{GL}(2, \mathbb{C})$. We will denote this representation V_n. We claim that it is also irreducible for SU(2). To see this, it suffices to note that the four unitary matrices

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} \mathrm{i} & 0 \\ 0 & -\mathrm{i} \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & \mathrm{i} \\ \mathrm{i} & 0 \end{bmatrix}$$

form a basis for $\mathrm{End}(V)$. Thus, if a subspace is invariant under SU(2) it is trivial.

A useful realization of V_n is as homogeneous polynomials of degree n in two variables x and y. To this end, we identify x and y with a basis for V and $x^k y^{n-k}$ with

$$\underbrace{x \odot \cdots \odot x}_{k} \odot \underbrace{y \odot \cdots \odot y}_{n-k}.$$

A matrix $A = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$ acts on the basis vectors by $Ax = ax + cy$, $Ay = bx + dy$ and more generally on homogeneous polynomials by

$$(\pi(A)p)(x, y) = p(ax + cy, bx + dy).$$

Note that if $b = c = 0$, then $x^k y^{n-k}$ is an eigenvector with eigenvalue $a^k d^{n-k}$. It follows that

$$\mathrm{Tr}_{V^{\odot n}}(\pi(A)) = \sum_{k=0}^{n} a^k d^{n-k} = \frac{a^{n+1} - d^{n+1}}{a - d}.$$

This is equal to the Schur polynomial $S_{(n)}(a, d)$, so we get a direct verification of Theorem 7.1.6 for V_n. If A is unitary, then it is conjugate to a diagonal matrix with eigenvalues $e^{\pm \mathrm{i}\theta}$. We will write the corresponding character as

$$\chi_n(\theta) = \frac{e^{\mathrm{i}(n+1)\theta} - e^{-\mathrm{i}(n+1)\theta}}{e^{\mathrm{i}\theta} - e^{-\mathrm{i}\theta}} = \frac{\sin(n+1)\theta}{\sin\theta}. \tag{8.3}$$

Equivalently,

$$\begin{aligned}\chi_n(\theta) &= \sum_{k=0}^{n} e^{\mathrm{i}(2k-n)\theta} \\ &= \begin{cases} 1 + 2\cos(2\theta) + 2\cos(4\theta) + \cdots + 2\cos(n\theta), & n \text{ even}, \\ 2\cos(\theta) + 2\cos(3\theta) + \cdots + 2\cos(n\theta), & n \text{ odd}. \end{cases}\end{aligned} \tag{8.4}$$

We will prove that any irreducible representation of $\mathrm{SU}(2)$ is equivalent to one of the representations V_n. We first prove a similar result for $\mathrm{U}(1) = \{z \in \mathbb{C};\, |z| = 1\} \simeq \mathbb{R}/\mathbb{Z}$. The Haar measure on $\mathrm{U}(1)$ is

$$I(f) = \frac{1}{2\pi} \int_0^{2\pi} f(e^{\mathrm{i}\theta})\, d\theta.$$

We let W_n be the one-dimensional representation of $\mathrm{U}(1)$ given by $\pi(z) = z^n \,\mathrm{Id}$, where $n \in \mathbb{Z}$. We will write the character of W_n as $\psi_n(\theta) = \mathrm{Tr}(\pi(e^{\mathrm{i}\theta})) = e^{\mathrm{i}n\theta}$.

Lemma 8.2.1. *Any irreducible representation of* $\mathrm{U}(1)$ *is equivalent to* W_n *for some* $n \in \mathbb{Z}$.

Proof. Let V be an irreducible representation, and suppose that V is not equivalent to W_n for any n. Then, the character $f(\theta) = \mathrm{Tr}_V(\pi(e^{\mathrm{i}\theta}))$ is a 2π-periodic continuous function such that

$$\langle f, \psi_n \rangle = \frac{1}{2\pi} \int_0^{2\pi} f(\theta) e^{-\mathrm{i}n\theta}\, d\theta = 0$$

for all $n \in \mathbb{Z}$. It is then a basic fact from Fourier series that $f = 0$, which gives the contradiction $\dim(V) = f(1) = 0$. □

Proposition 8.2.2. *Any irreducible representation of* $\mathrm{SU}(2)$ *is equivalent to* V_n *for some* $n \in \mathbb{Z}_{\geq 0}$.

Proof. Let V be a representation of $\mathrm{SU}(2)$ and let

$$R(z) = \begin{bmatrix} z & 0 \\ 0 & z^{-1} \end{bmatrix}.$$

Then, R is a homomorphism $\mathrm{U}(1) \to \mathrm{SU}(2)$, so $\rho = \pi_V \circ R$ is a representation of $\mathrm{U}(1)$ on V. By Lemma 8.2.1, it splits into a finite sum of the representations W_k, so $f(\theta) = \mathrm{Tr}_V(\rho(e^{\mathrm{i}\theta})) = \sum_k c_k e^{\mathrm{i}k\theta}$, where only finitely many terms are non-zero. Since any matrix in $\mathrm{SU}(2)$ is conjugate to $R(z)$ for some z, we can identify f with

the character of V. Since $f(\theta) = f(-\theta)$, the coefficients satisfy $c_k = c_{-k}$, so we can write

$$f(\theta) = c_0 + 2c_1\cos(\theta) + 2c_2\cos(2\theta) + \cdots + 2c_n\cos(n\theta)$$

for some n. It is then clear from (8.4) that f is a linear combination of $\chi_0, \ldots, \chi_n$. Since V is irreducible, this is only possible if $V \simeq V_j$ for some j. □

As Proposition 6.1.7 holds for compact groups, there is an SU(2)-invariant inner product on V_n. (One can also extend Exercise 6.1.3, so this inner product is unique up to normalization.) On $V_1 = \mathbb{C}^2$ we can take the standard inner product with orthonormal basis $e_1 = (1, 0)$, $e_2 = (0, 1)$. On $V_1^{\otimes n}$ we then have an invariant inner product

$$\langle u_1 \otimes \cdots \otimes u_n, v_1 \otimes \cdots \otimes v_n\rangle = \langle u_1, v_1\rangle \cdots \langle u_n, v_n\rangle.$$

There is an intertwining embedding $I : V^{\odot n} \to V^{\otimes n}$, given by

$$I(u_1 \odot \cdots \odot u_n) = \frac{1}{n!}\sum_{\sigma\in \mathrm{S}_n} u_{\sigma(1)} \otimes \cdots \otimes u_{\sigma(n)}.$$

A basis for $V^{\odot n}$ is given by $(E_k)_{k=0}^n$, where

$$E_k = \underbrace{e_1 \odot \cdots \odot e_1}_{k} \odot \underbrace{e_2 \odot \cdots \odot e_2}_{n-k}.$$

We then obtain an invariant inner product on $V^{\odot n}$ by defining

$$\langle E_k, E_l\rangle_{V^{\odot n}} = \langle I(E_k), I(E_l)\rangle_{V^{\otimes n}}.$$

It is clear that $\langle E_k, E_l\rangle = 0$ if $k \neq l$. If $k = l$, we write $I(E_k)$ and $I(E_l)$ as sums over σ, $\tau \in \mathrm{S}_n$, respectively. For fixed σ, there are then $k!(n-k)!$ permutations τ that contribute 1 to the sum; all other terms contribute 1. We conclude that

$$\langle E_k, E_k\rangle = \frac{1}{(n!)^2}\sum_{\sigma\in \mathrm{S}_n} k!(n-k)! = \frac{1}{\binom{n}{k}}.$$

This gives the following result.

Lemma 8.2.3. *There is an* SU(2)*-invariant inner product on V_n defined by $\langle E_k, E_l\rangle = \delta_{kl}/\binom{n}{k}$.*

We will give another proof of Lemma 8.2.3 in §8.3.

Exercise 8.2.1. Show that V_n can be realized as the space of all polynomials of degree at most n in z, with the action

$$(\pi(A)p)(z) = (bx+dy)^n p\left(\frac{ax+cy}{bx+dy}\right), \qquad A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{GL}(2,\mathbb{C}).$$

Show that, in this realization, the SU(2)-invariant inner product takes the form[3]

$$\langle p, q\rangle = \frac{n+1}{\pi}\int_{\mathbb{C}} \frac{p(z)\overline{q(z)}}{(1+|z|^2)^{n+2}}\,dxdy, \qquad z = x + \mathrm{i}y.$$

Exercise 8.2.2. Prove that any irreducible representation of a compact abelian group is one-dimensional. Then prove directly that any continuous homomorphism from U(1) to itself has the form $f(z) = z^n$. (This gives a proof of Lemma 8.2.1 that does not rely on the theory of Fourier series.)

Exercise 8.2.3. Show that V_n is a representation of $\mathrm{SO}(3) \simeq \mathrm{SU}(2)/\{\pm 1\}$ (see Exercise 8.1.1) if and only if n is even. (In quantum mechanics, $n/2$ is known as the *spin*. If a particle with half-integer spin, such an an electron, rotates 360°, then its wave function ψ transforms into $-\psi$.)

Exercise 8.2.4. Using characters, show the equivalence of representations

$$V_m \otimes V_n \simeq V_{m+n} \oplus V_{m+n-2} \oplus \cdots \oplus V_{|m-n|} = \bigoplus_{j=0}^{\min(m,n)} V_{m+n-2j}.$$

8.3 Matrix elements and Krawtchouk polynomials

We will now compute the matrix elements of $\pi_{V_n}(g)$ in the basis $(E_k)_{k=0}^n$ defined above. In the realization of V_n as a space of homogeneous polynomials, we have $E_k = x^k y^{n-k}$. The matrix elements then appear in the expansion

$$(ax+cy)^k(bx+dy)^{n-k} = \sum_{l=0}^{n} \pi(A)_{lk}\, x^l y^{n-l}, \qquad A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{GL}(2,\mathbb{C}). \quad (8.5)$$

Clearly, we can compute the matrix elements from (8.5) using the binomial theorem. This can be done in several ways. Rather than expanding $(ax+cy)^k$

[3] You will need the integral evaluation

$$\int_0^\infty \frac{t^k}{(1+t)^{n+2}}\,dt = \frac{k!(n-k)!}{(n+1)!}, \qquad k = 0,\ldots,n,$$

which is easy to prove by induction on k and partial integration.

directly into monomials, we will expand it into the elements $x^m(bx+dy)^{k-m}$. For simplicity, we assume that $ad-bc=1$, that is, $A \in \mathrm{SL}(2,\mathbb{C})$. Then,

$$ax+cy = \frac{ad-bc}{d}\,x + \frac{c}{d}\,(bx+dy) = \frac{x+c(bx+dy)}{d},$$

so we get

$$(ax+cy)^k = \frac{1}{d^k}\sum_{m=0}^{k}\binom{k}{m}c^{k-m}x^m(bx+dy)^{k-m}.$$

Multiplying this by $(bx+dy)^{n-k}$ and expanding $(bx+dy)^{n-m}$ using the binomial theorem gives

$$\pi(A)_{lk} = \sum_{m=0}^{\min(k,l)}\binom{k}{m}\binom{n-m}{l-m}b^{l-m}c^{k-m}d^{n-k-l}.$$

This sum is a special case of Gauss' hypergeometric function

$${}_2F_1\left(\begin{matrix}a,b\\ c\end{matrix};z\right) = \sum_{k=0}^{\infty}\frac{(a)_k(b)_k}{k!(c)_k}\,x^k,$$

where the shifted factorial (or Pochhammer symbol) is defined by

$$(a)_m = a(a+1)\cdots(a+m-1).$$

Indeed, using that

$$\binom{n}{m} = (-1)^m\frac{(-n)_m}{m!}$$

we get

$$\begin{aligned}\pi(A)_{lk} &= \binom{n}{l}b^lc^kd^{n-k-l}\sum_{m=0}^{\min(k,l)}\frac{(-k)_m(-l)_m}{m!(-n)_m}\left(-\frac{1}{bc}\right)^m\\ &= \binom{n}{l}b^lc^kd^{n-k-l}\,{}_2F_1\left(\begin{matrix}-k,-l\\ -n\end{matrix};-\frac{1}{bc}\right).\end{aligned}\tag{8.6}$$

(When we change the range of summation to $\sum_{m=0}^{\infty}$ we are formally adding a finite number of zero terms, for $\min(k,l) < m \le n$, and an infinite number of terms of the form $0/0$, for $m > n$. The convention is to treat all such terms as zero.)

Note that $(-k)_m$ is a polynomial in k of degree m so, except for the trivial factor $(c/d)^k$, (8.6) is a polynomial in k of degree l. These polynomials are known as *Krawtchouk polynomials*. The standard notation for these polynomials is

$$K_k(x;p;n) = {}_2F_1\left(\begin{matrix}-k,-x\\ -n\end{matrix};\frac{1}{p}\right).\tag{8.7}$$

We can then formulate (8.6) as follows.

Proposition 8.3.1. *The element* $A = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] \in \mathrm{SL}(2,\mathbb{C})$ *acts on the basis vectors* $E_k = x^k y^{n-k}$ *by* $\pi(A)E_k = \sum_l \pi(A)_{lk} E_l$, *where*

$$\pi(A)_{lk} = \binom{n}{l} b^l c^k d^{n-k-l}\, K_l(k; -bc; n).$$

In particular, $A = \left[\begin{smallmatrix} \alpha & \beta \\ -\bar\beta & \bar\alpha \end{smallmatrix}\right] \in \mathrm{SU}(2)$ *acts by*

$$\pi(A)_{lk} = \binom{n}{l} \beta^l (-\bar\beta)^k \bar\alpha^{n-k-l}\, K_l\left(k; |\beta|^2; n\right). \tag{8.8}$$

The invariance of the inner product is reflected in orthogonality relations for the matrix elements. Indeed, the identity $\pi(A)\pi(A^{-1}) = \mathrm{Id}$ can be written

$$\sum_{l=0}^{n} \pi(A)_{kl}\pi(A^{-1})_{lm} = \delta_{km}. \tag{8.9}$$

Assuming $A \in \mathrm{SU}(2)$, we have $A^{-1} = A^*$. Using (**??**), we can then write (8.9) as

$$\frac{1}{\|E_k\|^2}\,\delta_{km} = \sum_{l=0}^{n} \frac{1}{\|E_l\|^2}\,\pi(A)_{kl}\,\overline{\pi(A)_{ml}}.$$

Expressing the matrix elements as in (8.8) we get after simplification

$$\frac{1}{\binom{n}{k}}\left(\frac{1-|\beta|^2}{|\beta|^2}\right)^k \delta_{km} = \sum_{l=0}^{n}\binom{n}{l}|\beta|^{2l}(1-|\beta|^2)^{n-l} K_k(l;|\beta|^2;n)K_m(l;|\beta|^2;n).$$

Thus, the Krawtchouk polynomials are the orthogonal polynomials with respect to the binomial distribution.

Corollary 8.3.2. *For* n *a non-negative integer and* $0 < p < 1$, *let*

$$\int f(x)\, d\mu(x) = \sum_{l=0}^{n} \binom{n}{l} p^l (1-p)^{n-l} f(l).$$

Then, the Krawtchouk polynomials $K_k(x) = K_k(x; p; n)$ *satisfy the orthogonality relations*

$$\int K_k(x)K_l(x)\, d\mu(x) = \frac{(1-p)^k}{p^k\binom{n}{k}}\,\delta_{kl}, \qquad 0 \le k,\, l \le n.$$

Exercise 8.3.1. Let $S = \sum_{k=0}^{\infty} t_k$ be a series (convergent or divergent) such that t_{k+1}/t_k is a rational function in k. Show that there exist parameters such that

$$S = C \sum_{k=0}^{\infty} \frac{(a_1)_k \cdots (a_r)_k}{(b_1)_k \cdots (b_s)_k k!}\, x^k.$$

This is the hypergeometric series ${}_rF_s$.

Exercise 8.3.2. Replacing m by $k-m$ in the sum

$$ {}_2F_1\left(\begin{matrix}-k, a\\ b\end{matrix}; x\right) = \sum_{m=0}^{k} \frac{(-k)_m (a)_m}{(b)_m m!}\, x^m, $$

show that

$$ {}_2F_1\left(\begin{matrix}-k, a\\ b\end{matrix}; x\right) = (-1)^k \frac{(a)_k}{(b)_k}\, x^k\, {}_2F_1\left(\begin{matrix}-k, 1-b-k\\ 1-a-k\end{matrix}; \frac{1}{x}\right). $$

Exercise 8.3.3. The matrix elements of any group representation satisfy $\pi(AB)_{kl} = \sum_{m=0}^{n} \pi(A)_{km}\pi(B)_{ml}$. Write this explicitly for arbitrary matrices in $\mathrm{SL}(2,\mathbb{C})$. Deduce the addition formula for Krawtchouk polynomials:

$$ \sum_{m=0}^{n} \binom{n}{m} K_k(m;p;n) K_l(m;q;n) z^m \\ = (1+z)^{n-k-l} \left(1+z-\frac{z}{p}\right)^k \left(1+z-\frac{z}{q}\right)^l K_l(k;r;n), $$

where $r = -(p+pz-z)(q+qz-z)/z$.

8.4 Jacobi polynomials

The *Jacobi polynomials* are the orthogonal polynomials for the measure

$$ \int f(x)\, d\mu(x) = \int_{-1}^{1} f(x)(1-x)^a (1+x)^b\, dx, $$

where a, $b > -1$. The Jacobi polynomials together with their degenerate cases are sometimes called "classical orthogonal polynomials". Important special cases include the Gegenbauer or ultraspherical polynomials, $a = b$, the Legendre or spherical polynomials, $a = b = 0$, the Chebyshev polynomials of the first kind, $a = b = -1/2$ and second kind, $a = b = 1/2$, the Laguerre polynomials as a degenerate case when $a \to \infty$ and the Hermite polynomials as a subsequent degeneration $b \to \infty$. We will show that the Schur orthogonality relations (cf. (6.24)) for SU(2) are equivalent to the orthogonality for Jacobi polynomials, where a and b are arbitrary non-negative integers.

For a general compact group, the Schur orthogonality relations are

$$ \int_G \pi_V(g)_{ij}\, \overline{\pi_W(g)_{kl}}\, dg = \frac{\delta_{ik}\delta_{jl}\delta_{VW}}{\dim V}. \tag{8.10} $$

Here, V and W are irreducible representations equipped with invariant inner products, and the matrix elements are taken with respect to orthonormal bases. The proof of (6.24) extends *verbatim* to prove (8.10). A deeper statement, which we will not prove, is that the matrix elements form a complete system in $\mathrm{L}^2(G)$ (for finite groups we saw this by a dimension count).

Note that the basis elements $E_k = x^k y^{n-k}$ for V_n are not normalized. Compensating for this, we get the relation

$$\int_{\mathrm{SU}(2)} \pi(g)^m_{ij} \overline{\pi(g)^n_{kl}}\, dg = \frac{\delta_{ik}\delta_{jl}\delta_{mn}\binom{n}{i}}{(n+1)\binom{n}{j}},$$

where the upper index on the matrix elements indicates the choice of representation. Using (8.8), this takes the form

$$\frac{1}{2\pi^2}\int_{|\alpha|^2+|\beta|^2=1} \alpha^{n-k-l}\bar\alpha^{m-i-j}\beta^{i+l}\bar\beta^{j+k}\, {}_2F_1\left(\begin{matrix}-i,-j\\-m\end{matrix};\frac{1}{|\beta|^2}\right){}_2F_1\left(\begin{matrix}-k,-l\\-n\end{matrix};\frac{1}{|\beta|^2}\right) dS = \frac{\delta_{ik}\delta_{jl}\delta_{mn}}{(n+1)\binom{n}{j}\binom{n}{k}}. \quad (8.11)$$

This gives an orthogonal system on S^3. To understand it more clearly, we introduce polar coordinates $\alpha = re^{\mathrm{i}\theta}$, $\beta = se^{\mathrm{i}\phi}$. The change of variables

$$\big(\mathrm{Re}(\alpha), \mathrm{Im}(\alpha), \mathrm{Re}(\beta), \mathrm{Im}(\beta)\big) \mapsto (r, s, \theta, \phi)$$

has Jacobian rs. Thus, (8.11) takes the form

$$\frac{1}{2\pi^2}\int_Q \int_{\theta=0}^{2\pi}\int_{\phi=0}^{2\pi} r^{n+m-i-j-k-l+1}s^{i+j+k+l+1}e^{\mathrm{i}\theta(n+i+j-m-k-l)}e^{\mathrm{i}\phi(i+l-j-k)} \times {}_2F_1\left(\begin{matrix}-i,-j\\-m\end{matrix};\frac{1}{s^2}\right){}_2F_1\left(\begin{matrix}-k,-l\\-n\end{matrix};\frac{1}{s^2}\right) d\sigma d\theta d\phi = \frac{\delta_{ik}\delta_{jl}\delta_{mn}}{(n+1)\binom{n}{j}\binom{n}{k}}, \quad (8.12)$$

where $d\sigma$ is arc length measure on the quarter circle

$$Q = \{(r,s);\, r,\, s \geq 0,\, r^2+s^2 = 1\}.$$

Note that if $n+i+j \neq m+k+l$ then the integral over θ in (8.12) vanishes. Similarly, the integral over ϕ vanishes if $i+l \neq j+k$. Thus, the only non-trivial case is when $i+l = j+k$ and $n+i+j = m+k+l$. Rewriting this special case in

terms of the parameters i, k, $a = j - i = l - k$ and $b = m - i - j = n - k - l$ gives

$$2\int_Q r^{2b+1}s^{2(a+i+k)+1}\,{}_2F_1\left(\begin{matrix}-i,-a-i\\-a-b-2i\end{matrix};\frac{1}{s^2}\right){}_2F_1\left(\begin{matrix}-k,-a-k\\-a-b-2k\end{matrix};\frac{1}{s^2}\right)d\sigma \\ = \frac{\delta_{ik}}{(a+b+2k+1)\binom{a+b+2k}{a+k}\binom{a+b+2k}{k}}. \quad (8.13)$$

Let us now assume that $a \geq 0$ and $b \geq 0$ (using symmetries of the matrix elements, one can show that this is no essential restriction). By Exercise 8.3.2,

$$\begin{aligned}{}_2F_1\left(\begin{matrix}-i,-a-i\\-a-b-2i\end{matrix};\frac{1}{s^2}\right) &= (-1)^i\frac{(-a-i)_i}{(-a-b-2i)_i}\frac{1}{s^{2i}}\,{}_2F_1\left(\begin{matrix}-i,a+b+i+1\\a+1\end{matrix};s^2\right)\\ &= (-1)^i\frac{(a+i)!(a+b+i)!}{a!(a+b+2i)!}\frac{1}{s^{2i}}\,{}_2F_1\left(\begin{matrix}-i,a+b+i+1\\a+1\end{matrix};s^2\right).\end{aligned}$$

Using this and the same identity with i replaced by k, (8.13) takes the form

$$2\int_Q r^{2b+1}s^{2a+1}\,{}_2F_1\left(\begin{matrix}-i,a+b+i+1\\a+1\end{matrix};s^2\right){}_2F_1\left(\begin{matrix}-k,a+b+k+1\\a+1\end{matrix};s^2\right)d\sigma \\ = \frac{a!^2(b+k)!k!}{(a+k)!(a+b+k)!(a+b+2k+1)}\,\delta_{ik}.$$

To relate this to Jacobi polynomials, we introduce the parametrization $(r(x), s(x)) = (\sqrt{(1+x)/2}, \sqrt{(1-x)/2})$. Then, $(r(x), s(x))$ runs through Q as $-1 \leq x \leq 1$. Moreover,

$$d\sigma = \sqrt{r'(x)^2 + s'(x)^2}\,dx = \frac{1}{4rs}\,dx.$$

This leads to

$$\int_{-1}^{1}(1-x)^a(1+x)^b\,{}_2F_1\left(\begin{matrix}-i,a+b+i+1\\a+1\end{matrix};\frac{1-x}{2}\right){}_2F_1\left(\begin{matrix}-k,a+b+k+1\\a+1\end{matrix};\frac{1-x}{2}\right)dx \\ = \frac{2^{a+b+1}a!^2(b+k)!k!}{(a+k)!(a+b+k)!(a+b+2k+1)}\,\delta_{ik}.$$

The standard notation for Jacobi polynomials is

$$P_n^{(a,b)}(x) = \frac{(a+1)_n}{n!}\,{}_2F_1\left(\begin{matrix}-n,a+b+n+1\\a+1\end{matrix};\frac{1-x}{2}\right).$$

If a is an integer, the prefactor is $(a+n)!/a!n!$. Replacing (i,k) by (m,n), we have obtained a representation-theoretic proof of the following fact.

Proposition 8.4.1. *For a and b non-negative integers, the Jacobi polynomials satisfy the orthogonality relations*

$$\int_{-1}^{1} P_m^{(a,b)}(x) P_n^{(a,b)}(x)(1-x)^a(1+x)^b\,dx = \frac{2^{a+b-1}(a+n)!(b+n)!}{n!(a+b+n)!(a+b+2n+1)}\,\delta_{mn}.$$

This holds for arbitrary a, $b > -1$, with the factorials replaced by the gamma function. The general case can also be obtained from representation theory; for instance, using the non-compact group $\mathrm{SU}(1,1)$.

Index

action, 11
adjoint, 73
algebra
    associative, 38
    group, 90
    Hopf, 91
algebraic integer, 94
annihilator, 32, 57
antisymmetric map, 46
automorphism, 8
    inner, 17

basis, 34

canonical form
    Jordan, 64
    primary, 63
    rational, 63
Cauchy determinant, 117
Cayley's theorem, 11
Cayley–Hamilton theorem, 66
center
    of character, 87
    of group, 17
character, 75
character table, 81
characteristic, 27
class function, 80
class sum, 82
commutator subgroup, 86
composition (of a number), 119
conjugacy class, 14
conjugation, 14
coset, 9
cycle notation, 13
cycle structure, 13

degree equation, 80
degree of representation, 80
determinant, 15
dimension theorem, 31
direct sum
    of modules, 33
    of representations, 73
direct summand, 33
domain, 20
    integral, 20
    principal ideal, 22
    unique factorization, 25
dominance order, 105

End(G), 20
endomorphism, 8
equivalence
    of group actions, 12
equivalent representations, 70
exact sequence, 33
    short, 33
    split, 34
exterior power, 47

field, 20
    skew, 20
Fourier inversion formula, 90
Fourier transform
    on $\mathbb{Z}_n$, 85
    on finite group, 90

generated
    free module generated by, 35
    group generated by, 9
    ideal generated by, 22
    module generated by, 32
GL(V), 6
greatest common divisor, 26
group, 5

abelian, 6
alternating, 86
compact, 127
cyclic, 9
derived, 16, 86
general linear, 6
isotropy, 17
Lie, 102
quaternion, 16
quotient, 10
special unitary, 127
symmetric, 6
group algebra, 90
G-set, 11

Haar measure, 128
Hom_G, 70
homomorphism
of groups, 8
of modules, 30
of rings, 21
Hom_R, 31
hook length, 101
hypergeometric function, 134

ideal, 21
left, 22
principal, 22
idempotent, 98
inner product, 73
intertwining map, 70
invariant factor decomposition, 60
irreducible element, 24
$\mathrm{Irr}(G)$, 72, 88
isomorphism
of groups, 8
of modules, 30
of rings, 21
isotypic component, 97

kernel
of character, 85
of group homomorphism, 11
of module homomorphism, 31
of ring homomorphism, 23
Kostka number, 107

Lagrange's theorem, 10
least common multiple, 27
lexicographic order, 106
$\mathrm{L}^2(G)$, 76, 129
linear map, 31
linear subspace, 31

Maschke's theorem, 72
matrix
permutation, 12
module, 29
cyclic, 32
dual, 32
finitely generated, 32
free, 34
projective, 40
quotient, 31
right, 32
Specht, 111
torsion, 54
torsion free, 54
multiplicity, 76

orbit, 12, 17
free, 104
order
of group, 9
of group element, 9

partition, 14
permutation, 6
Peter–Weyl theorem, 94
PID, 22
Pochhammer symbol, 134
polynomial
Jacobi, 138
Krawtchouk, 134

minimal, 66
Schur, 102
Specht, 112
primary decomposition, 53
prime element, 24
pure tensor, 43

quotient
group, 10
module, 31
ring, 22

rank, 36
representation, 12, 70
alternating, 80, 83
defining, 77, 79, 83
dual, 74
of associative algebra, 91
permutation, 13
regular, 12, 79
standard, 77
trivial, 74
ring, 19
Boolean, 27
commutative, 20
division, 20
Noetherian, 25
quotient, 22
zero, 20

scalar, 29
change of, 45
Schur orthogonality relations, 94, 136
Schur's lemma, 72
Schur–Weyl duality, 102
shifted factorial, 134
sign, 15
similar matrices, 62
S_n, 6
spin, 133
stabilizer, 12, 17
subgroup, 7
commutator, 16
normal, 10
submodule, 31
subrepresentation, 71
$\mathrm{SU}(n)$, 127
symmetric map, 46
symmetric power, 47

tableau, 105
canonical, 100
row-strict, 105
semi-standard, 107
shape of, 105
standard, 112
weight of, 105
tensor product, 42
of representations, 74
torsion element, 54
transpose of partition, 100

UFD, 25
unit, 24
unitary, 73
universal property
of tensor products, 43

Vandermonde determinant, 116
vector space, 29

Young diagram, 99
Young symmetrizer, 100

$\mathrm{Z}(G)$, 17
$\mathbb{Z}_n$, 7, 20