Lecture Notes on Group Theory

Ronnie Rodgers

April 15, 2025

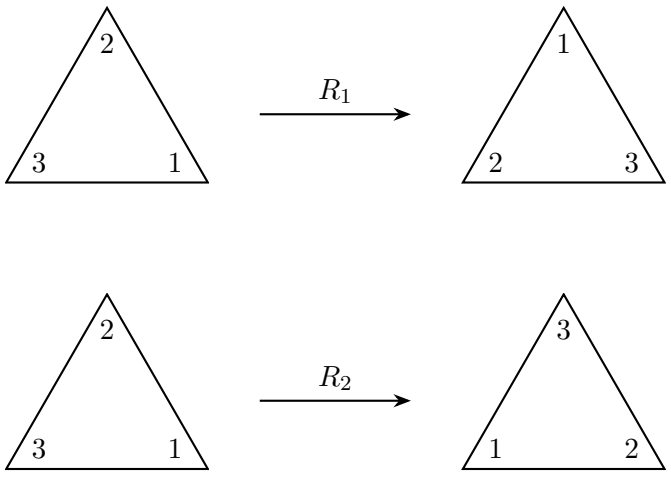
Contents

Preface

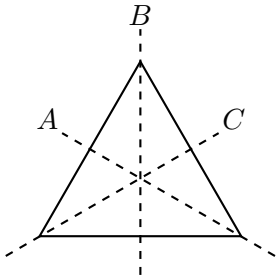
Many physical systems possess some degree of symmetry. A lone planet orbiting a star and the electron in a hydrogen atom both experience spherically-symmetric potentials. A crystal lattice exhibits a discrete set of rotational, reflection, and translational symmetries that together govern its electronic band structure. The Standard Model of particle physics is based upon a set of symmetries which constrain the how its particles can interact with one another. In all of these examples and more, understanding the symmetries is key to solving the equations which the physical system.

Group theory is the mathematical language of symmetry. A group consists of a set of elements with a certain structure. In physics, these elements are symmetry transformations. To motivate the structure of a group, let's consider the symmetries of an equilateral triangle. There are six possible transformations that we can perform on on an equilateral triangle that leave it looking the same:

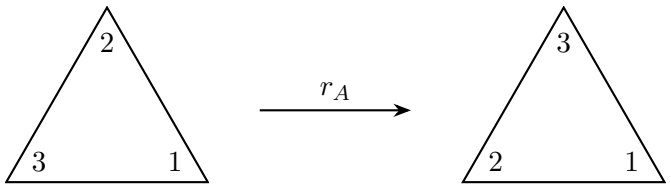
- We can do nothing to it, the "identity transformation". It may seem a little odd to count this as a transformation, but it certainly leaves the triangle invariant and it will be important in what follows. We will denote the identity transformation as e.
- We can rotate the triangle anticlockwise by either an angle of $2\pi/3$ or by an angle of $4\pi/3$. Let's denote these two rotations by R_1 and R_2, respectively. We can draw them by labelling the vertices of the triangle with numbers and showing where the numbers end up after the rotation:

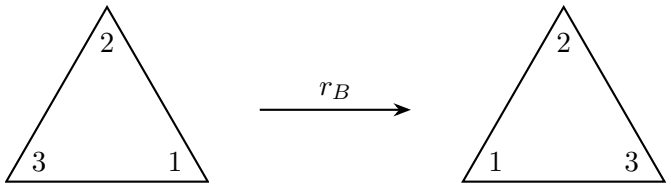


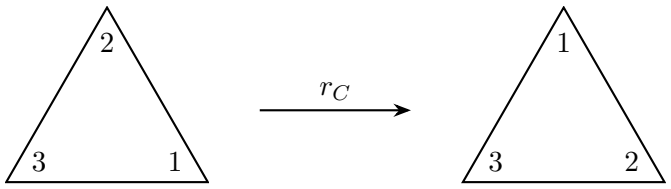
  Any other rotation is equivalent to one of these two rotations, or to leaving the triangle invariant. For instance, a clockwise rotation by $2\pi/3$ is equivalent to R_2.
- We can reflect the triangle through any of the three axes A, B, or C depicted below:



Let's denote these reflections by r_A, r_B, and r_C, respectively. Again we can draw them by labelling the vertices:

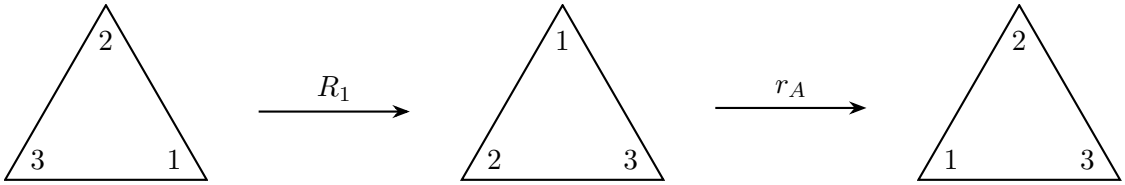




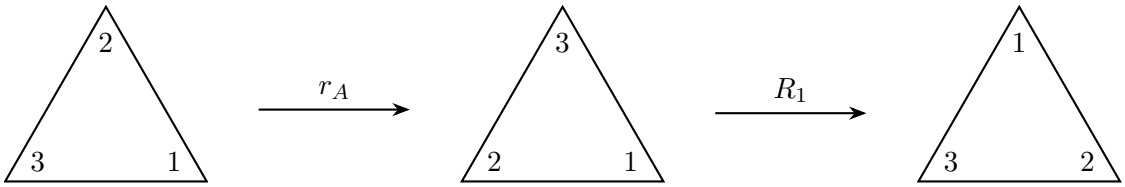


The symmetries described above have some structure. Notice that we can compose any two of the symmetry transformations to get another one. For instance, performing R_1 twice is the same as performing R_2 once. Performing R_1 followed by R_2 gives a total rotation of 2π, equivalent to the identity transformation e.

As another example, performing R_1 followed by r_A is equivalent to r_B, which we can see if we draw these transformations out:



Here the order of transformations matters; if we first perform r_A and then perform R_1, we instead get the reflection r_C:



We can introduce some notation to express this. We will denote the act of performing R_1 followed by r_A as $r_A \circ R_1$. Notice that we read from right-to-left.[1] The statements above can then be expressed as

$$r_A \circ R_1 = r_B, \qquad \text{and} \qquad R_1 \circ r_A = r_C. \tag{1}$$

In this notation, the composition of rotations discussed above is

$$R_1 \circ R_1 = R_2, \qquad \text{and} \qquad R_1 \circ R_2 = R_2 \circ R_1 = e. \tag{2}$$

A more subtle point about composing symmetry transformations arises when we consider performing three of them. For instance, consider the combination $R_1 \circ (r_A \circ R_1)$. This should be interpreted as (i) first perform the symmetry transformation obtained by performing R_1 followed by r_A, (ii) then perform R_1. From the examples in equation (1) we know that the symmetry transformation in point (i) is r_B. We then have that

$$R_1 \circ (r_A \circ R_1) = R_1 \circ r_B = r_A, \tag{3}$$

where the second equality may be checked by drawing out more triangles. Now let's consider the combination $(R_1 \circ r_A) \circ R_1$, meaning (i') first perform R_1, (ii') then perform the symmetry transformation corresponding to r_A followed by R_1. We've already seen that the transformation in (ii') is r_C. We find

$$(R_1 \circ r_A) \circ R_1 = r_C \circ R_1 = r_A, \tag{4}$$

where again we leave the second equality to be checked. We've found that

$$R_1 \circ (r_A \circ R_1) = r_A = (R_1 \circ r_A) \circ R_1. \tag{5}$$

In other words, it doesn't matter where we place the brackets and we may as well write both of the above products as $R_1 \circ r_A \circ R_1$. It turns out that this is a general feature of symmetry transformations — it doesn't matter where we place the brackets when composing three transformations. In other words, composition of transformations is *associative*.

The final piece of structure is that every symmetry transformation can be inverted. By this, we mean that after performing any symmetry transformation there is another transformation that may be performed to bring the triangle back to where it started. In other words, given any symmetry transformation, there is another that can be composed with it such that the final result equals the identity transformation. The identity and each of the reflections are their own inverse, while R_1 and R_2 are the inverses of each other.

In chapter 1 we will formalise the above structure into the mathematical definition of a group and begin exploring the general properties of groups. A group is a collection of abstract objects. In subsequent chapters we will develop the tools needed to apply the theory of groups to physical systems.

[1]The right-to-left convention is inspired by how matrices act on vectors. For example if A and B are matrices and $\mathbf{v}$ is a vector, then $AB\mathbf{v} = A(B\mathbf{v})$ can be thought of as the result of the matrix A acting on the vector $B\mathbf{v}$, in other words we act on the vector with B first.

Part I

Finite groups

Chapter 1

Introduction to groups

1.1 Definition and examples of groups

The properties of the symmetries of an equilateral triangle discussed in the preface exemplify the mathematical structure of a group. A group is a collection of abstract objects along with a way of composing two of these objects together to get a new one, with some structure imposed on this composition. In physics we may think of the objects forming the group as symmetry transformations and the composition as the process of performing two transformations in a row. We now give the mathematical definition of a group.

Definition 1.1.1. A **group** G is a set of elements equipped with an operation $\circ$ that acts on pairs of these elements, called **group multiplication** or **group composition**, satisfying the following properties (group axioms):

(i) **Closure:** For all $g, h \in G$ the product $g \circ h$ is an element of G.

(ii) **Associativity:** For all $f, g, h \in G$ we have $(f \circ g) \circ h = f \circ (g \circ h)$.

(iii) **Identity:** There is an element $e \in G$, such that for all $g \in G$ we have $e \circ g = g \circ e = g$. We call e the identity element.

(iv) **Inverse:** For every element $g \in G$, there is another element $g^{-1} \in G$, such that $g^{-1} \circ g = g \circ g^{-1} = e$. We call g^{-1} the inverse of g.

The number of elements of G is called the **order** of the group. This is sometimes denoted $|G|$. If the order is finite, then the group is a **finite group**.

Depending on context, the group multiplication may also be denoted by another symbol, such as $\cdot$, $\times$, or $+$. Frequently no symbol is used at all, for example writing $g \circ h = gh$. Associativity means that it doesn't matter where we put the brackets when multiplying more than two elements. As a consequence, we can simply drop the brackets, writing both the products in point (ii) above simply as $f \circ g \circ h$.

The group multiplication doesn't have to be commutative; it can be that for some group elements g and h we have $g \circ h \neq h \circ g$. We saw an example of this with the equilateral triangle, where we have $R_1 \circ r_A \neq r_A \circ R_1$. When the group multiplication *is* commutative this gets a special name:

Definition 1.1.2. When the group multiplication is commutative, i.e. when $g \circ h = h \circ g$ for all $g, h \in G$, the group is said to be **abelian**. Otherwise the group is **non-abelian**.[1]

We will now give some simple examples of groups. We can specify a finite group by listing its elements and specifying their multiplication laws in the form of a multiplication table. For instance, if the group consists of elements e and $g_{1,2,\ldots}$, the multiplication table is a table of the form

	e	g_1	g_2	$\cdots$
e	$e \circ e$	$e \circ g_1$	$e \circ g_2$	$\cdots$
g_1	$g_1 \circ e$	$g_1 \circ g_1$	$g_1 \circ g_2$	$\cdots$
g_2	$g_2 \circ e$	$g_2 \circ g_1$	$g_2 \circ g_2$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

where we replace each element $g_i \circ g_j$ in the table with the group element it produces.

Example 1.1.1. Consider the three elements $G = \{e, a, b\}$. There is only one possible multiplication table consistent with the group axioms,

	e	a	b
e	e	a	b
a	a	b	e
b	b	e	a

This satisfies the group axioms because the multiplication is closed and associative,[2] there is an identity element, e, and each element has an inverse, $a^{-1} = b$ and $b^{-1} = a$. The group is abelian, which you can read off from the fact that the multiplication table is symmetric about its leading diagonal (i.e. it is symmetric under reflections through its top-left to bottom-right axis).

Example 1.1.2. It turns out that any multiplication table for fewer than six elements that is consistent with the group axioms is abelian. On the other hand, with six elements we can write out a non-abelian multiplication table. Denoting $G = \{e, a, b, c, d, f\}$, we take the multiplication table to be

	e	a	b	c	d	f
e	e	a	b	c	d	f
a	a	b	e	f	c	d
b	b	e	a	d	f	c
c	c	d	f	e	a	b
d	d	f	c	b	e	a
f	f	c	d	a	b	e

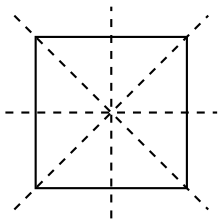
The multiplication table is not symmetric about its leading diagonal, indicating that the group is non-abelian. For example, we have $a \circ c = f$ but $c \circ a = d$. Notice that it

[1] Abelian groups take their name from Norwegian mathematician Niels Henry Abel. Somewhat unusually for a term originating from a person's name, "abelian" is commonly spelt with a lowercase "a".

[2] Associativity is a pain to check — one has to check that $f \circ (g \circ h) = (f \circ g) \circ h$ for each triple of elements of G — but indeed it holds for the multiplication table in example 1.1.1.

is not the case for a non-abelian group that $g \circ h \neq h \circ g$ for *all* non-identical elements g and h. For instance this group has $a \circ b = e = b \circ a$.

Example 1.1.3. The group formed by the symmetries of a regular n-sided polygon is called the **dihedral group** of order $2n$, and denoted D_n (sometimes this group is instead denoted Dih_n or D_{2n}). We discussed D_3, the symmetries of the equilateral triangle, in the preface. The group D_4 is the group of symmetries of the square. It consisists of the identity, three anticlockwise rotations through angles $\pi/2$, π, and $3\pi/2$, and four reflections, through these axes:



The multiplication table can be worked out in a similar manner to what we did for the equilateral triangle, by labelling vertices and seeing where they end up after composing two transformations. We won't write out the multiplication table here, since it's rather large and not so enlightening to look at, but if you'd like to see it you can find it on Wikipedia [1].

We can also sometimes define groups without having to write out a specific multiplication table, for example if we can define them in terms of elementary arithmetic operations. Here are some examples.

Example 1.1.4. Let N be a positive integer. The **cyclic group** of order N consists of the set of N-th roots of unity, with multiplication given by ordinary multiplication of complex numbers. In other words, it is the set of elements

$$e^{2\pi i n/N}, \qquad n = 0, 1, \ldots, N-1. \tag{1.1}$$

This forms a group since multiplication of complex numbers is associative, the multiplication is closed, there is an identity element (corresponding to $n = 0$), and every element has an inverse: the identity is its own inverse, and for $n > 0$ we have $(e^{2\pi i n/N})^{-1} = e^{2\pi i (N-n)/N}$. The multiplication is also abelian. The cyclic group of order N is conventionally denoted $\mathbb{Z}_N$.

Example 1.1.5. Let G be the group of all two-dimensional rotation matrices, with group multiplication given by ordinary matrix multiplication. These may be parameterised as

$$R(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \tag{1.2}$$

where θ is a real number. This is our first example of a group with an infinite number of elements. Each element $R(\theta) \in G$ corresponds to an anticlockwise rotation through angle θ. Let's check that G indeed satisfies the axioms necessary to form a group. To

check closure we note that

$$R(\theta)R(\phi) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} = \begin{pmatrix} \cos(\theta+\phi) & -\sin(\theta+\phi) \\ \sin(\theta+\phi) & \cos(\theta+\phi) \end{pmatrix}$$
$$= R(\theta+\phi), \tag{1.3}$$

thus multiplying two rotation matrices always produces another rotation matrix. The group multiplication is associative since matrix multiplication is always associative. We can also check associativity explicitly using equation (1.3), which implies that

$$R(\theta)\left(R(\phi)R(\psi)\right) = R(\theta)R(\phi+\psi) = R(\theta+\phi+\psi), \tag{1.4}$$

and

$$\left(R(\theta)R(\phi)\right)R(\psi) = R(\theta+\phi)R(\psi) = R(\theta+\phi+\psi). \tag{1.5}$$

The identity element is $e = R(0)$ and each element has an inverse $R(\theta)^{-1} = R(2\pi-\theta)$. This group is abelian, since $R(\theta)R(\phi) = R(\theta+\phi) = R(\phi+\theta) = R(\phi+\theta)$.

1.2 Elementary properties of groups

It turns out that definition 1.1.1 is slightly overprescriptive. We can actually define the identity element and the inverse using a more restrictive set of group axioms that only specify their action when multiplying from the left. The other group axioms then imply how they act when multiplying from the right:

Theorem 1.2.1. *We may replace points (iii) and (iv) in definition 1.1.1 by*

(iii′) There is an element $e \in G$, such that for all $g \in G$ we have $e \circ g = g$.

(iv′) For every element $g \in G$, there is another element $g^{-1} \in G$, such that $g^{-1} \circ g = e$.

This yields a definition of a group equivalent to definition 1.1.1.

Proof. We need to show that axioms (i), (ii), (iii′), and (iv′) imply (iii) and (iv), meaning that we need to show that they imply $g \circ e = g$ and $g \circ g^{-1} = e$ for all $g \in G$.

The proof that $g \circ g^{-1} = e$ works as follows. Using associativity we can manipulate the combination $g^{-1} \circ (g \circ g^{-1})$ to

$$g^{-1} \circ (g \circ g^{-1}) = (g^{-1} \circ g) \circ g^{-1} = e \circ g^{-1} = g^{-1}, \tag{1.6}$$

where the second equality follows from property (iv′) and the third equality follows from (iii′). Since $g^{-1} \in G$, property (iv′) tells us that it must have an inverse element $h = (g^{-1})^{-1} \in G$ such that $h \circ g^{-1} = e$. Using equation (1.6) we can then write

$$\begin{aligned} e &= h \circ g^{-1} \\ &= h \circ (g^{-1} \circ (g \circ g^{-1})) \\ &= (h \circ g^{-1}) \circ (g \circ g^{-1}) \\ &= e \circ (g \circ g^{-1}) \\ &= g \circ g^{-1}, \end{aligned} \tag{1.7}$$

where the final equality follows from (iii′). Using this result, we then have that for all $g \in G$ that

$$g \circ e = g \circ (g^{-1} \circ g) = (g \circ g^{-1}) \circ g = e \circ g = g. \tag{1.8}$$

□

We will now prove some basic properties about the general structure of groups that follow from the group axioms. From now on we will not use any symbol for group multiplication, to avoid clutter.

Theorem 1.2.2. *The inverse of any element of a group is unique.*

Proof. Given a group G with identity element e, consider an element $g \in G$. Suppose that there are $h, h' \in G$ such that $hg = gh = e$ and $h'g = gh' = e$. Then we have

$$h = he = h(gh') = (hg)h' = eh' = h'. \tag{1.9}$$

So we have $h = h'$, meaning in fact h and h' are the same group element. Thus g has only one inverse. □

Theorem 1.2.3. *The inverse of a product of group elements $g_i \in G$ is the product of the inverses in reverse order,*

$$(g_1 g_2 \cdots g_n)^{-1} = g_n^{-1} \cdots g_2^{-1} g_1^{-1} \tag{1.10}$$

Proof. Multiply out using associativity,

$$\begin{aligned} (g_n^{-1} \cdots g_2^{-1} g_1^{-1})(g_1 g_2 \cdots g_n) &= (g_n^{-1} \cdots g_2^{-1})(g_1^{-1} g_1)(g_2 \cdots g_n) \\ &= (g_n^{-1} \cdots g_2^{-1})(g_2 \cdots g_n) \\ &\vdots \\ &= e. \end{aligned} \tag{1.11}$$

□

Theorem 1.2.4. *Given any pair of elements $g, h \in G$, there are unique elements $a, b \in G$ such that*

$$ga = h \qquad \text{and} \qquad bg = h. \tag{1.12}$$

Proof. Applying the inverse of g from the left or right on the above equations, we find

$$a = g^{-1}h \qquad \text{and} \qquad b = hg^{-1}. \tag{1.13}$$

Since the inverse of g is unique, this uniquely specifies a and b. □

Theorem 1.2.5 (Cancellation law)**.** *Given $f, g, h \in G$, if $gf = gh$, then $f = h$. Similarly if $fg = hg$ then $f = h$.*

Proof. Apply g^{-1} from the left to $gf = gh$ and from the right to $fg = hg$. □

Theorem 1.2.6 (Permutation law)**.** *Multiplying all of the elements of a group from the left G by a fixed element $g \in G$ permutes the elements of G. Since the order of the elements in G is immaterial, we denote this by writing $gG = G$. Similarly, multiplying all of the elements of G from the right by a fixed element also permutes the elements of G, which we denote $Gg = G$.*

Proof. By closure, every element of gG must be an element of G. The only way $gG = G$ could fail to hold would be if there is an element of G not present in gG. Since G and gG contain the same number of elements, this would imply that there is at least one repeated element in gG, meaning that there are distinct elements $f, h \in G$ such that $gf = fh$. But by the cancellation law this can only happen if $f = h$, contradicting the assumption that f and h are distinct elements of G. Thus we conclude that $gG = G$. The proof that $Gg = G$ proceeds identically. □

The permutation law implies that each row and each column of a group multiplication table consists of a permutation of the group elements (in particular, there are no repeated elements in any one row or column), as is the case for the tables in examples 1.1.1 and 1.1.2. Let's demonstrate with the former:

Example 1.2.1. Suppose $G = \{e, a, b\}$ with the multiplication table given in example 1.1.1. Then we have three possible sets of the form gG with $g \in G$:

$$\begin{aligned} eG &= \{ee, ea, eb\} = \{e, a, b\}, \\ aG &= \{ae, aa, ab\} = \{a, b, e\}, \\ bG &= \{be, ba, bb\} = \{b, e, a\}. \end{aligned} \tag{1.14}$$

Each set on the right-hand side is simply a permutation of the elements $\{e, a, b\}$, as expected. Notice that by construction, each of the sets on the right-hand side corresponds to one of the rows of the multiplication table. Similarly, each of the possible Gg corresponds to one of the columns of the multiplication table.

Theorem 1.2.7. *In any finite group G of order $|G|$, given an element $g \in G$ we can consider forming the set of elements $g^2 = gg$, $g^3 = gg^2$, and so on. For a given g there exists a positive integer $n \leq |G|$ such that $g^n = e$.*

Proof. To see this, consider forming the set of the first $|G|$ powers of g

$$X_g = \{g, g^2, g^3, \cdots, g^{|G|}\}. \tag{1.15}$$

Since the group multiplication is closed, every element of X_g is an element of G. We want to prove that $e \in X_g$. Since both X_g and G have $|G|$ elements, the only way that e could fail to appear in X_g is if there is at least one element of G that appears multiple times in X_g. For this to happen, there must be two positive integers $a, b \leq |G|$ such that $g^a = g^b$. Let's take a to be the larger of the two integers. Then using associativity we have

$$g^{a-b} g^b = g^b. \tag{1.16}$$

Acting on both sides from the right with $(g^b)^{-1}$ we find that $g^{a-b} = e$. Since $a - b$ is a positive integer less than $|G|$, by construction $g^{a-b} \in X_g$. Thus we always have that $e \in X_g$. □

Definition 1.2.1. Given a finite group G of order $|G|$, the **order** of an element $g \in G$ is the smallest integer n such that $g^n = e$. In theorem 1.2.7 we have shown that $n \leq |G|$. The order of the identity element is always one, and the order of any group element that is its own inverse is two.

1.3 Group maps

Definition 1.3.1. A **group homomorphism** is a map from one group to another that preserves the group multiplication structure. Concretely, suppose we have two groups G and H, and a map $h : G \to H$. For each $g \in G$, let $h(g)$ denote the image of g under this map. The map is a homomorphism if

$$h(g_1)h(g_2) = h(g_1 g_2), \tag{1.17}$$

for all $g_{1,2} \in G$. This equation is what we mean by that map preserving the multiplication structure.

If $h : G \to H$ is a homomorphism, some important properties that follow from the requirement that h preserves the multiplication strucure are:

(i) If e is the identity element in G, then $h(e)$ is the identity element in H. This follows since we must have

$$h(e)h(g) = h(eg) = h(g)$$

for all $g \in G$. Since H is a group, if $h(g) \in H$ then there must be an inverse element $h(g)^{-1}$ such that $h(g)h(g)^{-1} = e'$ where e' is the identity element in H. Then applying $h(g)^{-1}$ from the right on the expression above, we obtain $h(e) = e'$.

(ii) For every $g \in G$ we have $h(g^{-1}) = h(g)^{-1}$. To prove this, note that we must have $h(g^{-1})h(g) = h(g^{-1}g) = h(e) = e'$, where e' again denotes the identity element in H.

Note that the definition of a homomorphism allows multiple elements of G to be mapped to a single element of H, so long as the multiplication structure is preserved. In other words, it may be that $h(g_1) = h(g_2)$ for some $g_{1,2} \in G$ with $g_1 \neq g_2$. The technical termology for this is that a homomorphism is not necessarily *injective*, or equivalently, not necessarily *one-to-one*.

The definition of a homorphism also does not require that every element of H is the image of some element of g. The technical terminology is that a homorphism is not necessarily *surjective*, or equivalently, not necessarily *onto*. See figure 1.1 for diagrams depcting injectivity and surjectivity.

Definition 1.3.2. A homomorphism that is both injective and surjective (equivalently, both one-to-one and onto) is called an **isomorphism**. This means that an isomorphism from a group G to a group H is a homomorphism that maps every element of G to a different element of H, and every element of H is the image of some element of G under the map. Two groups related by an isomorphism are said to be **isomorphic** to one-another. If G and H are isomorphic, we write $G \cong H$.

An immediate consequence of the definition is that two groups that are isomorphic to one-another must contain the same number of elements. We will think of groups that are

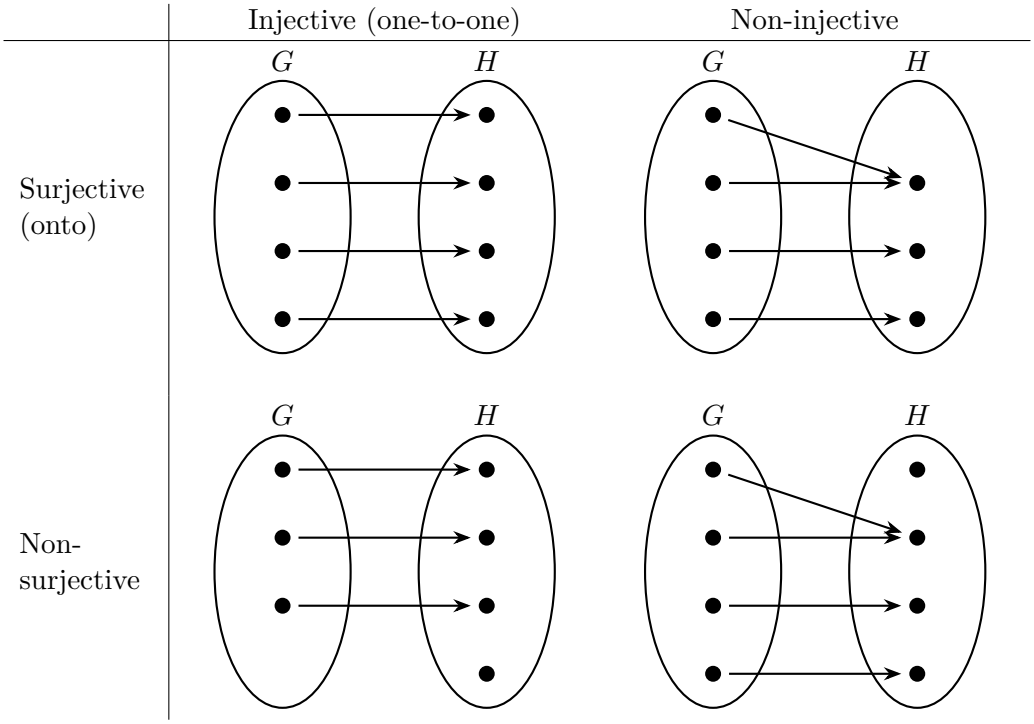


Figure 1.1: Maps from a group G to a group H. Each dot represents a group element, and the arrows show which element of H each element of G is mapped to. A group homomorphism may be non-injective and/or non-surjective. A homomorphism that is bijective (both injective and surjective) is called an isomorphism.

isomorphic as essentially the same group, written in different ways.

Example 1.3.1. A **cyclic group** is a group generated from a single element. If G is a cyclic group of order N, it takes the form

$$G = \{e, g, g^2, g^3, \cdots, g^{N-1}\}, \tag{1.18}$$

where $g^2 = gg$, $g^3 = gg^2$, and so on, with each of g, g^2, $\cdots$, g^{N-1} distinct. We say that G is **generated** by g. For the group multiplication to close we must have $g^N = e$. Any cyclic group of order N is isomorphic to the group $\mathbb{Z}_N$ defined in example 1.1.4 with the isomporphism provided by $\Phi : G \to \mathbb{Z}_N$ with

$$\Phi(g^n) = e^{2i\pi n/N}, \tag{1.19}$$

where we define $g^0 = e$ and $g^1 = g$. This is why we called $\mathbb{Z}_N$ *the* cyclic group of order N — any other cyclic group of the same order is isomorphic to it, and thus essentially the same group written in different notation.

The group $G = \{e, a, b\}$ in example 1.1.1 is cyclic, since $b = a^2$. The group is thus isomorphic to $\mathbb{Z}_3$.

Example 1.3.2. The group of six elements from example 1.1.2 is isomorphic to the dihedral group D_3, i.e. the group of symmetries of the equilateral triangle discussed in the preface. The isomorphism is

$$e \mapsto e, \qquad a \mapsto R_1, \qquad b \mapsto R_2, \qquad c \mapsto r_A, \qquad d \mapsto r_B, \qquad f \mapsto r_c. \tag{1.20}$$

From now on we will refer to the group $\{e, a, b, c, d, f\}$ as D_3, and we will think of (a, b) as rotations and (c, d, f) as reflections.

1.4 Subgroups

Sometimes a group may have a subset of elements that itself has the structure of a group. This motivates the definition of a subgroup:

Definition 1.4.1. Given a group G, a **subgroup** is a non-empty subset of elements $H \subset G$ that itself forms a group according to definition 1.1.1, with the same group multiplication law as G. The **order** of a subgroup is the number of elements it contains.

Since the multiplication is inherited from the group G, it is automatically associative. The important points are therefore: that H should be closed under multiplication, that is that $h_1 \circ h_2 \in H$ for all $h_{1,2} \in H$; that H should contain the identity element e; and that every element of H should have an inverse that is also in H. A group always has at least two subgroups:

- The definition of a subgroup implies that G is a subgroup of itself.
- The set containing only the identity element $H = \{e\}$ is also a subgroup.

Since these are rather trivial, we give a special name to non-trivial subgroups:

Definition 1.4.2. Given a group G, a **proper subgroup** of G is a subgroup that is not equal to G, and not equal to $\{e\}$.

Example 1.4.1. The group D_3 from example 1.1.2 has a proper subgroup consisting of the elements $\{e, a, b\}$. To see this, let's take another look at the multiplication table, putting in bold the part of it corresponding to these three elements:

	$\boldsymbol{e}$	$\boldsymbol{a}$	$\boldsymbol{b}$	c	d	f
$\boldsymbol{e}$	$\boldsymbol{e}$	$\boldsymbol{a}$	$\boldsymbol{b}$	c	d	f
$\boldsymbol{a}$	$\boldsymbol{a}$	$\boldsymbol{b}$	$\boldsymbol{e}$	f	c	d
$\boldsymbol{b}$	$\boldsymbol{b}$	$\boldsymbol{e}$	$\boldsymbol{a}$	d	f	c
c	c	d	f	e	a	b
d	d	f	c	b	e	a
f	f	c	d	a	b	e

We see that the part of the multiplication table corresponding only to multiplication of the elements $\{e, a, b\}$ is closed, since the bolded section does not contain any of the other group elements. All the other group axioms are satisfied by the subset $\{e, a, b\}$ too: in fact the bolded section of the multiplication table is identical to the multiplication table

of the group appearing in example 1.1.1. This shows us that this group is isomorphic to the subgroup that we are considering.

The existence of this subgroup can be understood from the isomorphism in example 1.3.2. We can think of a and b as rotations, while the other group elements (apart from the identity) are reflections. The subgroup is closed because you can never combine rotations to make a reflection.

Example 1.4.2. Let G be the group of two-dimensional rotation matrices considered in example 1.1.5. Let H be the subset of elements corresponding to rotations through angles which are multiples of $\pi/2$, $H = \{R(0), R(\pi/2), R(\pi), R(3\pi/2)\}$, explicitly,

$$H = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \right\}. \tag{1.21}$$

This forms a group since it is closed under multiplication — composing two rotations through multiples of $\pi/2$ gives another rotation through a multiple of $\pi/2$ — it contains the identity element $R(0)$, and each element has an inverse, $R(0) = R(0)^{-1}$, $R(\pi/2) = R(3\pi/2)^{-1}$ and $R(\pi) = R(\pi)^{-1}$. Thus H is a proper subgroup of G.

Definition 1.4.3. Given group elements g and h, forming the combination ghg^{-1} is called **conjugation** of h by g. A subgroup H of a group G is a **normal** subgroup if it is closed under conjugation by all group elements. This means that it satisfies $gHg^{-1} = H$ for all $g \in G$. By this notation, we mean

$$gHg^{-1} = \{gh_1g^{-1}, gh_2g^{-1}, \cdots\}, \tag{1.22}$$

where h_i are the elements of H. The trivial subgroup $\{e\} \subseteq G$ is always a normal subgroup, since $geg^{-1} = gg^{-1} = e$. Likewise, the trivial subgroup $G \subseteq G$ is always a normal subgroup, by closure of the group multiplication and the permutation law of theorem 1.2.6.

Example 1.4.3. The subgroup of D_3 in example 1.4.1 is a normal subgroup. Since a subgroup is closed under multiplication by definition, H is automatically closed under conjugation by any element of H: $eHe^{-1} = aHa^{-1} = bHb^{-1} = H$. We therefore just need to check conjugation by the other group elements. Noting that c, d, and f are each their own inverses, we have

$$\begin{aligned} cHc^{-1} &= cHc = c\{c, f, d\} = \{e, b, a\} = H, \\ dHd^{-1} &= dHd = d\{d, c, f\} = \{e, b, a\} = H, \\ fHf^{-1} &= fHf = f\{f, d, c\} = \{e, b, a\} = H. \end{aligned} \tag{1.23}$$

So H is indeed closed under conjugation, making it a normal subgroup.

Definition 1.4.4. Suppose a group G has a normal subgroup N and another subgroup H, such that:

(i) Every $g \in G$ may be written as $g = nh$ for some $n \in N$ and $h \in H$.

(ii) The only element of G belonging to both N and H is the identity.[3]

We then say that G is the **semi-direct product** of N and H. This is denoted

$$G = N \rtimes H, \tag{1.24}$$

Theorem 1.4.1. *If $G = N \rtimes H$, then the decomposition $g = nh$ is unique.*

Proof. Suppose that the decomposition was not unique. This would mean that we could write $g = nh = n'h'$ for two different $n, n' \in N$ and two different $h, h' \in H$. But then

$$n^{-1}gh'^{-1} = hh'^{-1} = n^{-1}n'. \tag{1.25}$$

By closure of subgroups, $hh'^{-1} \in H$ and $n^{-1}n' \in N$. Since the only group element shared between G and N is the identity, we must therefore have that $hh'^{-1} = n^{-1}n' = e$, and therefore $h' = h$ and $n' = n$. This contradicts the assumption that h is different from h' and/or n is different from n'. □

Example 1.4.4. We have already seen in example 1.4.3 that $N = \{e, a, b\} \cong \mathbb{Z}_3$ is a normal subgroup of D_3. We can form another subgroup H of D_3 consisting of the identity and one of the reflections. We choose $H = \{e, c\}$. Note that $H \cong \mathbb{Z}_2$. Every element of D_3 can be written as an element of N times an element of H,

$$\begin{aligned} e &= ee, \quad a = ae, \quad b = be, \\ c &= ec, \quad d = bc, \quad f = ac. \end{aligned} \tag{1.26}$$

Therefore D_3 is the semidirect product $\mathrm{D}_3 = N \rtimes H$.

1.5 Equivalence classes

In the next two sections we will discuss some results about the structure of group elements. To do so, we will need to introduce a mathematical concept from set theory called an equivalence class. Suppose we have a set of objects — for us the set will be a group — and a rule by which we can say that two elements of the set are related. It turns out that such a rule partitions the set into subsets called equivalence classes, with every element appearing in exactly one equivalence class.

For example, consider the set of integers $\mathbb{Z}$. If we define two integers to be equivalent if the difference between them is an even number, then this divides the set of integers into two equivalence classes. Once equivalence class is the set of even numbers, since the difference between two even numbers is again an even number. The other equivalence class is the set of all odd numbers, since the difference between two odd numbers is again an even number.

[3] More concisely, we could say that N and H have trivial intersection $N \cap H = \{e\}$.

Definition 1.5.1. Given a set S, suppose we have some rule by which we can say that two elements a, b of the set are related, which we write as $a \sim b$. This relation is an **equivalence relation** if it satisfies the following properties:

(i) Reflexivity: $a \sim a$.

(ii) Symmetry: $a \sim b$ implies that $b \sim a$.

(iii) Transitivity: $a \sim b$ and $b \sim c$ implies that $a \sim c$.

Returning to the example of the integers mentioned above, we would say that $a, b \in \mathbb{Z}$ are equivalent, $a \sim b$, if $a - b$ is an even number. This is reflexive since $a - a = 0$, which is even, so $a \sim a$. The equivalence is symmetric since if $a - b$ is even then so is $b - a$. Finally, it is transitive since if $a \sim b$ and $b \sim c$, then we have that $a - b$ and $b - c$ are even, and therefore $a - c = (a - b) - (b - c)$ is also even. Thus $\sim$ in this example is an equivalence relation.

Definition 1.5.2. Given a set S with an equivalence relation $\sim$, we define the **equivalence class** S_a of an element $a \in S$ as the subset of elements $b \in S$ satisfying $a \sim b$:

$$S_a = \{b \in S \mid a \sim b\}. \tag{1.27}$$

Reflexivity implies that $a \in S_a$.

We will frequently talk about equivalence classes being the same or different. By different we do *not* mean that they are defined using different elements a and b, we rather mean that they consist of different subsets of the elements of S. Conversely, if the equivalence classes S_a and S_b consist of the same subset of S then we think of them as the same equivalence class $S_a = S_b$. Returning again to the example of the integers, there are exactly two equvialence classes for the equivalence relation we defined: the even numbers and the odd numbers.

Theorem 1.5.1. *Two different equivalence classes can have no element in common.*

Proof. Consider two equivalence classes S_a and S_b. Suppose that they have an element in common, so that there is some $c \in S_{a,b}$, meaning $a \sim c$ and $b \sim c$. Then transitivity implies that $a \sim b$. Now:

- Suppose that there is some $d \in S_b$, so that $d \sim b$. Then transitivity implies that $a \sim d$, so we also have that $d \in S_a$. Thus every element of S_b is also an element of S_a.
- Similarly, suppose that there is some $d' \in S_a$. Then transitivity implies that $d' \in S_b$, so every element of S_a is also an element of S_b.

We have learned that S_a and S_b consist of the same subset of elements of S, and they are therefore the same equivalence class. Two different equivalence classes must therefore consist of two disjoint subsets. □

Every element of a set belongs to an equivalence class (since $a \in S_a$) but, according to theorem 1.5.1, no more than one equivalence class. An equivalence relation therefore partitions a set:

Theorem 1.5.2. *Given a set S and an equivalence relation $\sim$, every element of S belongs to exactly one equivalence class.*

1.6 Cosets

Group multiplication tables often show some structure. For instance — with the order in which we have chosen to write its rows and columns — the multiplication table in example 1.4.1 splits into 3×3 blocks, with the elements $\{e, a, b\}$ appearing only in the top-left and bottom-right blocks, and $\{c, d, f\}$ appearing only in the top-right and bottom-left blocks. We saw in the example that the elements $\{e, a, b\}$ are special because they form a subgroup, but what about $\{c, d, f\}$? We will now see that a subgroup can be used to sort the elements of a group into equal-sized sets called cosets, with each element of the group appearing in one and only one coset. This will be our first application of the machinery of equivalence classes.

Definition 1.6.1. Let G be a finite group with a proper subgroup H. A **left coset** of H in G is a set of elements of the form

$$gH = \{gh_1, gh_2, \cdots\}, \tag{1.28}$$

where $g \in G$ and $h_i \in H$. Similarly, a **right coset** is a set of elements of the form Hg. It is often the case that two elements $g_{1,2} \in G$ will lead to the cosets containing the same elements. If this is the case, we regard the cosets as the same: $g_1 H = g_2 H$.

Every element $g \in G$ must belong to at least one left coset, since by construction H contains the identity element, and therefore $g \in gH$. Similarly, every element belongs to at least one right coset $g \in Hg$.

We will focus on left cosets in the following, often dropping the prefix "left" where this will not cause ambiguity. Similar results and proofs apply to right cosets, with minor modifications. Notice that every coset gH has the same order as H, i.e. every coset contains $|H|$ elements. The only way this could be otherwise would be if $gh_i = gh_j$ for at least two distinct elements $h_{i,j} \in H$. However, the cancellation law from theorem 1.2.5 tells us that it is not possible to have $gh_i = gh_j$ for $h_i \neq h_j$.

Example 1.6.1. The $\mathrm{D}_3 = \{e, a, b, c, d, f\}$ with multiplication table given in example 1.1.2 has a subgroup $H = \{e, a, b\}$, as we saw in example 1.4.1. The cosets are

$$\begin{aligned} eH &= \{ee, ea, eb\} = \{e, a, b\}, & cH &= \{ce, ca, cb\} = \{c, d, f\}, \\ aH &= \{ae, aa, ab\} = \{a, b, e\}, & dH &= \{de, da, db\} = \{d, f, c\}, \\ bH &= \{be, ba, bb\} = \{b, e, a\}, & fH &= \{fe, fa, fb\} = \{f, c, d\}. \end{aligned} \tag{1.29}$$

Since the order of elements in the group is immaterial, we have that $eH = aH = bH$, and $cH = dH = fH$, so there are two distinct cosets.

Theorem 1.6.1. *Two group elements $g_{1,2} \in G$ belong to the same coset if and only if $g_1^{-1} g_2 \in H$.*

Proof. Let's first establish the "*if*" in the theorem. Suppose that two group elements satisfy $g_1^{-1}g_2 \in H$. Let gH be the coset containing g_2, where g is some element of G. By definition this means that we can write $g_2 = gh$ for some $h \in H$. Then we have that

$$g_2 = gh = gh(g_1^{-1}g_2)^{-1}(g_1^{-1}g_2). \tag{1.30}$$

If $g_1^{-1}g_2 \in H$ then we must also have that $(g_1^{-1}g_2)^{-1} \in H$, since H is a subgroup and so every element of H should have an inverse also in H. Then by closure of the subgroup, $h' \equiv h(g_1^{-1}g_2)^{-1} \in H$. We therefore have

$$g_2 = gh'g_1^{-1}g_2 \quad \Rightarrow \quad g_1 = gh', \tag{1.31}$$

where the equation on the right is obtained from the equation on the left by acting from the right first with g_2^{-1} and then with g_1. We therefore have that $g_1 \in gH$, so $g_{1,2}$ are part of the same coset if $g_1^{-1}g_2 \in H$.

Now we establish the "*only if*". Suppose that $g_{1,2}$ belong to the same coset, which we'll call gH for some $g \in G$. This means that we can write $g_1 = gh_1$ and $g_2 = gh_2$ for some $h_{1,2} \in H$. Then we have that

$$g_1^{-1}g_2 = h_1^{-1}g^{-1}gh_2 = h_1^{-1}h_2. \tag{1.32}$$

The right-hand side must belong to H, by closure of the subgroup. Thus $g_1^{-1}g_2 \in H$ if g_1 and g_2 belong to the same coset. □

We now prove the result mentioned above:

Theorem 1.6.2. *Given a finite group G with subgroup H, every element of G belongs to one and only one left coset.*

Proof. Consider two group elements $g_{1,2} \in G$. Let us define $g_1 \sim g_2$ if $g_1^{-1}g_2 \in H$. The relation $\sim$ satisfies all the conditions to be an equivalence relation:

(i) Reflexivity: $g \sim g$ since $g^{-1}g = e$, which must be a member of any subgroup.

(ii) Symmetry: If $g_1^{-1}g_2 \in H$ then we must also have that $g_2^{-1}g_1 = (g_1^{-1}g_2)^{-1} \in H$. Thus $g_1 \sim g_2$ implies that $g_2 \sim g_1$.

(iii) Transitivity: If $g_1^{-1}g_2 \in H$ and $g_2^{-1}g_3 \in H$, then by closure of the group multiplication we must have that $g_1^{-1}g_3 = (g_1^{-1}g_2)(g_2^{-1}g_3) \in H$. Thus $g_1 \sim g_2$ and $g_2 \sim g_3$ implies that $g_1 \sim g_3$.

Theorem 1.6.1 implies that the definition of $\sim$ given above is equivalent to saying that $g_1 \sim g_2$ if g_1 and g_2 belong to the same coset. The equivalence class of an element $g \in G$ is therefore a left coset as defined in equation (1.28). We see that the left cosets form a set of equivalence classes, so by theorem 1.5.2 we have that every element of G belongs to one and only one left coset. □

The result in theorem 1.6.2 implies that a proper subgroup partitions a group into equal size subsets, with a number of elements equal to the order of H (since every coset contains $|H|$ elements). This is only possible if $|G|$ is an integer multiple of $|H|$:

Corollary 1.6.1. *If a group G has a subgroup H, then $|H|$ is a divisor of $|G|$.*

Corollary 1.6.2. *If the order of a group G is a prime number, then G has no proper subgroups.*

The number $|G|/|H|$, which is an integer by corollary 1.6.1, is the total number of cosets. Corollary 1.6.2 implies a rather remarkable conclusion about groups with prime order:[4]

Theorem 1.6.3. *If the order of a group G is a prime number p, then G is isomorphic to the cyclic group $\mathbb{Z}_p$, and is therefore abelian.*

Proof. Consider an element $g \in G$ other than the identity. Let q be the order of g (recall that this means that q is the smallest positive integer such that $g^q = e$). Theorem 1.2.7 implies that $q \leq p$. Consider the set of elements

$$H = \{e, g, g^2, \cdots, g^q\}. \tag{1.33}$$

This set is closed under multiplication and thus forms a subgroup of H. This subgroup is cyclic, by construction. Corollary 1.6.2 implies that H cannot be a proper subgroup. Furthermore, H cannot be the trivial subgroup $H = \{e\}$ since by construction H contains $g \neq e$. Thus the only possibility is that H is the trivial subgroup $H = G$. Thus G is a cyclic group, and therefore as discussed in example 1.1.4 is isomorphic to $\mathbb{Z}_p$. □

Recall that a subgroup H of a group G is normal if $gHg^{-1} = H$ for every $g \in G$. Rearranging, we see that a normal subgroup satisfies $gH = Hg$. This tells us that the left and right cosets of a normal subgroup are identical. This fact implies that cosets defined using a normal subgroup have their own natural group structure. Given a group G and a normal subgroup H, we define the product of any two cosets g_1H and g_2H by

$$(g_1H)(g_2H) = g_1g_2H. \tag{1.34}$$

This product is well-defined for normal subgroups because it does not depend on the specific elements g_1 and g_2 we use to represent the cosets. Suppose that $g_2H = g_2'H$ for some other element $g_2' \in G$. From theorem 1.6.1 this implies that $g_2^{-1}g_2' \in H$. Then

$$(g_1H)(g_2'H) = g_1g_2'H = g_1g_2g_2^{-1}g_2'H = g_1g_2H = (g_1H)(g_2H) \tag{1.35}$$

Similarly, suppose that $g_1H = g_1'H$ for some $g_1' \in G$. Then

$$(g_1'H)(g_2H) = g_1'g_2H = g_1g_1^{-1}g_1'g_2H = g_1g_1^{-1}g_1'Hg_2, \tag{1.36}$$

where in the final equality we have made use of the fact that H is normal, so that $g_2H = Hg_2$. Since g_1 and g_1' are in the same coset, we have that $g_1^{-1}g_1' \in H$ and therefore $g_1^{-1}g_1'H = H$. Therefore, equation (1.36) implies that

$$(g_1'H)(g_2H) = g_1Hg_2 = g_1g_2H = (g_1H)(g_2H). \tag{1.37}$$

[4] As an aside, a finite group with no proper, *normal* subgroups is called simple. Finite simple groups have been classified: every finite simple group belongs to one of several infinite families, or is one of 26 exceptional cases (called sporadic groups). In theorem 1.6.3 we see that one of the infinite families is $\mathbb{Z}_p$ with p prime. You may have heard of the monster group, which is the largest sporadic group containing roughly 8×10^{53} elements.

Theorem 1.6.4. *Given a group G, the cosets of a normal subgroup H form a group under the group multiplication defined in equation* (1.36).

Proof. We need to check the group axioms in definition 1.1.1:

(i) Closure is guaranteed by the fact that every element of G belongs to a coset, so we can never end up with something on the right-hand side of equation (1.36) that is not a coset.

(ii) Associativity is inherited from associativity of the product on G. For any $g_{1,2,3} \in G$:

$$(g_1H)((g_2H)(g_3H)) = g_1(g_2g_3)H = (g_1g_2)g_3H = ((g_1H)(g_2H))(g_3H). \tag{1.38}$$

(iii) The identity element of the new group is the coset eH containing the identity element of G, since $(eH)(gH) = (eg)H = gH$.

(iv) The inverse of a coset gH is the coset $g^{-1}H$, since $(g^{-1}H)(gH) = (g^{-1}gH) = eH$.

□

Definition 1.6.2. Given a group G, the group of cosets of a normal subgroup H with multiplication law (1.34) is called the **quotient group**, and is denoted G/H.

The order of a quotient group is the number of cosets, $|G/H| = |G|/|H|$. Every group G has two trivial normal subgroups, $\{e\}$ and G. The corresponding quotient groups are $G/\{e\} \cong G$ — since the cosets of $\{e\}$ are just the group elements — and $G/G \cong \{e\}$ — since there is only one coset of G inside G, containing all of the group elements.

Example 1.6.2. The subgroup H in example 1.6.1 is normal, as was seen in example 1.4.3. The cosets of H therefore form a group. There are two distinct cosets, which we take to be eH and cH. The multiplication table of G/H is:

	eH	cH
eH	eH	cH
cH	cH	eH

The quotient group is isomorphic to the cyclic group of order two, $G/H \cong \mathbb{Z}_2$. Recall that $\mathbb{Z}_2 = \{1, -1\}$, with multiplication provided by ordinary multiplication of complex numbers. The isomorphism is $eH \mapsto 1$ and $cH \mapsto -1$.

1.7 Conjugacy classes

We will now discuss another way to partition the elements of a group. As we will see in the next chapter, conjugacy classes are important for the representation theory of finite groups.

Definition 1.7.1. We define two group elements to be **conjugate** to one another if they can be related by conjugation. This means that $a, b \in G$ are conjugate if there exists another group element $g \in G$ such that $a = gbg^{-1}$.

Theorem 1.7.1. *Conjugacy is an equivalence relation.*

Proof. We define $a \sim b$ if there exists another group element g such that $a = gbg^{-1}$. We need to show that $\sim$ satisfies the three defining properties of an equivalence relation:

(i) Reflexivity: Every group element is conjugate to itself, since $a = eae^{-1}$, so $a \sim a$.

(ii) Symmetry: Suppose $a \sim b$, so that $a = gbg^{-1}$ for some $g \in G$. Then $b = hah^{-1}$ where $h = g^{-1} \in G$, implying that $b \sim a$.

(iii) Transitivity: Suppose $a \sim b$ and $b \sim c$, so that $a = gbg^{-1}$ and $b = hch^{-1}$ for $g, h \in G$. Then we have that $a = ghch^{-1}g^{-1} = (gh)c(gh)^{-1}$, and since $gh \in G$ by closure of the group multiplication this implies that $a \sim c$.

Therefore, conjugacy satisfies all of the requirements to be an equivalence relation. □

Definition 1.7.2. The equivalence classes defined by conjugacy are called **conjugacy classes**. This means that the conjugacy class S_h of a group element $a \in G$ is the set of all group elements that may be written as ghg^{-1} for some $g \in G$:

$$S_h = \{a \in G \mid a = ghg^{-1}, g \in G\}. \tag{1.39}$$

The set S of elements in a conjugacy class is in general not a subgroup. Since conjugacy classes are equivalence classes, from theorem 1.5.2 we know the following:

Corollary 1.7.1. *Every element of a group belongs to one and only one conjugacy class.*

Conjugacy classes therefore provide another partition of the group elements. An important distinction from the partition into cosets is that not every conjugacy class has the same size. Note that the identity element of any group always sits in a conjugacy class all by itself, $S_e = \{e\}$. This is because any group element h conjugate to h would have to satisfy $h = geg^{-1}$ for some $g \in G$, but the right-hand side of this equation is $geg^{-1} = gg^{-1} = e$.

Example 1.7.1. Let's find the conjugacy classes of the group $D_3 = \{e, a, b, c, d, f\}$, with multiplication table given in example 1.1.2. We've already seen that one conjugacy class is $S_e = \{e\}$. To find S_a, we test the result of conjugating a by each group element,

$$\begin{aligned} eae^{-1} &= eae = a, & aaa^{-1} &= aab = a, & bab^{-1} &= baa = a, \\ cac^{-1} &= cac = b, & dad^{-1} &= dad = b, & faf^{-1} &= faf = b. \end{aligned} \tag{1.40}$$

The conjugacy class is the collection of elements on the right-hand side, so we see that $S_a = \{a, b\}$. By corollary 1.7.2 we know that b cannot belong to more than one conjugacy class, so we must have $S_b = S_a$. Now let's find S_c,

$$ece^{-1} = ece = c, \qquad aca^{-1} = acb = d, \qquad bcb^{-1} = bca = f, \tag{1.41}$$

$$ccc^{-1} = ccc = c, \qquad dcd^{-1} = dcd = f, \qquad fcf^{-1} = fcf = d. \tag{1.42}$$

We read off that $S_c = \{c, d, f\}$. Since d and f cannot belong to more than one conjugacy

class, we have that $S_d = S_f = S_c$. In summary, D_3 has three conjugacy classes,

$$\{e\}, \qquad \{a, b\}, \qquad \{c, d, f\}. \tag{1.43}$$

Chapter 2

Representations of finite groups

In this chapter we develop representation theory. Many of the definitions and results in this chapter apply both to finite and infinite groups, but some will only hold for finite groups. We will point these out along the way. Most of the chapter is quite light on examples. We will first develop the abstract machinery and then apply it to an number of groups towards the end of the chapter.

2.1 Definition of a representation

Definition 2.1.1. A **representation** of a group G on a complex vector space V is a map D from G to linear operators acting on V, satisfying the following properties:

(i) $D(e) = \mathbb{1}$, the identity operator of the space on which the linear operators act.

(ii) $D(g_1)D(g_2) = D(g_1 g_2)$ for any $g_{1,2} \in G$, so the map preserves the group multiplication structure.

These defining properties of a representation imply that

$$D(g)D(g^{-1}) = D(gg^{-1}) = D(e) = \mathbb{1}, \tag{2.1}$$

so the operator $D(g^{-1})$ is the inverse of the operator $D(g)$. If V is n-dimensional then we also describe the representation as n-dimensional.

In the following, we will frequently wish to say things like "D is a representation of a group G acting on some vector space V". We will abbreviate such statements as

$$D : G \to \mathrm{GL}(V), \tag{2.2}$$

leaving implicit that G is a group and V is a complex vector space. The initials GL stand for "general linear" and $\mathrm{GL}(V)$ is the group of invertible transformations that map the vector space V to itself. For example $\mathrm{GL}(\mathbb{R}^n)$ is the group of $n \times n$ real matrices and $\mathrm{GL}(\mathbb{C}^n)$ is the group of $n \times n$ complex matrices. The transformations need to be invertible dince $D(g^{-1}) = D(g)^{-1}$.

If we use a basis to write elements of the vector space as n-component column vectors then the representation will be a set of $n \times n$ matrices, with the multiplication $D(g_1)D(g_2)$

being ordinary matrix multiplication. Then $D(e)$ will be the n-dimensional identity matrix, and $D(g)^{-1}$ will be the matrix inverse of $D(g)$. We will frequently use this matrix notation.

Definition 2.1.2. Note that a valid representation is to set $D(g) = \mathbb{1}$ for all g. Such a representation is known as **trivial**, but don't let the name fool you, trivial representations are very important in physics. Degrees of freedom that transform in the trivial representation of a group corresponding to some symmetry are invariant under that symmetry.[1] In matrix notation, every group has a set of trivial representations in which we represent all of the group elements by the $n \times n$ identity matrix.

More generally, it may be that more than one group element is representated by the same operator. In other words, the map D may be non-injective (see figure 1.1).

Definition 2.1.3. If the map *is* injective, that is if every group element is represented by a different operator, then the representation is called **faithful**.

Example 2.1.1. Consider the group in example 1.1.1, containing three elements $\{e, a, b\}$ with multiplication table

	e	a	b
e	e	a	b
a	a	b	e
b	b	e	a

This is a two-dimensional faithful representation

$$D(e) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad D(a) = \begin{pmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}, \qquad D(b) = \begin{pmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}. \tag{2.3}$$

It satisfies the requirements to be a representation since (i) $D(e)$ is represented by the identity matrix $\mathbb{1}$, and (ii) it preserves the group multiplication. Since $D(e) = \mathbb{1}$, preservation of products involving $D(e)$ are trivial, $D(e)D(g) = \mathbb{1}D(g) = D(g) = D(eg)$ for any $g \in G$. The other products may be checked by explicit matrix multiplication, for example we have that $aa = b$ and

$$D(a)D(a) = \frac{1}{4}\begin{pmatrix} -1 & -\sqrt{3} \\ \sqrt{3} & -1 \end{pmatrix}\begin{pmatrix} -1 & -\sqrt{3} \\ \sqrt{3} & -1 \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix} = D(b). \tag{2.4}$$

You should check that the other products satisfy $D(a)D(b) = D(e)$ and $D(b)D(b) = D(a)$.

Here is a way to construct a faithful representation that exists for any finite group:[2]

[1] For example, scalars transform trivially under rotations, while vectors transform non-trivially.

[2] To see that the regular representation is faithful, suppose toward a contradiction that two different group elements g and h were represented by the same operator, $D(g) = D(h)$. Then we would have that for every group element f that $D(g)|f\rangle = D(h)|f\rangle$. By the defining property of the regular representation, this means that $|gf\rangle = |hf\rangle$ and hence $gf = hf$ for all f. By the cancellation law 1.2.5 this implies that $g = h$, contradicting the assumption that g and h are different group elements.

Definition 2.1.4. For any group G, the **regular** representation is a $|G|$-dimensional representation constructed by taking the group elements to form the basis elements of a vector space and defining the action of the operators through

$$D_R(g_1)|g_2\rangle = |g_1 g_2\rangle, \tag{2.5}$$

for all $g_{1,2} \in G$ and where we use the ket $|g\rangle$ to denote the vector corresponding to the group element g. This forms a representation since we have that for all $g \in G$

$$D_R(e)|g\rangle = |eg\rangle = |g\rangle, \tag{2.6}$$

so $D_R(e)$ is indeed the identity operator, while for all $g_{1,2,3} \in G$ we have

$$D_R(g_1)D_R(g_2)|g_3\rangle = D_R(g_1)|g_2 g_3\rangle = |g_1 g_2 g_3\rangle = |(g_1 g_2)g_3\rangle = D_R(g_1 g_2)|g_3\rangle, \tag{2.7}$$

so $D_R(g_1)D_R(g_2) = D_R(g_1 g_2)$ and hence the group multiplication is preserved.

Example 2.1.2. Let's construct the regular representation of the group in example 2.1.1. Since the group has three elements, its regular representation must be three-dimensional. As always, $D_R(e)$ must be the identity operator. To construct $D_R(a)$ we use the multiplication table to find[3]

$$D_R(a)|e\rangle = |a\rangle, \qquad D_R(a)|a\rangle = |b\rangle, \qquad D_R(a)|b\rangle = |e\rangle. \tag{2.8}$$

Similarly, we can use the multiplication table to find that $D_R(b)$ acts as

$$D_R(b)|e\rangle = |b\rangle, \qquad D_R(b)|a\rangle = |e\rangle, \qquad D_R(b)|b\rangle = |a\rangle. \tag{2.9}$$

Writing each basis element as a column vector,

$$|e\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \qquad |a\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \qquad |b\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \tag{2.10}$$

Then the regular representation is

$$D_R(e) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad D_R(a) = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \qquad D_R(b) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}. \tag{2.11}$$

It's straightforward to check that these matrices preserve the group multiplication.

Given a representation D and an invertible linear operator S, we can construct a new representation D' by the similarity transform

$$D'(g) = S^{-1}D(g)S, \tag{2.12}$$

[3] Linearity then defines how $D_R(a)$ acts on a general element of the vector space,

$$D_R(a)(c_e|e\rangle + c_a|a\rangle + c_b|b\rangle) = c_e|a\rangle + c_a|b\rangle + c_b|e\rangle.$$

for every group element g. It is straightforward to confirm that D' satisfies the two requirements to be a representation: $D'(e) = S^{-1}D(e)S = S^{-1}S = \mathbb{1}$, so the identity element is represented by the identity operator, and

$$D'(g_1)D'(g_2) = S^{-1}D(g_1)SS^{-1}D(g_2)S = S^{-1}D(g_1)D(g_2)S = S^{-1}D(g_1g_2)S = D'(g_1g_2), \tag{2.13}$$

so D' preserves the group multiplication. Similarity transformations of operators arise when we make a change of basis in the vector space on which the operators act (see section **??**). We therefore do not wish to think of the representations D and D' as different in any meaningful way, motivating the following definition:

Definition 2.1.5. Two representations D and D' of the same group are **equivalent** if they are related by a similarity transformation, $D'(g) = S^{-1}D(g)S$ for some invertible linear operator S for every $g \in G$.

Every complex vector space V admits an inner product. If $\{|e\rangle_i\}$ is a basis for V and $|u\rangle = \sum_i u_i|e_i\rangle$ and $|v\rangle = \sum_i v_i|e_i\rangle$, then

$$\langle u, v\rangle = \sum_i u_i^* v_i \tag{2.14}$$

satisfies the requirements to be an inner product on V. We will always assume this inner product.

An inner product induces a natural isomorphism between V and its dual space V^*, the space of linear functions on V. The image of $|u\rangle \in V$ under this isomorphism is $\langle u| \in V^*$ defined by

$$\langle u|v\rangle = \langle u, v\rangle, \tag{2.15}$$

for all $|v\rangle \in V$, where we use the braket $\langle u|v\rangle$ to denote the linear function $\langle u|$ evaluated on the vector $|v\rangle$. Using this isomorphism, we will always just write the inner product in equation (2.14) as $\langle u|v\rangle$.

Often operators which are unitary with respect to a given inner product are particularly important. For instance, symmetry transformations in quantum mechanics are generated by unitary operators. It is therefore useful to define:

Definition 2.1.6. A representation D of a group G is **unitary** (with respect to a given inner product) if $D(g)$ is a unitary operator $D(g)^\dagger = D(g)^{-1}$ for every $g \in G$.

For example, the regular representation for $G = \{e, a, b\}$ constructed in example 2.1.2 is unitary since it is straightforward to verify that it satisfies $D(e)D(e)^\dagger = D(a)D(a)^\dagger = D(b)D(b)^\dagger = \mathbb{1}$.

2.2 New representations from old

2.2.1 Direct sums

The direct sum is a particular way of combining vector spaces together:

Definition 2.2.1. Suppose we have two vector spaces U and V. We can form a new vector space $U \oplus V$, called the **direct sum** of these vector spaces, as follows. The elements of $U \oplus V$ are ordered pairs $(|u\rangle, |v\rangle)$, where $|u\rangle \in U$ and $|v\rangle \in V$. To multiply these vectors by a scalar λ, we multiply both components: $\lambda(|u\rangle, |v\rangle) = (\lambda|u\rangle, \lambda|v\rangle)$. To add vectors, we add the components:

$$(|u_1\rangle, |v_1\rangle) + (|u_2\rangle, |v_2\rangle) = (|u_1\rangle + |u_2\rangle, |v_1\rangle + |v_2\rangle). \tag{2.16}$$

The direct sum $U \oplus V$ inherits its vector space structure from its subspaces U and V. The zero vector is $(|0\rangle, |0\rangle)$, in other words both components are the zero vectors of the respective subspaces. Given linear operators A and B that act respectively on U and V, we can construct a linear operator $A \oplus B$ acting on $U \oplus V$ by defining

$$(A \oplus B)(|u\rangle, |v\rangle) = (A|u\rangle, B|v\rangle). \tag{2.17}$$

If $|e_i\rangle$ and $|f_j\rangle$ are bases for U and V, respectively, then a basis for $U \oplus V$ is given by the set $(|e_i\rangle, 0)$ and $(0, |f_j\rangle)$. To see this, we note that an arbitrary element $(|u\rangle, |v\rangle) \in U \oplus V$ can be uniquely decomposed as

$$(|u\rangle, |v\rangle) = (|u\rangle, 0) + (0, |v\rangle). \tag{2.18}$$

We then decompose $|u\rangle$ and $|v\rangle$ in terms of the basis elements on U and V, respectively. From the number of basis elements we find that $\dim(U \oplus V) = \dim(U) + \dim(V)$.

The direct sum may seem abstract, but hopefully it becomes clearer if we think about it in a particular basis. Suppose U is m-dimensional and V is n-dimensional. In some basis, we can write elements of these vector spaces as

$$|u\rangle = \begin{pmatrix} u_1 \\ \vdots \\ u_m \end{pmatrix}, \qquad |v\rangle = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}. \tag{2.19}$$

Elements of $U \oplus V$ can then be written as

$$(|u\rangle, |v\rangle) = \begin{pmatrix} u_1 \\ \vdots \\ u_m \\ v_1 \\ \vdots \\ v_m \end{pmatrix}. \tag{2.20}$$

This naturally implements the multiplication and addition properties in the definition of the direct sum. Representing the vectors like this, the linear operator $A \oplus B$ becomes a block diagonal matrix,

$$A \oplus B = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}. \tag{2.21}$$

Notice that to compose two linear operators in the direct sum, we simply compose the operators acting on the subspaces,

$$(A_1 \oplus B_1)(A_2 \oplus B_2)(|u\rangle, |v\rangle) = (A_1 \oplus B_1)(A_2|u\rangle, B_2|v\rangle) = (A_1A_2|u\rangle, B_1B_2|v\rangle), \tag{2.22}$$

so we see that $(A_1 \oplus B_1)(A_2 \oplus B_2) = A_1A_2 \oplus B_1B_2$. The block diagonal matrix form in equation (2.21) natarually implements this, since

$$(A_1 \oplus B_1)(A_2 \oplus B_2) = \begin{pmatrix} A_1 & 0 \\ 0 & B_1 \end{pmatrix} \begin{pmatrix} A_2 & 0 \\ 0 & B_2 \end{pmatrix} = \begin{pmatrix} A_1A_2 & 0 \\ 0 & B_1B_2 \end{pmatrix} = (A_1A_2) \oplus (B_1B_2), \tag{2.23}$$

We can use this to combine two representations to form a new one:

Definition 2.2.2. Given two representations of the same group, $D : G \to \mathrm{GL}(V)$ and $D' : G \to \mathrm{GL}(V')$, we can construct a new representation $D'' = D \oplus D'$, called the **direct sum**, which acts on $V \oplus V'$ in the natural way. Expressed as a matrix, the element of D'' corresponding to a group element g is

$$D''(g) = \begin{pmatrix} D(g) & 0 \\ 0 & D'(g) \end{pmatrix}. \tag{2.24}$$

You should convince yourself that the direct sum satisfies the requirements to be a representation.

We can construct direct sums of more than one vector space, and thus more than one representation, inductively. For example, for three vector spaces we could form $U \oplus V \oplus W$ by first forming the sum $U \oplus V$, and then forming the direct sum of that vector space with W. Alternatively we could first form $V \oplus V$, and then form the direct sum of this with W, the result is the same. Elements of $U \oplus V \oplus W$ are triples $(|u\rangle, |v\rangle, |w\rangle)$. Written as column vectors in some basis they would be

$$(|u\rangle, |v\rangle, |w\rangle) = (u_1, \cdots, u_{d_U}, v_1, \cdots, v_{d_V}, w_1, \cdots, w_{d_W})^T, \tag{2.25}$$

where $d_X = \dim(X)$ is the dimension of the vector space X. In this basis, linear operators (including representation matrices) acting on $U \oplus V \oplus W$ would be block diagonal with three blocks.

2.2.2 Tensor products

The tensor product is a different way of combining vector spaces:

Definition 2.2.3. Given two vector spaces U and V, we can form a new vector space $U \otimes V$, called the **tensor product**, as follows. The building blocks of the tensor product are ordered pairs of elements of U and V, which we denote $|u\rangle \otimes |v\rangle$. The tensor product space $U \otimes V$ is then the space of all possible linear combinations of such pairs, of the form

$$a_1|u_1\rangle \otimes |v_1\rangle + a_2|u_2\rangle \otimes |v_2\rangle + \cdots, \tag{2.26}$$

where the a_i are scalar coefficients, $|u_i\rangle \in U$, and $|v_i\rangle \in V$. We require the tensor product to be bilinear,

$$\begin{aligned} (a_1|u_1\rangle + a_2|u_2\rangle) \otimes |v\rangle &= a_1|u_1\rangle \otimes |v\rangle + a_2|u_2\rangle \otimes |v\rangle, \\ |u\rangle \otimes (a_1|v_1\rangle + a_2|v_2\rangle) &= a_1|u\rangle \otimes |v_1\rangle + a_2|u\rangle \otimes |v_2\rangle. \end{aligned} \tag{2.27}$$

Given linear operators A and B acting on U and V, respectively, we define a linear

operator $A \otimes B$ acting on $U \otimes V$ by

$$(A \otimes B)|u\rangle \otimes |v\rangle = (A|u\rangle) \otimes (B|v\rangle), \tag{2.28}$$

with the action on linear combinations such as (2.26) determined by the requirement of linearity.

The direct sum and tensor product are both constructed out of pairs of elements from two vector spaces, and the action of linear operators looks very similar as well. The difference between the direct sum and the tensor product is how we add the pairs of vectors. For the direct sum, we had $(|u_1\rangle + |u_2\rangle, |v_1\rangle + |v_2\rangle) = (|u_1\rangle, |v_1\rangle) + (|u_2\rangle, |v_2\rangle)$, while for the tensor product we have

$$(|u_1\rangle + |u_2\rangle) \otimes (|v_1\rangle + |v_2\rangle) = |u_1\rangle \otimes |v_1\rangle + |u_1\rangle \otimes |v_2\rangle + |u_2\rangle \otimes |v_1\rangle + |u_2\rangle \otimes |v_2\rangle. \tag{2.29}$$

Not every element of the tensor product space can be written as a product of a single pair of states.

In physics, tensor products are familiar from the quantum mechanics of more than one particle. For instance, consider a system of two qubits, each with possible states $|0\rangle$ and $|1\rangle$. A general state in the two qubit system is

$$|\psi\rangle = a|0\rangle \otimes |0\rangle + b|0\rangle \otimes |1\rangle + c|1\rangle \otimes |0\rangle + d|1\rangle \otimes |1\rangle. \tag{2.30}$$

In quantum mechanics we call the states that cannot be written as a single product *entangled*. An example is the maximally entangled state $|\psi\rangle \propto |0\rangle \otimes |1\rangle + |1\rangle \otimes |0\rangle$.

Theorem 2.2.1. *If $\{|e_i\rangle\}$ and $\{|f_j\rangle\}$ are bases for the vector spaces U and V, respectively, then the set $\{|e_i\rangle \otimes |f_j\rangle\}$, with $i = 1, \cdots, \dim(U)$ and $j = 1, \cdots, \dim(V)$ forms a basis for the tensor product space $U \otimes V$. This implies that* $\dim(U \otimes V) = \dim(U)\dim(V)$.

Proof. We need to show that every vector in $U \otimes V$ can be written as a unique linear combination of the vectors $|e_i\rangle \otimes |f_j\rangle$. This follows from the bilinearity of the tensor product; given a generic element of $U \otimes V$, of the form

$$|\psi\rangle = \sum_{a,b} c_{ab}|u_a\rangle \otimes |v_b\rangle, \tag{2.31}$$

with $|u_a\rangle \in U$ and $|v_b\rangle \in V$, we decompose the vectors appearing in the product in terms of the basis elements of U and V,

$$|\psi\rangle = \sum_{a,b} c_{ab} \left(\sum_i x_{ai}|e_i\rangle \right) \otimes \left(\sum_j y_{bj}|f_j\rangle \right) = \sum_{i,j} C_{ij}|e_i\rangle \otimes |f_j\rangle, \tag{2.32}$$

where $C_{ij} = \sum_{a,b} c_{ab}x_{ai}y_{bj}$. Thus every vector in $U \otimes V$ may indeed be decomposed as a linear combination of the $|e_i\rangle \otimes |f_j\rangle$. This decomposition is unique because the vectors $|e_i\rangle \otimes |f_j\rangle$ are linearly independent (because the $|e_i\rangle$ and $|f_j\rangle$ are separately linearly independent in their respective vector spaces). Thus $\{|e_i\rangle \otimes |f_j\rangle\}$ indeed forms a basis. □

Tensor products can also be written down in terms of column vectors and matrices, but this works a little less cleanly than for direct sums, so it is not done so frequently. Writing $m = \dim(U)$ and $n = \dim(V)$, we can represent the vectors as follows:

$$\sum_{i,j} c_{ia}|e_i\rangle \otimes |f_a\rangle = (c_{11}, c_{12}, \cdots, c_{1n}, c_{21}, c_{22}, \cdots, c_{2n}, \cdots c_{mn})^T . \tag{2.33}$$

If we have linear operators with matrix elements A_{ij} and B_{ab} in the $|e_i\rangle$ and $|f_a\rangle$ bases, then $A \otimes B$ takes the block form

$$A \otimes B = \begin{pmatrix} A_{11}B & A_{12}B & \cdots & A_{1m}B \\ A_{12}B & A_{22}B & \cdots & A_{2m}B \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1}B & A_{m2}B & \cdots & A_{mm}B \end{pmatrix}, \tag{2.34}$$

where each of the B's is an $n \times n$ block with elements B_{ab}. This combination of matrices in is called the Kronecker product.

Definition 2.2.4. Given two representations of the same group, $D : G \to \mathrm{GL}(V)$ and $D' : G \to \mathrm{GL}(V')$, we can construct a new representation $D'' = D \otimes D'$, called the **tensor product**, which acts on $V \otimes V'$ in the natural way.

2.3 Reducible and irreducible representations

We have seen that a group can have many representations and that we can combine representations together to make new ones. It would be helpful to know if there are some basic representations that can act as a set of building blocks with which we can construct other representations. For many groups there is indeed such a set of building blocks. In this chapter we will describe a class of representations called *irreducible representations* (sometimes known as irreps), that for many groups can be used to construct all other representations using direct sums.

Definition 2.3.1. Consider a representation $D : G \to \mathrm{GL}(V)$. An **invariant subspace** W is a subspace of V such that $D(g)|w\rangle \in W$ for all $g \in G$ and $|w\rangle \in W$. The invariant subspace is said to be **non-trivial** if $W \neq \{0\}$ and $W \neq V$.

Any subspace has a projection operator P satisfying $P^2 = P$ where a vector $|w\rangle$ is in the subspace if and only if $P|w\rangle = |w\rangle$. The projection operator gets its name from the fact that P acting on an arbitrary vector $|v\rangle \in V$ produces a vector in the subspace $|W\rangle$, since

$$P(P|v\rangle) = P^2|v\rangle = P|v\rangle, \tag{2.35}$$

using the defining property $P^2 = P$. Thus, if a representation has an invariant subspace then we must have that for every $g \in G$.

$$PD(g)P|v\rangle = D(g)P|v\rangle. \tag{2.36}$$

Since this equation must hold for every $|v\rangle \in V$, this implies that a representation has an invariant subspace if and only if

$$PD(g)P = D(g)P, \quad \text{for all } g \in G. \tag{2.37}$$

Definition 2.3.2. A representation is said to be **irreducible** if it has no non-trivial invariant subspaces. A representation with at least one non-trivial invariant subspace is said to be **reducible**.

Suppose a representation is reducible, and thus has invariant subspace W. Denoting the dimension of the subspace $m = \dim(W)$ and the dimension of its orthogonal complement by $n = \dim(V) - \dim(W)$, if we employ a basis in which vectors in the invariant subspace take the form

$$|w\rangle = (\underbrace{0, 0, \cdots, 0}_{m \text{ zeroes}}, w_1, \cdots, w_n)^T, \tag{2.38}$$

then the projector onto the subspace is then the block matrix

$$P = \begin{pmatrix} 0 & 0 \\ 0 & \mathbb{1}_n \end{pmatrix}, \tag{2.39}$$

with $\mathbb{1}_n$ the $n \times n$ identity matrix. The condition in equation (2.37) implies that the representation preserves the invariant subspace if the representation matrices take the block form

$$D(g) = \begin{pmatrix} A(g) & 0 \\ C(g) & B(g) \end{pmatrix}, \tag{2.40}$$

where $A(g)$ is an $m \times m$ matrix, $B(g)$ is $n \times n$, and $C(g)$ is $n \times m$. Since $D(e)$ should be the identity matrix, we must have

$$A(e) = \mathbb{1}_m, \qquad B(e) = \mathbb{1}_n, \tag{2.41}$$

as well as $C(e) = 0$. The product of two matrices in this representation is

$$D(g_1)D(g_2) = \begin{pmatrix} A(g_1)A(g_2) & 0 \\ C(g_1)A(g_2) + B(g_1)C(g_2) & B(g_1)B(g_2) \end{pmatrix}. \tag{2.42}$$

In order that $D(g_1)D(g_2) = D(g_1 g_2)$, we see that we must have

$$A(g_1)A(g_2) = A(g_1 g_2), \qquad B(g_1)B(g_2) = B(g_1 g_2), \tag{2.43}$$

along with $C(g_1)A(g_2) + B(g_1)C(g_2) = C(g_1 g_2)$.

Equations (2.41) and (2.43) together imply that $A(g)$ and $B(g)$ both form representations of G. This is why we call D *reducible* — it contains at least two lower-dimensional representations (*at least* two because A and/or B might also be reducible).

If the off-diagonal block in equation (2.40) vanishes, $C(g) = 0$ for all $g \in G$, then the representation is block diagonal in this basis. In this case the representation D is a direct sum of the smaller representations A and B, $D = A \oplus B$. If A and B are reducible, it may be that they can also be written in block diagonal form, so that they are direct sums of yet smaller representations. If we can iterate this process to eventually write D as a direct sum of irreducible representations, then this gets a special name:

Definition 2.3.3. A finite-dimensional representation is **completely reducible** if it is equivalent to a direct sum of finitely many irreducible representations.

By this definition an irreducible representation is completely reducible, since it is isomorphic to a direct sum of one irreducible representation — itself. Completely reducible representations are also sometimes referred to as semisimple.

We soon show that every finite dimensional representation of a finite group is completely reducible. We will also show that every finite group has a finite number of irreducible representations. Thus, one can determine all of the irreducible representations of a finite group, and then construct any other representation as a direct sum of these irreducible representations.

2.4 Important properties of representations of finite groups

Theorem 2.4.1. *Every representation of a finite group is equivalent to a unitary representation.*

Proof. Consider an arbitrary representation of a finite group $D : G \to \mathrm{GL}(V)$. We wish to show that D is equivalent to a unitary representation, which we will do by explicitly constructing the necessary similarity transformation. The first step is to define the linear operator

$$A = \sum_{g \in G} D(g)^\dagger D(g). \tag{2.44}$$

This operator is Hermitian and positive definite. To see the latter, consider an arbitrary non-zero vector $|v\rangle \in V$. We have

$$\langle v|A|v\rangle = \sum_{g \in G} \langle v|D(g)^\dagger D(g)|v\rangle = \sum_{g \in G} ||D(g)|v\rangle||^2, \tag{2.45}$$

where $||\cdot||$ denotes the norm of the vector $|v\rangle$. Each term in the sum is non-negative, so we have $\langle v|A|v\rangle \geq 0$ with equality only if $D(g)|v\rangle = 0$ for all g. However, $D(e)$ is the identity matrix and by assumption $D(e)|v\rangle = |v\rangle \neq 0$ so there is at least one element in the sum that is non-zero and therefore the inequality is strict,

$$\langle v|A|v\rangle > 0, \qquad \forall |v\rangle \in V, \tag{2.46}$$

which is the statement that A is postive definite.

The fact that A is Hermitian means that it can be diagonalised, in other words there is an orthonormal basis $\{|e_i\rangle\}$ in which $A = \sum_i \lambda_i |e_i\rangle\langle e_i|$ where λ_i are the eigenvalues of A. The fact that A is positive definite means that every eigenvalue is positive, $\lambda_i > 0$. We can therefore define a square root S, i.e. an operator satisfying $S^2 = A$, which in the $\{|e_i\rangle\}$ basis takes the form $S = \sum_i \sqrt{\lambda_i}|e_i\rangle\langle e_i|$. Since the $\lambda_i > 0$, the eigenvalues of S, $\sqrt{\lambda_i}$, are real and non-zero so S is therefore Hermitian and invertible.

We now define a new representation D' by the similarity transformation

$$D'(g) = SD(g)S^{-1}. \tag{2.47}$$

By definition, D and D' are equivalent representations. Since S is Hermitian, we have that $D'(g)^\dagger = S^{-1}D(g)^\dagger S$, and therefore for any $g \in G$,

$$D'(g)^\dagger D'(g) = S^{-1}D(g)^\dagger S^2 D(g)S^{-1} = S^{-1}D(g)^\dagger A D(g)S^{-1}. \tag{2.48}$$

But

$$D(g)^\dagger A D(g) = \sum_{h\in G} D(g)^\dagger D(h)^\dagger D(h) D(g) = \sum_{h\in G} D(hg)^\dagger D(hg). \tag{2.49}$$

By the permutation law, in summing hg over all $h \in G$ for fixed g we are summing over all of the group elements exactly once, and therefore $\sum_{h\in G} D(hg)^\dagger D(hg) = \sum_{h\in G} D(h)^\dagger D(h)$, so that

$$D(g)^\dagger A D(g) = \sum_{h\in G} D(h)^\dagger D(h) = A = S^2. \tag{2.50}$$

Substituting this result into equation (2.48), we obtain

$$D'(g)^\dagger D'(g) = S^{-1} S^2 S^{-1} = \mathbb{1}, \tag{2.51}$$

and therefore $D'(g)$ is a unitary representation. □

Theorem 2.4.2. *Any finite-dimensional unitary representation is completely reducible.*

Proof. Consider a unitary representation $D : G \to \mathrm{GL}(V)$. If D is irreducible then it is completely reducible, by definition. Otherwise, if D is reducible then it has an invariant subspace which we denote W. Given an orthonormal basis $\{|e_i\rangle\}$ for W, we can define a projection operator

$$P = \sum_i |e_i\rangle\langle e_i|. \tag{2.52}$$

This satisfies $P^2 = P$ and $P^\dagger = P$. The condition that W is preserved by D is

$$PD(g)P = D(g)P, \tag{2.53}$$

for all $g \in G$. The expression in (2.53) projects out the smaller representation of G acting on the subspace W. Let's denote it $B(g) = D(g)P$. If D is unitary, then so is B; taking the Hermitian conjugate of equation (2.53) and using that P is Hermitian we find that

$$B(g)^\dagger B(g) = PD(g)^\dagger D(g)P = PD(g)^{-1} D(g)P = P\mathbb{1}P, \tag{2.54}$$

which is the projection of the identity onto W.

Taking the Hermitian conjugate of equation (2.53), we obtain

$$PD(g)^\dagger P = PD(g)^\dagger. \tag{2.55}$$

By assumption D is a unitary representation, and therefore $D(g)^\dagger = D(g)^{-1} = D(g^{-1})$. Since equation (2.55) holds for all $g \in G$ and the set of all g is the same as the set of all g^{-1}, we can therefore write equation (2.55) as

$$PD(g)P = PD(g). \tag{2.56}$$

Now consider the combination

$$\begin{aligned}(\mathbb{1} - P)D(g)(\mathbb{1} - P) &= D(g) - D(g)P - PD(g) + PD(g)P \\ &= D(g)(\mathbb{1} - P),\end{aligned} \tag{2.57}$$

where in the second line we have used equation (2.55) to cancel the last two terms. The combination $\mathbb{1} - P$ is also a projector,[4] projecting onto the orthogonal complement $W^\perp$ of the subspace W. This tells us that $D(g)$ preserves $W^\perp$ as well as W, and thus $D(g)$ is a direct sum of representations B acting on W and $A = D(\mathbb{1} - P)$ acting on $W^\perp$. Note that A is a unitary representation for the same reason that B is.

If the representations A and B are irreducible then we have decomposed D as a direct sum of irreducible representations and we are finished. Otherwise, we can repeat the proof on these sub-representations, decomposing A and B as direct sums of yet smaller representations, and so on. Since the representation is assumed to be finite-dimensional, this process must eventually stop, so we can eventually express D as a direct sum of irreducible representations. □

This proof is perhaps easier to understand in matrix notation. In the block notation used in equations (2.39) and (2.40), we wrote a reducible representation and the projector onto the preserved subspace as

$$D(g) = \begin{pmatrix} A(g) & 0 \\ C(g) & B(g) \end{pmatrix}, \qquad P = \begin{pmatrix} 0 & 0 \\ 0 & \mathbb{1} \end{pmatrix}. \tag{2.58}$$

The projector onto $W^\perp$ is then

$$\mathbb{1} - P = \begin{pmatrix} \mathbb{1} & 0 \\ 0 & 0 \end{pmatrix}. \tag{2.59}$$

If D satisfies equation (2.57), $(\mathbb{1} - P)D(g)(\mathbb{1} - P) = D(g)(\mathbb{1} - P)$ then multiplying out the matrices in equation (2.59) we see that this requires $C(g) = 0$ and hence $D(g)$ takes the block diagonal form

$$D(g) = \begin{pmatrix} A(g) & 0 \\ 0 & B(g) \end{pmatrix}. \tag{2.60}$$

In other words $D(g) = A(g) \oplus B(g)$.

To see the connection to unitarity, we take the inverse and the Hermitian conjugate of $D(g)$ in equation (2.58),

$$D(g)^{-1} = \begin{pmatrix} A(g)^{-1} & 0 \\ -B(g)^{-1}C(g)A(g)^{-1} & B(g)^{-1} \end{pmatrix}, \qquad D(g)^\dagger = \begin{pmatrix} A(g)^\dagger & C(g)^\dagger \\ 0 & B(g)^\dagger \end{pmatrix}. \tag{2.61}$$

If D is a unitary representation then we must have $D(g)^{-1} = D(g)^\dagger$. Comparing the matrices in equation (2.61) we see that this occurs if $A(g)^\dagger = A(g)^{-1}$, $B(g)^\dagger = B(g)^{-1}$, and $C(g) = 0$. Thus we see again that the block diagonal form is required by unitarity, and also that the diagonal blocks must themselves be unitary.

Taken together, theorem 2.4.1 and theorem 2.4.2 imply the important result:

Corollary 2.4.1. *Every finite-dimensional representation of a finite group is completely reducible.*

[4] Since $(\mathbb{1} - P)^2 = \mathbb{1} - 2P + P^2 = \mathbb{1} - P$ using that $P^2 = P$.

In other words, up to a similarity transformation, every finite-dimensional representation of a finite group may be expressed as the direct sum of irreducible representations.

Theorem (2.4.1) and corollary (2.4.1) only apply to finite groups because the sum over G in equation (2.44) is not guaranteed to converge if G contains infinitely many elements. Here is an explicit example of a representation of an infinite group that is not completely reducible:

Example 2.4.1. The set of real numbers $\mathbb{R}$ forms an infinite group, with the "group multiplication" given by addition, $x \circ y = x + y$ for two real numbers x and y. The identity element is the number zero. This group has a two-dimensional representation

$$D(x) = \begin{pmatrix} 1 & 0 \\ x & 1 \end{pmatrix}. \tag{2.62}$$

Plainly $D(0) = \mathbb{1}$ and it is straightforward to verify that $D(x)D(y) = D(x+y)$, so this is indeed a representation. It is also reducible, since it preserves the one-dimensional subspace consisting of vectors with their first component vanishing. Suppose we try to construct a similarity transformation that makes $D(x)$ diagonal, thus expressing this representation as the direct sum of two one-dimensional representations. Defining the matrix

$$S = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \tag{2.63}$$

we have that

$$S^{-1}D(x)S = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}\begin{pmatrix} 1 & 0 \\ x & 1 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \mathbb{1} - \frac{x}{ad-bc}\begin{pmatrix} ab & b^2 \\ -a^2 & -ab \end{pmatrix}, \tag{2.64}$$

and we see that there is no possible choice of the elements of S that makes $S^{-1}D(x)S$ diagonal.

2.5 Schur's lemma

In this section we will prove Schur's lemma, a two-part result that is useful for proving many other things in representation theory. One the main uses of Schur's lemma is to develop conditions for determining which representations are irreducible.

Lemma 2.5.1 (Schur's lemma)**.** *For any group G:*

(i) Suppose that $D_1 : G \to \mathrm{GL}(V_1)$ and $D_2 : G \to \mathrm{GL}(V_2)$ are two irreducible representations. Suppose further that there is a map $A : V_2 \to V_1$ satisfying $D_1(g)A = AD_2(g)$ for all $g \in G$. Then either $A = 0$ or A is invertible, in which case $D_1(g) = AD_2(g)A^{-1}$ and the representations are equivalent.

(ii) Suppose that $D : G \to \mathrm{GL}(V)$ is an irreducible representation and that $A : V \to V$ is an invertible map from V to itself, such that $D(g)A = AD(g)$ for all $g \in G$. Then $A = \lambda\mathbb{1}$ for some $\lambda \in \mathbb{C}$.

Proof of part (i) of Schur's lemma. Let $W \subset V_2$ be the subspace of V_2 defined by $|w\rangle \in W$

if and only if $A|w\rangle = 0$. The subspace W is an invariant subspace of the representation D_2 since

$$AD_2(g)|w\rangle = D_1(g)A|w\rangle = 0, \tag{2.65}$$

showing that $AD_2(g)|w\rangle \in W$ for all $g \in G$ and $|w\rangle \in W$. By assumption D_2 is irreducible, so has no nontrivial invariant subspaces. Thus we have two options, either $W = V_2$ or $W = \{0\}$.

If $W = V_2$ then A annihilates every vector in its domain, $A|v\rangle = 0$ for all $|v\rangle$, and hence $A = 0$.

On the other hand if $A \neq 0$ then we have found that must have $W = \{0\}$. In other words, A does not annihilate any non-zero vector. This implies that A is injective (maps every element of V_2 to a different element of V_1), since if we were to have that $A|v\rangle = A|v'\rangle$ for two vectors $|v\rangle, |v'\rangle \in V_2$ then we could take the difference of both sides and get that $A(|v\rangle - |v'\rangle) = 0$, which is only possible if $|v\rangle - |v'\rangle = 0$. We will show that A is also surjective (every element of V_1 may be written as $A|v\rangle$ for some $|v\rangle \in V_2$).

To see this, consider the image of A in V_1, which we denote U. By definition, U is the set of all vectors in V_1 that can be written as $A|v\rangle$ for some $|v\rangle \in V_1$. But U is an invariant subspace of D_1, since

$$D_1(g)A|v\rangle = AD_2(g)|v\rangle = A|v'\rangle, \tag{2.66}$$

where $|v'\rangle = D_2(g)|v\rangle$. Since D_2 is assumed to be irreducible it can have no nontrivial subspaces, so, either $U = \{0\}$ or $U = V_1$. If $U = \{0\}$ then this means $A|v\rangle = 0$ for all $|v\rangle \in V_2$, i.e. $A = 0$, which was the case above. For $A \neq 0$ the only option is then $U = V_1$. Then A is surjective.

We have shown that if $A \neq 0$ then A is both injective and surjective, and is thus bijective. A is therefore invertible, meaning that D_1 and D_2 are equivalent representations of G. □

Proof of part (ii) of Schur's lemma. Since A is assumed to be invertible its eigenvalues are non-zero. Let λ be one of those eigenvalues and U the corresponding eigenspace. Thus U is the space of the vectors $|u\rangle$ satisfying

$$A|u\rangle = \lambda|u\rangle \qquad \Rightarrow \qquad (A - \lambda\mathbb{1})|u\rangle = 0. \tag{2.67}$$

Since $D(g)A = AD(g)$ we also have that

$$D(g)(A - \lambda\mathbb{1}) = (A - \lambda\mathbb{1})D(g). \tag{2.68}$$

By the same logic as in the proof of part (i), this implies that U is an invariant subspace of D; if $|u\rangle \in U$ then

$$(A - \lambda\mathbb{1})D(g)|u\rangle = D(g)(A - \lambda\mathbb{1})|u\rangle = 0 \tag{2.69}$$

and so $D(g)|u\rangle \in U$.

Since D is assumed to be irreducible it can have no non-trivial invariant subspaces, so either $U = \{0\}$ or $U = V$. But it cannot be that $U = \{0\}$ since every eigenvalue corresponds to at least one non-zero eigenvector, so the only option is that $U = V$. Thus $(A - \lambda\mathbb{1})|v\rangle = 0$ for all $|v\rangle \in V$, and thus $(A - \lambda\mathbb{1}) = 0$ and so $A = \lambda\mathbb{1}$. □

Here is something that follows almost immediately from Schur's lemma:

Corollary 2.5.1. *All irreducible representations of a finite abelian group are one-dimensional.*

Proof. Let D be an n-dimensional complex irreducible representation of an abelian group G. Pick any group element $g \in G$. Since the group is abelian $D(g)$ commutes with all of the operators in this representation, $D(g)D(h) = D(gh) = D(hg) = D(h)D(g)$ for all $h \in G$, so part (ii) of Schur's lemma implies that $D(g)$ is proportional to the identity. Thus the representation takes the form

$$D(g) = \lambda(g)\mathbb{1}_n, \qquad \lambda(g) \in \mathbb{C}. \tag{2.70}$$

This implies that D is the direct sum of n copies of the one-dimensional representation in which each group element is represented by $\lambda(g)$. But D was assumed to be irreducible, so the only possibility is that $n = 1$. □

2.6 The orthogonality theorem

Theorem 2.6.1 (Orthogonality theorem)**.** *Let $C : G \to \mathrm{GL}(U)$ and $D : G \to \mathrm{GL}(V)$ be inequivalent irreducible unitary representations of a finite group G. Denote the matrix elements in these representations as $[C(g)]_{ij}$ and $[D(g)]_{ab}$. Then*

$$\sum_{g\in G} [C(g)]_{ij}[D(g)^\dagger]_{ab} = 0. \tag{2.71}$$

Alternatively, if we use the same representation in this sum, we find

$$\sum_{g\in G} [D(g)]_{ab}[D(g)^\dagger]_{cd} = \frac{|G|}{n}\delta_{ad}\delta_{bc}, \tag{2.72}$$

where as usual $|G|$ is the order of G, while n is the dimension of the representation D.

Proof. Consider an arbitrary map $X : V \to U$ and construct another map $A : V \to U$ through

$$A = \sum_{g\in G} C(g)XD(g)^\dagger = \sum_{g\in G} C(g)XD(g^{-1}), \tag{2.73}$$

where the second equality follows from the fact that D is unitary. Then we have that

$$\begin{aligned} C(h)A &= \sum_{g\in G} C(h)C(g)XD(g^{-1}) \\ &= \sum_{g\in G} C(h)C(g)X(D(h)D(g))^{-1}D(h) \\ &= \sum_{g\in G} C(hg)XD(hg)^{-1}D(h) \\ &= \sum_{g\in G} C(g)XD(g)^{-1}D(h) \\ &= AD(h), \end{aligned} \tag{2.74}$$

where the final equality follows from the permutation law 1.2.6, which tells us that summing over all hg with fixed h is the same as summing over all g.

If C and D are inequivalent then Schur's lemma 2.5.1 implies that $A = 0$,

$$\sum_{g\in G} C(g)XD(g)^\dagger = 0, \tag{2.75}$$

for any X. Now let's adopt a basis $\{|e_i\rangle\}$ for U and $\{|f_a\rangle\}$ for V, and take $X = |e_j\rangle\langle f_a|$ for some fixed j and a. The matrix elements of the left-hand side of equation (2.75) in this basis are

$$\left[\sum_{g\in G} C(g)XD(g)^\dagger\right]_{ib} = \sum_{g\in G}\langle e_i|C(g)|e_j\rangle\langle f_a|D(g)^\dagger|f_b\rangle = \sum_{g\in G}[C(g)]_{ij}[D(g)^\dagger]_{ab}. \tag{2.76}$$

Equation (2.75) implies that these matrix elements vanish, establishing the first half of the orthogonality theorem.

Suppose instead that $C = D$. Part (ii) of Schur's lemma then tells us that A is proportional to the identity,

$$\sum_{g\in G} D(g)XD(g)^\dagger = \lambda_X \mathbb{1}_n, \tag{2.77}$$

where $\lambda_X \in \mathbb{C}$ depends on what we chose for the map X. To determine λ_X we take the trace of both sides. The trace of the right-hand side is $n\lambda_X$, while the trace of the left-hand side is

$$\sum_{g\in G} \operatorname{tr}\left(D(g)XD(g)^\dagger\right) = \sum_{g\in G} \operatorname{tr}\left(D(g)^\dagger D(g)X\right) = |G| \operatorname{tr} X \tag{2.78}$$

where we have used clyclicity of the trace and unitrarity of the representation. This fixs λ_X, so we have

$$\sum_{g\in G} D(g)XD(g)^\dagger = \frac{|G|}{n}\mathbb{1}_n \operatorname{tr} X. \tag{2.79}$$

Finally we adopt an orthonormal basis $\{|f_a\rangle\}$ for V and choose $X = |f_b\rangle\langle f_c|$, then $\operatorname{tr} X = \delta_{bc}$ and we find

$$\sum_{g\in G}[D(g)]_{ab}[D(g)^\dagger]_{cd} = \frac{|G|}{n}\delta_{ad}\delta_{bc}, \tag{2.80}$$

establishing the second half of the orthogonality theorem. □

2.7 Characters

2.7.1 Definition of characters

Definition 2.7.1. Given a representation D of a group G, the **character** $\chi_D(g)$ of a group element $g \in G$ is the trace of $D(g)$,

$$\chi_D(g) = \operatorname{tr} D(g). \tag{2.81}$$

The cyclic property of the trace implies that characters are unchanged under similarity transformations, and therefore the character of a group element is the same in equivalent representations. If D' is equivalent to D, with $D'(g) = S^{-1}D(g)S$, then

$$\chi_{D'}(g) = \operatorname{tr}\left(S^{-1}D(g)S\right) = \operatorname{tr}\left(D(g)SS^{-1}\right) = \operatorname{tr} D(g) = \chi_D(g). \tag{2.82}$$

The same cyclic property implies that characters are constant on conjugacy classes. Recall that if two group elements g and g' are in the same conjugacy class, this means that there is another group element h such that $g' = hgh^{-1}$. Then

$$\chi_D(g') = \operatorname{tr} D(g') = \operatorname{tr}\left(D(h)D(g)D(h)^{-1}\right) = \operatorname{tr} D(g) = \chi_D(g). \tag{2.83}$$

Theorem 2.7.1. *Given two representations D and D' of a group, the characters of the direct sum $D \oplus D'$ and the tensor product $D \otimes D'$ are*

$$\chi_{D\oplus D'}(g) = \chi_D(g) + \chi_{D'}(g), \qquad \chi_{D\otimes D'}(g) = \chi_D(g)\chi_{D'}(g). \tag{2.84}$$

Proof. Write the matrix representation of the direct sum in block diagonal form as

$$(D \oplus D')(g) = \begin{pmatrix} D(g) & 0 \\ 0 & D'(g) \end{pmatrix}. \tag{2.85}$$

Taking the trace of this expression immediately gives the stated result for the direct sum. For the tensor product, if we employ bases $\{|e_i\rangle\}$ and $\{|f_a\rangle\}$ for the vector spaces on which D and D' act, then the matrix elements of $(D \otimes D')(g)$ are

$$\begin{aligned} [(D \otimes D')(g)]_{ia,jb} &= \left(\langle e_i| \otimes \langle f_a|\right)\left(D(g) \otimes D'(g)\right)\left(|e_j\rangle \otimes |f_b\rangle\right) \\ &= \langle e_i|D(g)|e_j\rangle\langle f_a|D'(g)|f_b\rangle \\ &= [D(g)]_{ij}[D(g)]_{ab}. \end{aligned} \tag{2.86}$$

Taking the trace of both sides reproduces the stated result for the tensor product. □

2.7.2 Character orthogonality

Taking the traces separately over the ij and ab indices appearing in the orthogonality theorem 2.6.1 replaces the representation matrices with their characters. We thus arrive at the following result.

Theorem 2.7.2 (Character orthogonality I)**.** *Consider two inequivalent irreducible representations $C : G \to \mathrm{GL}(U)$ and $D : G \to \mathrm{GL}(V)$ of the same finite group G. The characters of these representations satisfy*

$$\sum_{g\in G} \chi_C(g)\chi_D(g)^* = |G|\delta_{CD}, \tag{2.87}$$

where $\delta_{CD} = 0$ if C and D are inequivalent, and $\delta_{CD} = 1$ if they are equivalent.

Proof. Take the trace of equations (2.71) and (2.72). □

Since the characters are constant on the conjugacy classes, we can rewrite equation (2.87) as a sum over conjugacy classes. Labelling each conjugacy class by an index α and denoting for example $\chi_D(g) = \chi_{D,\alpha}$ if the element g is in the αth conjugacy class, equation (2.87) becomes

$$\sum_{p=1}^{c} m_\alpha \chi_{C,\alpha}\chi^*_{D,\alpha} = |G|\delta_{CD}, \tag{2.88}$$

where m_p is the number of group elements in the pth conjugacy class and c is the total number of conjugacy classes.

This character orthogonality theorem implies a bound on the numer of inequivalent irreducible representations:

Theorem 2.7.3. *The number ρ of inequivalent irreducible representations of a finite group is bounded from above by the number c of conjugacy classes of that group, $\rho \leq c$.*

Proof. For each representation D of the group, define the "normalised character" of the αth conjugacy class as $\bar{\chi}_{D,\alpha} = \sqrt{m_\alpha/|G|}\,\chi_{D,\alpha}$. Then equation (2.88) becomes

$$\sum_{i=1}^{c} \bar{\chi}_{C,p}\bar{\chi}^*_{D,p} = \delta_{CD}. \tag{2.89}$$

Thinking of the $\bar{\chi}_{D,\alpha}$ as the components of a c-component vector

$$\bar{\chi}_D = \begin{pmatrix} \bar{\chi}_{D,1} \\ \bar{\chi}_{D,2} \\ \vdots \\ \bar{\chi}_{D,c} \end{pmatrix}, \tag{2.90}$$

equation (2.89) becomes

$$\langle \bar{\chi}_D, \bar{\chi}_C \rangle = \delta_{CD}, \tag{2.91}$$

where $\langle \cdot, \cdot \rangle$ denotes the usual inner product on a complex vector space. There are ρ column vectors of the form (2.90), one for each inequivalent irreducible representation, and these vectors are mutually orthogonal by equation (2.90). But we can have at most c mutually orthogonal vectors in a c-dimensional vector space, and therefore $\rho \leq c$. □

This establishes one result promised earlier, that the number of inequivalent irreducible representations of a finite group is finite. But we can actually go further and show that the bound on ρ is always saturated so that once we know the conjugacy classes of a group we know exactly how many irreducible representations it has. To do so we need another orthogonality property of characters:

Theorem 2.7.4 (Character orthogonality II)**.** *Let r label the inequivalent irreducible representations D_r of a finite group G and α label the conjugacy classes of G. The characters of these representations satisfy*

$$\sum_{r=1}^{\rho} \chi_{r,\alpha}\chi^*_{r,\beta} = \frac{|G|}{m_\alpha}\delta_{\alpha\beta}, \tag{2.92}$$

where ρ is the number of inequivalent irreducible representations and m_α is the number of elements in the αth conjugacy class.

The proof of this theorem is a little long. We give it in section 2.10.

Using this second character orthogonality theorem, we can now prove a key result.

Theorem 2.7.5. *The number ρ of inequivalent irreducible representations of a finite group is equal to the number c of conjugacy classes of that group, $\rho = c$.*

Proof. We prove this in the same way as we proved theorem 2.7.3. For each α we think of the index r in equation (2.92) as labelling components of a vector[5]

$$\chi_\alpha = \begin{pmatrix} \chi_{1,\alpha} \\ \chi_{2,\alpha} \\ \vdots \\ \chi_{\rho,\alpha} \end{pmatrix}. \tag{2.93}$$

Equation (2.92) then becomes

$$\langle \chi_\beta, \chi_\alpha \rangle = \frac{|G|}{m_\alpha} \delta_{\alpha\beta}, \tag{2.94}$$

where $\langle \cdot, \cdot \rangle$ is the usual inner product on a complex vector space. We thus have c mutually orthogonal r-component vectors χ_α. This is only possible if $\rho \geq c$. But from theorem 2.7.3 we have that $\rho \leq c$. Combinining these inequalities gives $\rho = c$. □

Theorem 2.7.6. *Let n_r be the dimension of the irreducible representation D_r. Summing over all inequivalent irreducible representations, we have*

$$\sum_{r=1}^{c} n_r^2 = |G|. \tag{2.95}$$

Proof. This follows from the second character orthogonality theorem in equation (2.92). Choosing both α and β to correspond to the conjugacy class of the identity element. The identity element is always represented by the identity operator, and so we have

$$\chi_{r,\alpha} = \chi_{r,\beta} = \operatorname{tr} \mathbb{1}_{n_r} = n_r. \tag{2.96}$$

Recalling that the identity element always sits in a conjugacy class all by itself, we have that $m_\alpha = 1$. Substituting these results into equation (2.92) gives the result stated in the theorem. □

2.8 Reducibility of representations

Let G be a finite group and let $D_r : G \to \mathrm{GL}(V)$ be the inequivalent irreducible representations, where $r = 1, 2, \cdots, c$ with c the number of conjugacy classes of G. Let$D : G \to \mathrm{GL}(V)$ be any representation. From theorem 2.4.2 we know that D must be completely reducible, so it mus be possible to write D as a direct sum

$$D = D_1 \oplus \cdots \oplus D_1 \oplus D_2 \oplus \cdots \oplus D_2 \oplus \cdots \oplus D_c \oplus \cdots \oplus D_c. \tag{2.97}$$

Let's simplify this notation and write

$$D = \bigoplus_{r=1}^{c} k_r D_r, \tag{2.98}$$

[5]Note that the components of this vector are different to the components of the vector appearing in the proof of theorem 2.7.3.

where for each r the non-negative integer k_r is the number of times D_r appears in the direct sum. An irreducible representation has one $k_r = 1$ and all the rest vanishing.

We will use character orthogonality to find a criterion for irreducibility and to find a formula for k_r. Taking the trace of equation (2.98) acting on a group element $g \in G$, we find that

$$\chi_D(g) = \sum_{r=1}^{c} k_r \chi_r(g), \tag{2.99}$$

where χ_r are the characters of the irreducible representations. Our task in this section is to prove the following two theorems:

Theorem 2.8.1 (Irreducibility criterion)**.** *A representation $D : G \to \mathrm{GL}(V)$ of a finite group G is irreducible if and only if its characters satisfy*

$$\sum_{\alpha=1}^{c} m_\alpha |\chi_{D,\alpha}|^2 = |G|, \tag{2.100}$$

where $\chi_{D,\alpha}$ is the character of the αth conjugacy class and m_α is the number of group elements in this conjugacy class.

Theorem 2.8.2 (Decomposition theorem)**.** *The coefficients k_r appearing in the decomposition* (2.98) *are given by*

$$k_r = \frac{1}{|G|} \sum_{\alpha=1}^{c} m_\alpha \chi_{D,\alpha} \chi^*_{r,\alpha}. \tag{2.101}$$

If you are given any representation of a group, you can use theorem 2.8.1 to determine whether it is irreducible or not. For example, one use for this is when looking for all of the inequivalent representations of a finite group, we can use theorem 2.8.1 to determine whether a candidate representation is indeed irreducible. Once all of the irreducible representations are known we can then use theorem 2.8.2 to understand how any other representation decomposes into a direct sum of them. We will see some examples in the next section. We now prove these theorems.

Proof of theorem 2.8.1. Using equation (2.99) we can expand the sum appearing the theorem as

$$\sum_{r=1}^{c} m_\alpha |\chi_{D,\alpha}|^2 = \sum_{g\in G} |\chi_D(g)|^2 = \sum_{g\in G} \sum_{r=1}^{c} \sum_{r'=1}^{c} k_r k'_r \chi_r(g) \chi_{r'}(g)^*. \tag{2.102}$$

We then use the first character orthgonality theorem (2.87) to perform the sum over g

$$\sum_{r=1}^{c} m_\alpha |\chi_{D,\alpha}|^2 = \sum_{r=1}^{c} \sum_{r'=1}^{c} k_r k_{r'} |G| \delta_{rr'} = |G| \sum_{r=1}^{c} k_r^2. \tag{2.103}$$

An irreducible representation has exactly one $k_r = 1$ and the rest zero, in which case the right-hand side equals $|G|$. A reducible representation will have at least one $k_r > 1$ and/or more than one k_r non-zero, so the right-hand side of equation (2.103) will be larger than $|G|$ for a reducible representation. □

Proof of theorem 2.8.2. If g in equation (2.99) is in the αth conjugacy class, the equation becomes

$$\chi_{D,\alpha}(g) = \sum_{r'=1}^{c} k_{r'} \chi_{r',\alpha}. \tag{2.104}$$

Multiplying by $m_\alpha \chi^*_{r,\alpha}$ and summing over the conjugacy classes, we find

$$\sum_{\alpha=1}^{c} m_\alpha \chi_{D,\alpha} \chi^*_{r,\alpha} = \sum_{r'=1}^{c} k_{r'} \sum_{\alpha=1}^{c} m_\alpha \chi_{r',\alpha} \chi^*_{r,\alpha}. \tag{2.105}$$

Performing the sum over α using the first character orthgonality theorem (2.87), we obtain

$$\sum_{\alpha=1}^{c} m_\alpha \chi_{D,\alpha} \chi^*_{r,\alpha} = \sum_{r'=1}^{c} k_{r'} |G| \delta_{rr'} = |G| k_r, \tag{2.106}$$

which is the result in the theorem. □

Example 2.8.1. Recall the regular representation from definition 2.1.4, which exists for any group G and is $|G|$-dimensional. The characters for the regular representation are straightforward to compute. As always, the character of the conjugacy class containing the identity element is equal to the dimension of the representation, $\chi_{D_R}(e) = |G|$. On the other hand, for any $g_i \in G$ other than the identity, all of the diagonal elements of $D_R(g_i)$ must vanish and therefore the character of g_i must vanish. To see this, note that the diagonal elements of $D_R(g_i)$ are

$$[D_R(g_i)]_{jj} = \langle g_j | D_R(g_i) | g_j \rangle = \langle g_j | g_i g_j \rangle, \tag{2.107}$$

where we use the basis from representation 2.1.4, and the index j runs over the group elements. The right-hand side of equation (2.107) vanishes unless $g_j = g_i g_j$, which is only possible if $g_i = e$.

To summarise, the characters of the regular representation are $\chi_{D_R} = |G|$ and $\chi_{D_R}(g) = 0$ for any $g \neq e$. The left-hand side of the irreducibility criterion is (2.100)

$$\sum_{\alpha=1} m_\alpha |\chi_{D,\alpha}|^2 = |G|^2. \tag{2.108}$$

This is not equal to $|G|$ except for groups of order one, so we have learned that the regular representation is reducible for any group containing more than one element. Substituting the characters into the decomposition theorem 2.8.2, we find that

$$k_r = \chi_r(e)^* = n_r, \tag{2.109}$$

In other words the number of times an irreducible representation appears in the decomposition of the regular representation is equal to the dimension of the irreducible representation.

2.9 Examples

2.9.1 $\mathbb{Z}_3$

Let's consider the group of three elements $G = \{e, a, b\}$ from example 1.1.1, which has multiplication table:

	e	a	b
e	e	a	b
a	a	b	e
b	b	e	a

This group is isomorphic to the cyclic group of order three $\mathbb{Z}_3$. Since the group is abelian:

- each conjugacy class consists of a single element, they are $\{e\}$, $\{a\}$, and $\{b\}$;
- the number of irreducible representations is three, the order of group, since there are as many irreducible representations as conjugacy classes; and
- all of the irreducible representations are one-dimensional, by corollary 2.5.1.

One of the irreducible representations is the trivial representation, which we label D_1,

$$D_1(e) = D_1(a) = D_1(b) = 1. \tag{2.110}$$

The other two irreducible representations can be guessed from the fact G is isomorphic to $\mathbb{Z}_3 = \{1, e^{2\pi i/3}, e^{4\pi i/3}\}$. Under this isomorphism $e \mapsto 1$, and we have a choice whether $a \mapsto e^{2\pi i/3}$ or $e^{4\pi i/3}$, giving two more irreducible representations $D_{2,3}$,

$$\begin{aligned} &D_2(e) = 1, && D_2(a) = e^{2\pi i/3}, && D_2(b) = e^{4\pi i/3}, \\ &D_3(e) = 1, && D_3(a) = e^{4\pi i/3}, && D_3(b) = e^{2\pi i/3}. \end{aligned} \tag{2.111}$$

These representations are unitary since each $D_r(g)$ is the inverse of its complex conjugate.

Let's put all of these representations in a table,

	e	a	b
D_1	1	1	1
D_2	1	$e^{2\pi i/3}$	$e^{4\pi i/3}$
D_3	1	$e^{4\pi i/3}$	$e^{2\pi i/3}$

Since the representations are one-dimensional, the characters are equal to the representation matrices and so the elements of this table are also the characters of the irreducible representations. According to the first character orthgonality theorem 2.7.2, the rows of this table must be orthogonal to each other and indeed they are, for example

$$\chi_1(e)^*\chi_2(e) + \chi_1(a)^*\chi_2(a) + \chi_1(b)^*\chi_2(b) = 1 + e^{2\pi i/3} + e^{4\pi i/3} = 0. \tag{2.112}$$

Similarly, the second character orthogonality theorem 2.7.4 tells us that the columns of this table must be one-dimensional and indeed they are.

Now let's consider the regular representation of this group, which we constructed in example 2.1.2,

$$D_R(e) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad D_R(a) = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \qquad D_R(b) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}. \tag{2.113}$$

Taking the traces of these matrices, we find the corresponding characters,

$$\chi_{D_R}(e) = 3, \qquad \chi_{D_R}(a) = 0, \qquad \chi_{D_R}(b) = 0. \tag{2.114}$$

This is what we expected from the general analysis in example 2.8.1. Explicitly evaluating the left-hand side of the irreducibility criterion 2.8.1, we find

$$\sum_{\alpha=1}^{3} m_\alpha |\chi_{D_R,\alpha}|^2 = |\chi_{D_R}(e)|^2 + |\chi_{D_R}(a)|^2 + |\chi_{D_R}(b)|^2 = 9 \neq 3 = |G|, \tag{2.115}$$

so theorem 2.8.1 confirms that the regular representation is not irreducible.

Since the irreducible representations are all one-dimensional, from example 2.8.1 we expect them to each appear once in the direct sum decomposition of the regular representation. We can check this explictly using the characters we have found. From theorem 2.8.2:

$$\begin{aligned} k_1 &= \frac{1}{3}(\chi_{D_R}(e) + \chi_{D_R}(a) + \chi_{D_R}(b)) = 1, \\ k_2 &= \frac{1}{3}(\chi_{D_R}(e) + e^{-2\pi i/3}\chi_{D_R}(a) + e^{-4\pi i/3}\chi_{D_R}(b)) = 1, \\ k_3 &= \frac{1}{3}(\chi_{D_R}(e) + e^{-4\pi i/3}\chi_{D_R}(a) + e^{-2\pi i/3}\chi_{D_R}(b)) = 1. \end{aligned} \tag{2.116}$$

Thus, the regular representation must be equivalent to $D' = D_1 \oplus D_2 \oplus D_3$. Explicitly, this representation is

$$D'(e) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad D'(a) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{2\pi i/3} & 0 \\ 0 & 0 & e^{4\pi i/3} \end{pmatrix}, \qquad D'(b) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{4\pi i/3} & 0 \\ 0 & 0 & e^{2\pi i/3} \end{pmatrix}. \tag{2.117}$$

The characters of this representation match the characters of the regular representation, which is a necessary condition for equivalence. We can prove their equivalence by constructing a similarity transformation relating them. It is straightforward to check that $D'(g) = S^{-1}D(g)S$ where

$$S = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 1 & 1 \\ 1 & e^{4\pi i/3} & e^{2\pi i/3} \\ 1 & e^{2\pi i/3} & e^{4\pi i/3} \end{pmatrix}, \tag{2.118}$$

and thus D' is indeed equivalent to the regular representation (for checking this, it is helpful to know that we have constructed S to be unitary, so $S^{-1} = S^\dagger$).

2.9.2 D_3

For a simple non-abelian example, let's consider the six-dimensional group from example 1.1.2, which is isomorphic to D_3, the group of symmetries of the equilateral triangle. Its multiplication table is

	e	a	b	c	d	f
e	e	a	b	c	d	f
a	a	b	e	f	c	d
b	b	e	a	d	f	c
c	c	d	f	e	a	b
d	d	f	c	b	e	a
f	f	c	d	a	b	e

We saw in example 1.7.1 that this group has three conjugacy classes, $\{e\}$, $\{a, b\}$, and $\{c, d, f\}$. It therefore has three inequivalent irreducible representations. One of these irreducible representations is, as always, the trivial one,

$$D_1(g) = 1, \tag{2.119}$$

for all g. We wish to determine the other two irreducible representations.

It turns out that we can determine the dimensions of these representations using theorem (2.7.6), which tells us that the the sum of the squares of the dimensions of the irreducible representations must add up to six. Thus the dimensions n_2 and n_3 of the as yet to be determined irreducible representations satisfy

$$n_2^2 + n_3^2 = 5. \tag{2.120}$$

Since $n_{2,3}$ must be positive integers, the only possibility is that one of them is one and the other is two. Let's label the other one-dimensional irreducible representation as D_1' and the two-dimensional one as D_2.

Since D_1' is one-dimensional, its representation "matrices" are equal to its characters, so $D_1'(g)$ must be constant on conjugacy classes. It's straightforward to check that the following is indeed a representation,

$$D_1'(e) = D_1'(a) = D_1'(b) = 1, \qquad D_1'(c) = D_1'(d) = D_1'(f) = -1. \tag{2.121}$$

This acts as the trivial representation on the subgroup $H = \{e, a, b\}$. It preserves the group multiplication since multiplying any element of H by an element of its complement $X = \{c, d, f\}$ yields an element of X, while multiplying two elements of X together yields an element of H.

Now we turn to the two-dimensional representation D_2. We can gain some information about this representation from character orthogonality. Let's write out what we know so far of the character table of D_3:

	$\{e\}$	$\{a, b\}$	$\{c, d, f\}$
D_1	1	1	1
D_1'	1	1	-1
D_2	2	$\chi_{3,2}$	$\chi_{3,3}$

The 2 in the bottom-left corner comes from the fact that we know that D_3 is two-dimensional, and therefore $\chi_{3,1} = \operatorname{tr} D_2(e) = 2$. The second character orthgonality theorem 2.7.4 tells us that different columns of this table must be orthogonal, so we can immediately read off that $\chi_{3,2} = -1$ and $\chi_{3,3} = 0$. The complete character table is thus:

	$\{e\}$	$\{\begin{smallmatrix}a\\b\end{smallmatrix}\}$	$\{\begin{smallmatrix}c\\d\\f\end{smallmatrix}\}$
D_1	1	1	1
D_1'	1	1	-1
D_2	2	-1	0

It is straightforward to check that the first orthogonality theorem 2.7.2 — which says that the rows should be orthogonal when weighted by the number of elements in the conjugacy class — is satisfied by this table.

Character orthogonality has fixed the traces of the representation matrices, but not the representation matrices themselves. However, one can check by tedious calculation that the following is a representation of D_3,

$$D_2(e) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad D_2(a) = \begin{pmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}, \qquad D_2(b) = \begin{pmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix},$$

$$D_2(c) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad D_2(d) = \begin{pmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}, \qquad D_2(f) = \begin{pmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}. \tag{2.122}$$

The characters of this representation match those written in the character table, so we can take this representation to be D_2. In particular, we know that this representation is irreducible because its characters satisfy the criterion in theorem 2.8.1,

$$\sum_{\alpha=1}^{c} m_\alpha |\chi_{D_2,\alpha}|^2 = 1 \times 2^2 + 2 \times (-1)^2 + 3 \times 0^2 = 6 = |G|. \tag{2.123}$$

We can construct a four-dimensional representation of D_3 by taking the tensor product of D_2 with itself, $D_4 = D_2 \otimes D_2$. The characters of a tensor product are the product of the characters of its factors, $\chi_{D_4}(g) = \chi_{D_2}(g)^2$, and thus the characters of D_4 are:

	$\{e\}$	$\{\begin{smallmatrix}a\\b\end{smallmatrix}\}$	$\{\begin{smallmatrix}c\\d\\f\end{smallmatrix}\}$
D_4	4	1	0

We already know that D_4 cannot be irreducible since we have already determined that there are no four-dimensional irreducible representations. We can confirm this with theorem 2.8.1,

$$\sum_{\alpha=1}^{3} m_\alpha |\chi_{D_4,\alpha}|^2 = 18 \neq |\mathrm{D}_4|. \tag{2.124}$$

Thus it must be possible to reexpress D_4 as the direct sum of the irreducible representations. Using theorem 2.8.2 we can determine how many times each irreducible representation appears in this direct sum. Using the character tables written above one finds $k_{D_1} = k_{D_1'} = k_{D_2} = 1$, so each irreducible representation appears exactly once,

$$D_2 \otimes D_2 = D_1 \oplus D_1' \oplus D_2, \tag{2.125}$$

where this equality holds up to equivalence. This is an example of something that one does all the time when working with representation theory in physics: taking tensor products of representations and rewriting them as direct sums of irreducible representations. Tensor products occur when physical quantities have multiple indices that transform other

the symmetry group. The physical significance of the irreducible representations should become clear when we study applications in the next chapter.

The decomposition of D_4 can be made explicit if we write out the representation matrices using equation (2.34). We won't write out all six 4×4 matrices, but for example

$$D_4(a) = \frac{1}{4}\begin{pmatrix} 1 & \sqrt{3} & \sqrt{3} & 3 \\ -\sqrt{3} & 1 & -3 & \sqrt{3} \\ -\sqrt{3} & -3 & 1 & \sqrt{3} \\ 3 & -\sqrt{3} & -\sqrt{3} & 1 \end{pmatrix}, \qquad D_4(c) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{2.126}$$

On the other hand, constructing the direct sum $D_4' = D_1 \oplus D_1' \oplus D_2$, we have

$$D_4'(a) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ 0 & 0 & -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}, \qquad D_4'(c) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{2.127}$$

The matrices $D_4(a)$ and $D_4'(a)$ are not the same because the bases that we have used to write the tensor product and the direct sum in matrix form are different. However, the two representations are equivalent, $D_4'(g) = S^{-1}D_4(g)S$. A similarity transformation that works to achieve this is

$$S = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ 1 & 0 & 0 & -1 \end{pmatrix}. \tag{2.128}$$

2.10 Additional proofs*

Proof of theorem 2.7.4. Let $F : G \to \mathbb{C}$ be a function acting on the group, so that for every $g \in G$ we have an associated complex number $F(g)$. Recalling that the regular representation $D_R : G \to \mathrm{GL}(V)$ acts on the vector space V with basis consisting of the group elements, we can think of F as a members of the dual space V^*,

$$F(g) = \langle F|g\rangle = \langle F|D_R(g)|e\rangle, \tag{2.129}$$

where in the second equality we use that $D_R(g)|e\rangle = |ge\rangle = |e\rangle$. We can expand $\langle F|$ in the natural basis on V^*

$$\langle F| = \sum_{g' \in G} F(g')\langle g'|, \tag{2.130}$$

so then equation (2.129) becomes

$$F(g) = \sum_{g' \in G} F(g')[D_R(g)]_{g'e}, \tag{2.131}$$

where $[D_R(g)]_{g'e}$ is the matrix element $\langle g'|D_R(g)|e\rangle$. By theorem 2.4.2 the regular representation must be completely reducible, so its matrix elements must be linear combinations of matrix elements of irreducible representations. Thus

$$F(g) = \sum_{r=1}^{\rho} \sum_{i,j} c_{r,ij}[D_r(g)]_{ij}, \tag{2.132}$$

for some coefficients $c_{r,ij}$.

Let's specialise to functions F that are class functions, meaning that they are constant on conjugacy classes $F(g) = F(h^{-1}gh)$ for any $g, h \in G$, and therefore

$$F(g) = \frac{1}{|G|} \sum_{h \in G} F(h^{-1}gh), \tag{2.133}$$

Using equation (2.132) to expand the right-hand side, this implies that

$$\begin{aligned} F(g) &= \frac{1}{|G|} \sum_{r=1}^{\rho} \sum_{h \in G} \sum_{i,j} c_{r,ij} [D_r(h^{-1}gh)]_{ij} \\ &= \frac{1}{|G|} \sum_{r=1}^{\rho} \sum_{h \in G} \sum_{i,j,k,l} c_{r,ij} [D_r(h)^\dagger]_{ik} [D_r(g)]_{kl} [D_r(h)]_{lj}, \end{aligned} \tag{2.134}$$

where we have taken the irreducible representations to be unitary. We can then use the orthogonality theorem 2.6.1 to perform the sum over h,

$$F(g) = \sum_{r=1}^{\rho} \sum_{i,j,k,l} \frac{c_{r,ij}}{n_r} [D_r(g)]_{kl} \delta_{ij} \delta_{kl} = \sum_{r=1}^{\rho} \sum_{j,k} \frac{c_{r,jj}}{n_r} [D_r(g)]_{kk} = \sum_{r=1}^{\rho} \sum_{j} \frac{c_{r,jj}}{n_r} \chi_r(g), \tag{2.135}$$

where n_r is the dimension of D_r. Thus, every class function may be written as a linear combination of the characters.

Let's consider the following class function. For any fixed conjugacy class labelled by β, the function evaluated on a group element g in the αth conjugacy class if $F(g) = 1$ if $\alpha = \beta$ and $F(g) = 0$ otherwise. In other words we are describing $\delta_{\alpha\beta}$ as a class function. We can therefore expand $\delta_{\alpha\beta}$ in terms of the characters of the irreducible representations

$$\delta_{\alpha\beta} = \sum_{r=1}^{\rho} c_{r,\beta} \chi_r(g) = \sum_{r=1}^{\rho} c_{r,\beta} \chi_{r,\alpha} \tag{2.136}$$

for some coefficients $c_{r,\beta}$. To determine these coefficients, multiply both sides by $\chi_{r'}(g)^*$ for some irreducible representation $D_{r'}$ and sum over g. On the left-hand side we get

$$\sum_{g \in G} \delta_{\alpha\beta} \chi_{r'}(g)^* = m_\beta \chi^*_{r',\beta}, \tag{2.137}$$

where m_β is the number of elements in the βth conjugacy class. On the right-hand side we get

$$\sum_{r=1}^{\rho} c_{r,\beta} \sum_{g \in G} \chi^*_{r'}(g) \chi_r(g) = \sum_{r=1}^{\rho} c_r |G| \delta_{rr'} = |G| c_{r'}, \tag{2.138}$$

where the first equality is obtained using the character orthogonality theorem 2.7.2. Comparing these two expressions, we read off that $c_r = m_\beta \chi_{r,\beta}/|G|$ and therefore the character expansion of $\delta_{\alpha\beta}$ is

$$\delta_{\alpha\beta} = \frac{m_\beta}{|G|} \sum_{r=1}^{\rho} \chi_{r,\alpha} \chi^*_{r,\beta}, \tag{2.139}$$

which is equivalent to the expression in the theorem. □

Chapter 3

Symmetric groups

In this chapter we will study an important class of groups, the symmetric groups S_n, which exist for each positive integer n. A given symmetric group S_n is the group formed by the permutations of n objects. One reason why symmetric groups are important is a result called Cayley's theorem, which implies that every group G is isomorphic to a subgroup of the symmetric group $S_{|G|}$, where as usual $|G|$ denotes the order of G.

3.1 Definitions and notation

Definition 3.1.1. The **symmetric group** of degree n, denoted S_n is the group of permutations of n objects. It has order $|\mathrm{S}_n| = n!$.

We will think of the permutations as acting on the set of numbers $X = \{1, 2, \cdots, n\}$. An element of S_n takes the set X and gives another set consisting of the same numbers but in a different order. The order of the group is $n!$ because there are $n!$ different ways of ordering these numbers.

A given permutation π maps 1 to some number $\pi(1)$, 2 to some number $\pi(2)$, and so on. The list of which number maps to which is often written out as

$$\pi = \begin{pmatrix} 1 & 2 & \cdots & n \\ \pi(1) & \pi(2) & \cdots & \pi(n) \end{pmatrix}. \tag{3.1}$$

The product $\pi\pi'$ of two permutations would have $\pi(\pi'(1))$, $\pi(\pi'(2))$, and so on along the bottom row. The product is closed because permuting the elements of a set twice yields another permutation. It is associative because function composition is always associative. There is an identity element, the trivial permutation $\pi(x) = x$ for all x, and every permutation has an inverse permutation. Thus S_n indeed forms a group.

Example 3.1.1. The $3! = 6$ elements of S_3 are

$$e = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \qquad \pi_a = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \qquad \pi_b = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix},$$
$$\pi_c = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \qquad \pi_d = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}, \qquad \pi_f = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}. \tag{3.2}$$

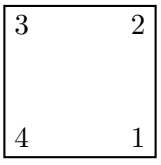
The identity element is π_e, the trivial permutation. Here's some example products,

$$\pi_a\pi_a = \begin{pmatrix} 1 & 2 & 3 \\ \pi_a(2) & \pi_a(3) & \pi_a(1) \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} = \pi_b,$$
$$\pi_a\pi_b = \begin{pmatrix} 1 & 2 & 3 \\ \pi_a(3) & \pi_a(1) & \pi_a(2) \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix} = e, \tag{3.3}$$
$$\pi_b\pi_b = \begin{pmatrix} 1 & 2 & 3 \\ \pi_b(3) & \pi_b(1) & \pi_b(2) \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} = \pi_a. \tag{3.4}$$

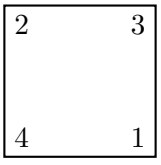
From these products we see that $\{e, \pi_a, \pi_b\}$ form a proper subgroup of S_3. This subgroup is isomorphic to the group considered in example 1.1.1 (and thus also isomorphic to $\mathbb{Z}_3$). The isomorphism is $e \mapsto e$, $\pi_a \mapsto a$, and $\pi_b \mapsto b$. This provides an example of Cayley's theorem, mentioned above.

The whole group S_3 is also isomorphic to some groups we've seen before. It is isomorphic to the group considered in example 1.1.2, with the isomorphism provided by $e \mapsto e$ and $\pi_g \mapsto g$ for $g \in \{a, b, c, d, f\}$. The group S_3 is also isomorphic to D_3, the symmetries of the equilateral triangle. This isomorphism is made manifest if we think about the symmetries of the triangle by labelling the vertices with numbers $\{1, 2, 3\}$ and seeing where these vertices end up after a symmetry transformation. The different symmetry transformations permute the vertices, and thus permute the numbers $\{1, 2, 3\}$. The permutations $\pi_{a,b}$ correspond to rotations while $\pi_{c,d,f}$ correspond to reflections.

Note that it cannot be true in general that S_n is isomorphic to D_n, since $|S_n| = n!$ while $|D_n| = 2n$. Instead D_n is usually just a proper subgroup of S_n. For example, consider D_4, the symmetries of the square. If we number the vertices like this,



then the permutation $\pi = \left(\begin{smallmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 2 & 4 \end{smallmatrix}\right) \in S_4$ would turn the square into this,



which cannot be generated by any symmetry transformation since there is no rotation or reflection that puts the diagonally opposite vertices 1 and 3 next to each other. The

equilateral triangle is special because all of the vertices are next to one another.

From theorem 1.2.7, we know that for each $\pi \in \mathrm{S}_n$ there exists a smallest positive integer r such that $\pi^r = e$. Thus, if we start with an element x of the set on which the permutations act and repeatedly apply π, we will eventually get back to the original element. Thus, repeated application of π must create a chain of elements of X

$$x \to \pi(x) \to \pi(\pi(x)) \to \cdots \to \pi^{r-1}(x) \to x. \tag{3.5}$$

Every element in the middle of this chain must be distinct, because if $\pi^j(x) = \pi^k(x)$ for some positive $j, k < r$, then letting j be the larger of these two integers we can apply π^{-1} repeatedly to obtain $\pi^{j-k}(x) = x$, contradicting the assertion that r is the smallest positive integer such that $\pi^r(x) = x$. The distinctness of these elements also implies that $r \leq n$.[1]

Definition 3.1.2. The chain of distinct elements in equation (3.5), not including the final x, is called the **cycle** generated by x. A cycle containing r elements is called an ***r*-cycle**. We denote the cycle by listing its elements in order in parentheses, $(x\, \pi(x)\, \pi^2(x)\, \cdots\, \pi^{r-1}(x))$. In this list the order of elements is important, but we think of the list as periodic so for exxample we would say

$$(x\, \pi(x)\, \pi^2(x)\, \cdots\, \pi^{r-1}(x)) = (\pi(x)\, \pi^2(x)\, \cdots\, \pi^{r-1}(x)\, x)$$

For a given permutation π, let us define $x \sim y$ if y belongs to the cycle generated by x. This is an equivalence relation, since it satisfies all of the necessary properties:

(i) Reflexivity: $x \sim x$ since x belongs to the cycle it generates by definition.

(ii) Symmetry: Suppose that $x \sim y$. Then $y = \pi^k(x)$ for some postive integer $k < r$. Then dropping the first k terms of the chain in equation (3.5), we have

$$y \to \pi^{k+1}(x) \to \cdots \to x.$$

We then apply π a total of k more times to produce the cycle generated by y, which plainly includes x,

$$y \to \pi^{k+1}(x) \to \cdots \to x \to \pi(x) \to \cdots \to \pi^{k-1}(x) \to y.$$

Thus $x \sim y$ implies that $y \sim x$. Notice that this new cycle contains all of the same elements as the cycle generated by x.

(iii) Transitivity: Suppose that $x \sim y$ and $y \sim z$. We've just pointed out that if $x \sim y$ then the cycles generated by x and y contain all of the same elements. Thus if $y \sim z$, meaning that z is in the cycle generated by y, then z is also in the cycle generated by x and thus $x \sim z$.

The equivalence classes of this equivalence relation are the cycles. Theorem 1.5.2 then tells us that each permutation partitions the set $\{1, 2, \cdots, n\}$ into cycles, with each element of this set appearing in exactly one cycle.

[1] This is stronger than the bound $r \leq |\mathrm{S}_n| = n!$ given by theorem 1.2.7.

Example 3.1.2. We will compute the cycles of the group S_3. The identity permutation has only one-cycles, since for each x the identity permutation generates the chain $x \to e(x) = x$.

$$e = (1)(2)(3). \tag{3.6}$$

The reason why we write that e is equal to its cycles will become clear very soon. For now, just interpret "=" as relating a permutation and its cycles.

To compute the cycles generated by π_a, we first consider the chain generated by acting with π_a on 1. We have $\pi_a(1) = 2$, then $\pi_a^2(1) = \pi_a(2) = 3$, and finally $\pi_a^3(1) = \pi_a(3) = 1$. Thus, for π_a the chain is $1 \to 2 \to 3 \to 1$, so we obtain a single three-cycle

$$\pi_a = (123) \tag{3.7}$$

Proceeding in a similar manner, one finds

$$\pi_b = (132), \qquad \pi_c = (1)(23), \qquad \pi_d = (13)(2), \qquad \pi_f = (12)(3). \tag{3.8}$$

The one-cycles are often omitted when writing permutations in the cycle notation since they act trivially and their presence can be inferred if all of the k-cycles with $k > 1$ are given. For instance, one may write $\pi_c = (23)$ in the previous example.

Every permutation defines a partition into cycles. Conversely, the cycles uniquely specify the permutation. For instance, for S_3 given the cycles $(1)(23)$ we can read off that the corresponding permutation π has $\pi(1) = 1$, $\pi(2) = 3$, and $\pi(3) = 2$. Thus $(1)(23)$ uniquely specifies π_c. This is why we wrote equalities between permutations and cycles in example 3.1.2. Specifying cycles provides an alternative notation to equation (3.1) for defining permutations.

Definition 3.1.3. If a permutation has one cycle of length $r \geq 1$ and all other cycles (if there are any) of length one, then it is a **cyclic permutation** of length r.

From example 3.1.2 we see that all of the permutations in S_3 are cyclic. On the other hand, S_4 has non-cyclic permutations such as $(12)(34)$.

Definition 3.1.4. A cyclic permutation of length two is called a transposition.

A transposition exchanges two objects while leaving the rest invariant. For example, in S_3 the permutations $\pi_{c,d,f}$ are transpositions.

Here is an example of how we can multiply permutations. For example let's consider the product $\pi_a\pi_c$ in S_3. In cycle notation this is

$$\pi_a\pi_c = (123)(23). \tag{3.9}$$

To determine the result of the product we consider each number and look at where it ends up after performing the permutations sequentially, reading from right to left. For example, 1 does not appear in the cycle first from the right, while the second cycle sends $1 \to 2$. The first cycle sends $2 \to 3$ and then the second cycle sends $3 \to 1$, so the net effect of $\pi_a\pi_c$ is to send $2 \to 1$. Finally, the first cyle sends $3 \to 2$ and then the second cycle sends $2 \to 3$, for a net effect of $3 \to 3$. We conclude that

$$\pi_a\pi_c = (12) = \pi_f, \tag{3.10}$$

where we have left implicit the one-cycle (3). As another example, consider the product of the permutations (243)(15) and (13)(45) in S_5,

$$(243)(15)(13)(45). \tag{3.11}$$

Working through the cycles from right to left, we see for example that $1 \to 3 \to 2$, $2 \to 4$, and $4 \to 5 \to 1$, so one of the cycles appearing in the product is (124). One can check in the same way that that 3 and 5 form a two-cycle, and so the product is

$$(243)(15)(13)(45) = (124)(35). \tag{3.12}$$

The inverse of a permutation is obtained by reversing the order of its cycles. For example $((243)(15))^{-1} = (342)(51)$.

Definition 3.1.5. The **parity** of a permutation π consisting of k cycles, each with length r_i, is **even** or **odd** if its **sign** $\mathrm{sgn}(\pi)$

$$\mathrm{sgn}(\pi) = (-1)^P, \qquad P = \sum_{i=1}^{k} (r_i - 1). \tag{3.13}$$

is a $+1$ or -1, respectively. In other words, P is even for an even permutation and odd for an odd permutation.

The sign of a product of two permutations is the product of the two signs. Thus, the product of two even permutations or two odd permutations is even, while the product of an even permutation with an odd permutation is odd. The identity permutation is always even, since it consists only of one-cycles and thus has $P = 0$. This means that the sign has the correct structure to form a representation:

Definition 3.1.6. For every symmetric group S_n, $n \geq 2$, the **sign representation** $D : S_n \to \mathrm{GL}(\mathbb{R})$ is a non-trivial one-dimensional representation defined by

$$D(\pi) = \mathrm{sgn}(\pi), \tag{3.14}$$

for all $\pi \in S_n$.

Example 3.1.3. The even permutations of S_3 are e and $\pi_{a,b}$. The odd permutations are $\pi_{c,d,f}$. Thus, the sign representation of S_3 is

$$D(e) = D(\pi_a) = D(\pi_b) = 1, \qquad D(\pi_c) = D(\pi_d) = D(\pi_f) = -1. \tag{3.15}$$

Since the product of two even permutations is even, the identity element is even, and the inverse of an even permutation is also even, the even permutations of S_n form a subgroup:

Definition 3.1.7. The subgroup of S_n consisting of all of the even permutations is called the **alternating group** and denoted A_n.

3.2 Conjugacy classes

The conjugacy classes of a finite group are very important for the representation theory of that group. The conjugacy classes of the symmetric groups are determined by the following theorem:

Theorem 3.2.1. *Two permutations in* S_n *are conjugate if and only if they have the same cycle structure.*

Two permutations have the same cycle structure if they have the same number of r-cycles for each r, for example $(123)(45)$ and $(153)(24)$ have the same cycle structure, so belong to the same conjugacy class of S_5. Before proving the theorem, we will prove the following lemma:

Lemma 3.2.1. *Suppose* α *is an* r*-cycle in* S_n*,*

$$\alpha = (a_1\, a_2\, \cdots\, a_r), \tag{3.16}$$

for some numbers a_i*, where* $a_{i+1} = \alpha(a_i)$ *with* $a_{r+1} \equiv a_1$*. Then given a permutation* $\sigma \in \mathrm{S}_n$*, the result of* $\sigma\alpha\sigma^{-1}$ *is an* r*-cycle*

$$\sigma\pi\sigma^{-1} = (\sigma(a_1)\, \sigma(a_2)\, \cdots\, \sigma(a_r)). \tag{3.17}$$

Proof. Suppose that b is any number not appearing in the cycle α, meaning that $\alpha(b) = b$. Consider $\sigma\alpha\sigma^{-1}$ acting on $\sigma(b)$. We have

$$\sigma\alpha\sigma^{-1}(\sigma(b)) = \sigma\alpha(b) = \sigma(b). \tag{3.18}$$

This tells us that any number of the form $\sigma(b)$ with b not in the cycle α is left unchanged by the permutation $\sigma\alpha\sigma^{-1}$ and thus belongs in a one-cycle of $\sigma\alpha\sigma^{-1}$. Now consider what happens when we act with this permuation on $\sigma(a_i)$, where a_i is a member of the cycle α,

$$\sigma\alpha\sigma^{-1}(\sigma(a_i)) = \sigma\alpha(a_i) = \sigma(a_{i+1}). \tag{3.19}$$

This tells us that $\sigma(a_{i+1})$ appears to the right of $\sigma(a_i)$ in a cycle, for all i. Thus $\sigma\alpha\sigma^{-1}$ consists of the single r-cycle stated in the lemma. □

Proof of theorem 3.2.1. We first prove that two permutations that are conjugate have the same cycle structure. Consider a permutation π consisting of m cycles and think of it as a product of its individual cycles,

$$\pi = \alpha_1\alpha_2\cdots\alpha_m, \tag{3.20}$$

where the α_i are disjoint r_i-cycles, i.e. they have no elements in common. The permutation π' obtained by conjugating π with another element σ is then

$$\begin{aligned} \pi' &= \sigma\pi\sigma^{-1} \\ &= \sigma\alpha_1\alpha_2\cdots\alpha_m\sigma^{-1} \\ &= (\sigma\alpha_1\sigma^{-1})(\sigma\alpha_2\sigma^{-1})\cdots(\sigma\alpha_m\sigma^{-1}). \end{aligned} \tag{3.21}$$

From lemma 3.2.1 we know that each of the $(\sigma\alpha_i\sigma^{-1})$ are r_i-cycles, with elements given by σ acting on the elements of α_i. Since the α_i are disjoint, and since σ is a bijection and thus in particular does not map two objects to the same object, the $(\sigma\alpha_i\sigma^{-1})$ are also disjoint.

It remains to prove that any two permutations π and τ with the same cycle structure are conjugate. Let us write the cycles of π and τ as α and β, respectively,

$$\pi = \alpha_1\alpha_2\cdots\alpha_m, \qquad \tau = \beta_1\beta_2\cdots\beta_m, \tag{3.22}$$

where both α_i and β_i are r_i-cycles for each i. Let us write the elements of the cycles as

$$\alpha_i = (a_{i1}\, a_{i2},\, \cdots, a_{ir_i}), \qquad \beta_i = (b_{i1},\, b_{i2},\, \cdots\, b_{ir_i}). \tag{3.23}$$

The objects a_{ij} consist of the set of numbers $\{1, 2, \cdots, n\}$ acted on by permutations in S_n, and by construction no two a_{ij} with different indices are equal. The same is true for the b_{ij}. Thus there must exist a permutation $\sigma \in \mathrm{S}_n$ such that $b_{ij} = \sigma(\alpha_{ij})$ for all i and j. From lemma 3.2.1 we then have that

$$b_i = \sigma\alpha_i\sigma^{-1}, \tag{3.24}$$

and therefore

$$\tau = \beta_1\beta_2\cdots\beta_m = (\sigma\alpha_1\sigma^{-1})(\sigma\alpha_2\sigma^{-1})\cdots(\sigma\alpha_m\sigma^{-1}) = \sigma\alpha_1\alpha_2\cdots\alpha_m\sigma^{-1} = \sigma\alpha\sigma^{-1}, \tag{3.25}$$

meaning that π and τ are conjugate to one another. □

Theorem 3.2.1 tells us that the conjugacy classes of S_n consist of the set of all permutations with a given cycle structure. Thus a conjugacy class is specified by the number ν_r of r-cycles for each r. Sometimes the conjugacy class corresponding to a given set of $\{\nu_r\}$ will be denoted $(1^{\nu_1}, 2^{\nu_2}, \cdots, n^{\nu_n})$.

The total number of elements that we can distribute among the cycles is n, so a given conjugacy class has

$$\sum_{r=1}^{n} r\nu_r = n. \tag{3.26}$$

Rewriting the sum as

$$(\nu_1 + \nu_2 + \cdots + \nu_n) + (\nu_2 + \nu_3 + \cdots + \nu_n) + \cdots + (\nu_n) = n, \tag{3.27}$$

we see that by defining the non-negative integers $\lambda_j = \sum_{r=j}^{n} \nu_r$ the condition above becomes

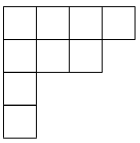
$$\sum_{j=1}^{n} \lambda_j = n. \tag{3.28}$$

The conjugacy class is equally well specified by listing the set $\{\lambda_j\}$, since this set determines $\{\nu_r\}$ through the relation $\nu_r = \lambda_r - \lambda_{r+1}$ for $r < n$ and $\lambda_n = n$. Since the ν_r are non-negative, this implies a chain of inequalities $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$.

A decreasing sequence of non-negative integers satisfying (3.28) is called a partition of n. The number of conjugacy classes of S_n is then equal to the number of partitions of

n. There is no known closed form expression for the number of partitions of n, and thus the number of conjugacy classes of S_n. Integer partitions, and therefore also conjugacy classes of symmetric groups, can be represented by Young diagrams, consisting of n boxes arranged in a left-justified table, with λ_1 boxes in the first row, λ_2 boxes in the second row, and so on. The number of r-cycles ν_r is then equal to the difference between the numbers of boxes in the rth and $(r+1)$th rows. This means that the number of r-cycles is equal to the number of columns in the diagram containing r boxes.

Example 3.2.1. Consider the conjugacy class $(1, 2^2, 4)$ of S_9. This has $\nu_1 = 1$, $\nu_2 = 2$, $\nu_3 = 0$, and $\nu_4 = 1$, along with $\nu_r = 0$ for $5 \leq r \leq 9$. This corresponds to $\lambda_1 = 4$, $\lambda_2 = 3$, and $\lambda_3 = \lambda_4 = 1$ with all other λ_j vanishing. The Young diagram is then:



Every Young diagram with n boxes describes a partition of n, and therefore also a conjugacy class of S_n. The identity permutation of S_n consists of n one-cycles, so its conjugacy class is (n). This has $\nu_n = 1$ with all other ν_r zero, so $\lambda_j = 1$ for all j. The corresponding Young diagram is a single column of n boxes.

Theorem 3.2.2. *The number of permutations in the conjugacy class of* S_n *with* ν_r *r-cycles for each r is*

$$\frac{n!}{\prod_{r=1}^{n} r^{\nu_r} \nu_r!}. \tag{3.29}$$

Proof. Let us count how many ways we can construct a given element of this conjugacy class by filling up the cycles one-by-one. We begin with the one-cycles. We need to choose ν_1 of the numbers $\{1, \cdots, n\}$ to go in these cycles. Since the order in which we list the cycles doesn't matter, the number of ways of doing this is the binomial coefficient

$$\binom{n}{\nu_1} = \frac{n!}{\nu_1!(n-\nu_1)!}. \tag{3.30}$$

After filling up the one-cycles, we are left with $(n - \nu_1)$ numbers from which to choose the $2\nu_2$ elements of the two-cycles. The number of ways of doing this is not simply given by a binomial coefficient, however, since we also need to pair up the numbers we choose into cycles. We can instead build the two-cycles one-by-one. There are $(n - \nu_1)$ possible choices for the first element of the first two-cycle and $(n - \nu_1 - 1)$ choices for the second element, but the which element of the two-cycle we list first is irrelevant, and thus the total number of ways of building the first two-cycle is $\frac{1}{2}(n - \nu_1)(n - \nu_1 - 1)$. Continuing in this way, and remembering to divide by a factor of $\nu_2!$ to account for the fact that the order in which we list the two-cycles doesn't matter, we find that the number of different

ways to construct the two-cycles is

$$\frac{1}{\nu_2!}\times\frac{(n-\nu_1)(n-\nu_1-1)}{2}\times\frac{(n-\nu_1-2)(n-\nu_1-3)}{2}\times\cdots \\ \times\frac{(n-\nu_1-2\nu_2+2)(n-\nu_1-2\nu_2+1)}{2}=\frac{(n-\nu_1)!}{2^{\nu_2}\nu_2!(n-\nu_1-2\nu_2)!}. \quad (3.31)$$

Following the same procedure, the number of three-cycles is[2]

$$\frac{(n-\nu_1-2\nu_2!)}{3^{\nu_3}\nu_3!(n-\nu_1-2\nu_2-3\nu_3)!}, \quad (3.32)$$

and so on. Multiplying the number of r-cycles together, we find that the number of permutations in the conjugacy class is

$$\frac{n!}{\nu_1!(n-\nu_1)!}\times\frac{(n-\nu_1)!}{2^{\nu_2}\nu_2!(n-\nu_1-2\nu_2)!}\times\cdots\times\frac{(n-\nu_1-2\nu_2-\cdots-(r-1)\nu_{r-1})!}{r^{\nu_r}\nu_r!(n-\nu_1-2\nu_2-\cdots-r\nu_r)!}, \quad (3.33)$$

which simplifies to equation (3.29), using equation (3.26) to note that the right-most factorial in the right-most denominator equals one. □

3.3 Representation theory of S_n and Young tableaux

We have seen that every conjugacy class of S_n can be associated with a Young diagram with n boxes, and vice versa. Since the number of inequivalent irreducible representations of a finite group is equal to the number of conjugacy classes, we also have that the number of Young diagrams for a given n is equal to the number of irreducible representations. It turns out that there is a natural bijection between Young diagrams and irreducible representations, so that we can naturally associate an irreducible representation to each Young diagram. Moreover, the Young diagram encodes some properties of its corresponding irreducible representation. There is not time to cover the theory behind this in this course, so we will simply state some results without proof and see how the story works for some simple examples.

Definition 3.3.1. A **Young tableau** is a Young diagram with the numbers $1, 2, \cdots, n$ written in the boxes, where n is the number of boxes.

For example, these are two different Young tableaux corresponding to the same Young diagram:

1	2	3	4
5	6		
7			

7	2	5	6
4	3		
1			

Permutations act naturally on Young tableaux by permuting the numbers written in the boxes. For example, we can obtain the Young tableau on the right from the one on the left using the permutation

$$\pi=\begin{pmatrix}1&2&3&4&5&6&7\\7&2&5&6&4&3&1\end{pmatrix}=(17)(3546). \quad (3.34)$$

[2]The factor of 3^{ν_3} in equation (3.32) accounts for overcounting that occurs because three-cycles satisfy $(xyz)=(yzx)=(zxy)$.

Acting with this permutation, we find:

$$\pi \left(\begin{array}{|c|c|c|c|} \hline 1 & 2 & 3 & 4 \\ \hline 5 & 6 \\ \cline{1-2} 7 \\ \cline{1-1} \end{array} \right) = \begin{array}{|c|c|c|c|} \hline 7 & 2 & 5 & 6 \\ \hline 4 & 3 \\ \cline{1-2} 1 \\ \cline{1-1} \end{array} \tag{3.35}$$

For a given Young diagram, we can naturally associate each possible Young tableau with an element of S_n and vice versa, as follows. We can construct a canonical Young tableau by filling in the numbers from 1 to n sequentially from left-to-right and top-to-bottom, as in the Young tableau:

$$\begin{array}{|c|c|c|c|} \hline 1 & 2 & 3 & 4 \\ \hline 5 & 6 \\ \cline{1-2} 7 \\ \cline{1-1} \end{array}$$

Any Young tableau t of the same shape is then associated to the permutation that maps this canonical Young tableau into t. For example

$$\begin{array}{|c|c|c|c|} \hline 7 & 2 & 5 & 6 \\ \hline 4 & 3 \\ \cline{1-2} 1 \\ \cline{1-1} \end{array}$$

is associated to the permutation π in equation (3.34).

This one-to-one correspondence between Young tableaux and permutations assumes we have fixed the shape of the underlying Young diagram (i.e. that we have fixed the arrangement of boxes). For example, any of these Young tableaux also correspond to the permutation (3.34):

$$\begin{array}{|c|c|c|c|c|c|c|} \hline 7 & 2 & 5 & 6 & 4 & 3 & 1 \\ \hline \end{array} \qquad \begin{array}{|c|c|} \hline 7 & 2 \\ \hline 5 & 6 \\ \hline 4 & 3 \\ \hline 1 \\ \cline{1-1} \end{array} \qquad \begin{array}{|c|c|c|} \hline 7 & 2 & 5 \\ \hline 6 & 4 & 3 \\ \hline 1 \\ \cline{1-1} \end{array}$$

We have seen that all of the n-box Young tableaux of a given shape correspond to element of S_n and that we can act with elements of S_n on these Young tableaux. This means that if we think of the Young tableaux as basis elements of a vector space, then the acting with permutations on Young tableaux is just a fancy way of writing the regular representation. For example, let's consider S_3 and Young tableaux of shape $\begin{array}{|c|c|} \hline & \\ \hline \\ \cline{1-1} \end{array}$, of which there are six:

$$\begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 \\ \cline{1-1} \end{array} \qquad \begin{array}{|c|c|} \hline 2 & 3 \\ \hline 1 \\ \cline{1-1} \end{array} \qquad \begin{array}{|c|c|} \hline 3 & 1 \\ \hline 2 \\ \cline{1-1} \end{array} \qquad \begin{array}{|c|c|} \hline 1 & 3 \\ \hline 2 \\ \cline{1-1} \end{array} \qquad \begin{array}{|c|c|} \hline 3 & 2 \\ \hline 1 \\ \cline{1-1} \end{array} \qquad \begin{array}{|c|c|} \hline 2 & 1 \\ \hline 3 \\ \cline{1-1} \end{array}$$

We think of these as basis elements, and then a general vector transforming in the regular representation is

$$|v\rangle = c_1 \begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 \\ \cline{1-1} \end{array} + c_2 \begin{array}{|c|c|} \hline 2 & 3 \\ \hline 1 \\ \cline{1-1} \end{array} + c_3 \begin{array}{|c|c|} \hline 3 & 1 \\ \hline 2 \\ \cline{1-1} \end{array} + c_4 \begin{array}{|c|c|} \hline 1 & 3 \\ \hline 2 \\ \cline{1-1} \end{array} + c_5 \begin{array}{|c|c|} \hline 3 & 2 \\ \hline 1 \\ \cline{1-1} \end{array} + c_6 \begin{array}{|c|c|} \hline 2 & 1 \\ \hline 3 \\ \cline{1-1} \end{array}, \tag{3.36}$$

with complex coefficients c_i. Actubg with the representation matrix of the permutation $\pi_a = \left(\begin{smallmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{smallmatrix}\right) = (123)$ gives

$$D(\pi_a)|v\rangle = c_1 \begin{array}{|c|c|} \hline 2 & 3 \\ \hline 1 \\ \cline{1-1} \end{array} + c_2 \begin{array}{|c|c|} \hline 3 & 1 \\ \hline 2 \\ \cline{1-1} \end{array} + c_3 \begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 \\ \cline{1-1} \end{array} + c_4 \begin{array}{|c|c|} \hline 2 & 1 \\ \hline 3 \\ \cline{1-1} \end{array} + c_5 \begin{array}{|c|c|} \hline 1 & 3 \\ \hline 2 \\ \cline{1-1} \end{array} + c_6 \begin{array}{|c|c|} \hline 3 & 2 \\ \hline 1 \\ \cline{1-1} \end{array}, \tag{3.37}$$

since π_a sends $1 \to 2$, $2 \to 3$, and $3 \to 1$.

We have seen that the regular representation is reducible, and contains all of the irreducible representations. We will now see how to use Young tableaux to construct all of the irreducible subspaces of the regular representation, thus obtaining all of the irreducible representations of S_n.

Definition 3.3.2. A **standard Young tableau** is a Young tableau such that the numbers in every row increase from left to right, and the numbers in every column increase from top to bottom.

For example, this is a standard Young tableau:

1	2	5	6
3	4	7	
8			

These are not:

1	2	5	6
4	3	7	
8			

1	3	6	7
2	4	5	
8			

The tableau on the left is non-standard because 3 appears after 4 on the second row. The tableau on the right is non-standard because 5 appears below 6 in the third column.

Given a standard Young tableau t with n boxes, which we have seen can be thought of as a state in the regular representation, we define another state $|e_t\rangle$ in the regular representation of S_n as

$$|e_t\rangle = \sum_{\pi \in C_t} \sum_{\rho \in R_t} \mathrm{sgn}(\pi) \rho(\pi(t)), \tag{3.38}$$

where C_t is the set of permutations that preserve the columns, meaning that they shuffle the elements in each column of t, without moving elements between different columns, and R_t is the set of permutations preserve the rows. By $\rho(\pi(t))$ are the Young tableaux corresponding obtained by acting on t with π followed by ρ. For example, consider the standard Young tableau

$$t = \begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 \\ \cline{1-1} \end{array}.$$

There are two permutations in C_t. One is the identity permutation, which has positive sign. The other is (13), which has negative sign. Likewise, there are two permutations in R_t, the identity and 12. Thus

$$|e_t\rangle = \begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 \\ \cline{1-1} \end{array} + \begin{array}{|c|c|}\hline 2 & 1 \\ \hline 3 \\ \cline{1-1} \end{array} - \begin{array}{|c|c|}\hline 3 & 2 \\ \hline 1 \\ \cline{1-1} \end{array} - \begin{array}{|c|c|}\hline 2 & 3 \\ \hline 1 \\ \cline{1-1} \end{array}. \tag{3.39}$$

Note that for any standard Young tableau t the identity permutation is an element of both C_t and R_t, and so t will always appear in $|e_t\rangle$. The other Young tableaux in $|e_t\rangle$ will be non-standard, since the permutations creating them shuffle the columns or rows and so at least one column or row will be non-increasing.

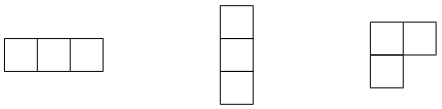
Here is the key theorem for constructing the irreducible representations of S_n:

Theorem 3.3.1. *Given a Young diagram with n boxes, the set of states* $\{|e_t\rangle\}$ *formed from the set of standard Young tableaux* $\{t\}$ *using equation* (3.38) *forms a basis for an irreducible subspace of the regular representation of* S_n*, so we can construct an irreducible representation of* S_n *by considering the action of the regular representation on this subspace. Irreducible representations formed in this way from different Young diagrams are inequivalent.*

Since the number of Young diagrams is equal to the number of irreducible representations, we can construct all of the irreducible representations of S_n in this way. Since we construct a basis element $|e_t\rangle$ for each standard Young tableau t, the dimension of the irreducible representation is equal to the number of standard Young tableaux.

Let's see how this works for S_3, which is isomorphic to D_3. We constructed the irreducible representations of D_3 in section 2.9.2. We will reproduce this using Young diagrams.

Example 3.3.1 (Irreducible representations of S_3)**.** There are three possible Young diagrams with three boxes:



Each of these will correspond to a different irreducible representation of S_3.

We begin with the Young diagram on the left, consisting of a single row of three boxes. There is only one possible standard Young tableau

$$t = \boxed{1}\boxed{2}\boxed{3}.$$

The corresponding irreducible representation will be one-dimensional. We will call this representation D_1.

The only element of C_t is the trivial permutation; since there columns each contain only a single box there is no non-trivial permutation that doesn't exchange elements between columns. On the other hand, $R_t = S_3$; since there is only a single row, every permutation will preserve the rows. Thus the basis element corresponding to this Young tableau is

$$|e_t\rangle = \boxed{1}\boxed{2}\boxed{3} + \boxed{2}\boxed{3}\boxed{1} + \boxed{3}\boxed{1}\boxed{2} + \boxed{1}\boxed{3}\boxed{2} + \boxed{3}\boxed{2}\boxed{1} + \boxed{2}\boxed{1}\boxed{3}. \tag{3.40}$$

Recalling that each Young tableau may be identified with a permutation, we can just as well write this as

$$|e_t\rangle = \sum_{\pi \in S_3} |\pi\rangle \tag{3.41}$$

where for instance the identity permutation $|(1)(2)(3)\rangle$ corresponds to $\boxed{1}\boxed{2}\boxed{3}$ and $|(123)\rangle$ would correspond to $\boxed{2}\boxed{1}\boxed{3}$.

Since all of the possible Young tableaux appear in this sum, all with plus signs, acting with any permutation on this state will simply reorder the elements of the sum.

We thus have $D_1(\pi)|e_t\rangle = |e_t\rangle$ for any $\pi \in \mathrm{S}_3$. This is easy to see from equation (3.41), we have that

$$D_1(\pi)|e_t\rangle = \sum_{\pi' \in \mathrm{S}_3} D_1(\pi)|\pi'\rangle = \sum_{\pi' \in \mathrm{S}_3} |\pi\pi'\rangle = \sum_{\pi' \in \mathrm{S}_3} |\pi'\rangle = |e_t\rangle, \tag{3.42}$$

where the second-to last equality arises because of the permutation law 1.2.6, which tells us that summing over all $\pi\pi'$ is equivalent to summing over all π'. Since $D_1(\pi)$ leaves the basis state invariant for all π, we thus identify D_1 as the trivial representation.

Now consider the next Young diagram, consisting of a single column of three boxes. The only possible standard Young tableau is

$$t = \begin{array}{|c|}\hline 1 \\ \hline 2 \\ \hline 3 \\ \hline \end{array}.$$

Again since there is only one standard Young tableau, the irreducible representation we construct will be one-dimensional. We will call this irreducible representation D'_1.

This time R_t consists only of the trivial representation, since the rows each contain only a single box. On the other hand $C_t = \mathrm{S}_3$. Thus the single basis state corresponding to $|e_t\rangle$ is obtained by summing over all of the permutations weighted by their signs,

$$|e_t\rangle = \begin{array}{|c|}\hline 1 \\ \hline 2 \\ \hline 3 \\ \hline \end{array} + \begin{array}{|c|}\hline 2 \\ \hline 3 \\ \hline 1 \\ \hline \end{array} + \begin{array}{|c|}\hline 3 \\ \hline 1 \\ \hline 2 \\ \hline \end{array} - \begin{array}{|c|}\hline 1 \\ \hline 3 \\ \hline 2 \\ \hline \end{array} - \begin{array}{|c|}\hline 3 \\ \hline 2 \\ \hline 1 \\ \hline \end{array} - \begin{array}{|c|}\hline 2 \\ \hline 1 \\ \hline 3 \\ \hline \end{array} = \sum_{\pi \in \mathrm{S}_3} \mathrm{sgn}(\pi)|\pi\rangle. \tag{3.43}$$

We then have that

$$\begin{aligned} D'_1(\pi)|e_t\rangle &= \sum_{\pi' \in \mathrm{S}_3} \mathrm{sgn}(\pi') D'_1(\pi)|\pi'\rangle \\ &= \sum_{\pi' \in \mathrm{S}_3} \mathrm{sgn}(\pi')|\pi\pi'\rangle \\ &= \mathrm{sgn}(\pi) \sum_{\pi' \in \mathrm{S}_3} \mathrm{sgn}(\pi\pi')|\pi\pi'\rangle \\ &= \mathrm{sgn}(\pi) \sum_{\pi' \in \mathrm{S}_3} \mathrm{sgn}(\pi')|\pi'\rangle \\ &= \mathrm{sgn}(\pi')|e_t\rangle. \end{aligned} \tag{3.44}$$

In the third line we have made use of the facts that $\mathrm{sgn}(\pi\pi') = \mathrm{sgn}(\pi)\mathrm{sgn}(\pi')$ and $\mathrm{sgn}(\pi) = \pm 1$ so that $\mathrm{sgn}(\pi)^2 = 1$. This one-dimensional representation is thus $D'_1(\pi) = \mathrm{sgn}(\pi)$. With the labelling of S_3 permutations used in example 3.1.1 we have

$$D'_1(e) = D'_1(\pi_a) = D'_1(\pi_b) = 1, \qquad D'_1(\pi_c) = D'_1(\pi_d) = D'_1(\pi_f) = -1. \tag{3.45}$$

Using the isomorphism between S_3 and D_3 this matches the one-dimensional irreducible representation of D_3 in equation (2.121) that we also called D'_1.

There are two possible standard Young tableaux corresponding to the final Young diagram,

$$t_1 = \begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 \\ \cline{1-1}\end{array}, \qquad t_2 = \begin{array}{|c|c|}\hline 1 & 3 \\ \hline 2 \\ \cline{1-1}\end{array}. \tag{3.46}$$

The corresponding irreducible representation will therefore be two-dimensional. We will call it D_2'. We already constructed the basis state corresponding to t_1 in equation (3.39). We will record it here again, along with the basis state corresponding to t_2 which may be constructed the same way,

$$\begin{aligned} |e_{t_1}\rangle &= \begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 \\ \cline{1-1}\end{array} + \begin{array}{|c|c|}\hline 2 & 1 \\ \hline 3 \\ \cline{1-1}\end{array} - \begin{array}{|c|c|}\hline 3 & 2 \\ \hline 1 \\ \cline{1-1}\end{array} - \begin{array}{|c|c|}\hline 2 & 3 \\ \hline 1 \\ \cline{1-1}\end{array}, \\ |e_{t_2}\rangle &= \begin{array}{|c|c|}\hline 1 & 3 \\ \hline 2 \\ \cline{1-1}\end{array} + \begin{array}{|c|c|}\hline 3 & 1 \\ \hline 2 \\ \cline{1-1}\end{array} - \begin{array}{|c|c|}\hline 2 & 3 \\ \hline 1 \\ \cline{1-1}\end{array} - \begin{array}{|c|c|}\hline 3 & 2 \\ \hline 1 \\ \cline{1-1}\end{array}. \end{aligned} \tag{3.47}$$

Let's consider the action of $D_2'(\pi_a)$ on these states. Since π_a is the permutation (123), it maps $1 \to 2$, $2 \to 3$, and $3 \to 1$, so we have

$$\begin{aligned} D_2'(\pi_a)|e_{t_1}\rangle &= \begin{array}{|c|c|}\hline 2 & 3 \\ \hline 1 \\ \cline{1-1}\end{array} + \begin{array}{|c|c|}\hline 3 & 2 \\ \hline 1 \\ \cline{1-1}\end{array} - \begin{array}{|c|c|}\hline 1 & 3 \\ \hline 2 \\ \cline{1-1}\end{array} - \begin{array}{|c|c|}\hline 3 & 1 \\ \hline 2 \\ \cline{1-1}\end{array} = -|e_{t_2}\rangle, \\ D_2'(\pi_a)|e_{t_2}\rangle &= \begin{array}{|c|c|}\hline 2 & 1 \\ \hline 3 \\ \cline{1-1}\end{array} + \begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 \\ \cline{1-1}\end{array} - \begin{array}{|c|c|}\hline 3 & 1 \\ \hline 2 \\ \cline{1-1}\end{array} - \begin{array}{|c|c|}\hline 1 & 3 \\ \hline 2 \\ \cline{1-1}\end{array} = |e_{t_1}\rangle - |e_{t_2}\rangle. \end{aligned} \tag{3.48}$$

If we write the basis elements as column vectors, $|e_{t_1}\rangle = \left(\begin{smallmatrix}1\\0\end{smallmatrix}\right)$ and $|e_{t_2}\rangle = \left(\begin{smallmatrix}0\\1\end{smallmatrix}\right)$, then this tells us that $D_2'(\pi_a)$ is the matrix

$$D_2'(\pi_a) = \begin{pmatrix} 0 & 1 \\ -1 & -1 \end{pmatrix}. \tag{3.49}$$

Carrying out the same procedure for the other permutations, one finds

$$\begin{aligned} D_2'(\pi_b) &= \begin{pmatrix} -1 & -1 \\ 1 & 0 \end{pmatrix}, & D_2'(\pi_c) &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \\ D_2'(\pi_d) &= \begin{pmatrix} -1 & -1 \\ 0 & 1 \end{pmatrix}, & D_2'(\pi_f) &= \begin{pmatrix} 1 & 0 \\ -1 & -1 \end{pmatrix}. \end{aligned} \tag{3.50}$$

This is not the same as the two-dimensional irreducible representation of D_3 written in equation (2.122) but the two representations are equivalent, $D_2 = S^{-1} D_2' S$. The similarity transformation relating the two is given by

$$S = \begin{pmatrix} 1 & -\frac{1}{\sqrt{3}} \\ -1 & -\frac{1}{\sqrt{3}} \end{pmatrix}. \tag{3.51}$$

The Young tableaux method has managed to reproduce all of the irreducible representations of $\mathrm{S}_3 \cong \mathrm{D}_3$ that we previously found. The calculations work for S_n for any n, although they will quickly become impractical to do by hand. However, this method is purely algorithmic and could programmed on a computer.

In the previous example we saw that for S_3 the Young diagram consisting of a single

row corresponded to the trivial representation. This generalises to S_n for any n. The only standard Young tableau consisting of a single row of n boxes is

$$\boxed{1}\boxed{2}\boxed{3}\cdots\boxed{n}$$

which gives a basis state for the corresponding irreducible subspace same form as for S_3,

$$|e\rangle = \sum_{\pi \in S_n} |\pi\rangle. \tag{3.52}$$

The regular representation of any permutation leaves this state invariant, $D(\pi)|e\rangle = |e\rangle$, and thus this state transforms in the trivial representation.

Similarly, there is only one standard Young tableau consisting of a single column of n boxes,

$$\begin{array}{|c|}\hline 1 \\ \hline 2 \\ \hline 3 \\ \hline \end{array}\\ \vdots \\ \begin{array}{|c|}\hline n \\ \hline \end{array}$$

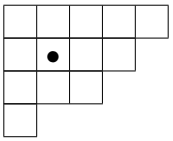
which gives a corresponding basis state

$$|e\rangle = \sum_{\pi \in S_n} \mathrm{sgn}(\pi)|\pi\rangle. \tag{3.53}$$

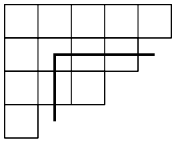
Just like for S_3, the regular representation of any permutation acts as $D(\pi)|e\rangle = \mathrm{sgn}(\pi)|e\rangle$, and thus this representation corresponds to the sign representation.

The dimension of the irreducible representation corresponding to a Young diagram of a given shape is equal to the number of standard Young tableaux with that shape. There is a formula that may be used to compute the dimension of this irreducible representation without having to determine all of the standard Young tableaux.

The hook length associated to any box b in a Young diagram is equal to one plus the number of boxes to the right of b in the same row plus the number of boxes below b in the same column. So for example the hook length of the marked box in this Young diagram



is four, since there are two boxes to the right and one box below. In other words, the hook length is the length (measured in number of boxes) of the line formed from the horizontal segment and a vertical segment that meet at the box, like this:

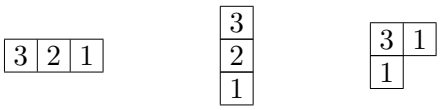


Theorem 3.3.2 (Hook length formula)**.** *The dimension of the irreducible representation of* S_n *associated to a Young diagram is*

$$\frac{n!}{H}, \tag{3.54}$$

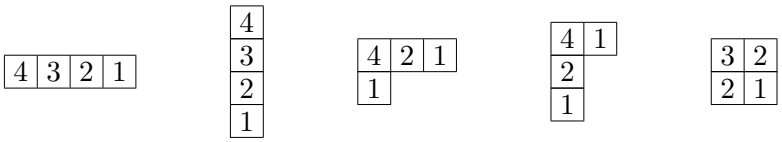
where H *is the product of the hook lengths of all of the boxes in the Young diagram.*

Example 3.3.2. Here are all of the Young diagrams for S_3, with the hook lengths of the boxes written in them (sorry that these look so much like Young tableaux):



We have $n! = 6$, and the first two tableaux have $H = 6$ and therefore correspond to representations of dimension $6/6 = 1$, as expected. The right-most tableau has $H = 3$ and so corresponds to a two-dimensional representation.

Example 3.3.3. Here are all of the Young diagrams for S_4 with the hook lengths written in the boxes, and the corresponding H factors written below:



H :	24	24	8	8	12
$\frac{4!}{H}$:	1	1	3	3	2

The first two Young diagrams therefore correspond to one-dimensional representations. This is as expected, as discussed above these Young diagrams correspond to the trivial and the sign representations. The next two Young diagrams correspond to three-dimensional representations. The final Young diagram corresponds to a two-dimensional representation.

As a check that this makes sense, we can use theorem 2.7.6 which tells us that the sum of the squares of the dimensions of the irreducible representations equals the order of the group, which is $4! = 24$. The theorem is indeed satisfied by the dimensions we have found.

Part II

Lie groups

Chapter 4

Lie groups

Lie groups are a class of infinite groups that are continuous, meaning that they can parameterised by a set of real or complex numbers, with the property that the group multiplication is differentiable with respect to those parameters. Lie groups arise in physics whenever we want to describe a continuous symmetry such as the rotational symmetry of the sphere. In this example, the group elements are all the possible rotations and the parameters are the angles of rotation. Other examples of continuous symmetries are the translational symmetry of the plane, Lorentz invariance of Minkowski space, and the gauge symmetry of electromagnetism.

The precise mathematical definition of a Lie group is a little abstract and requires mathematics that goes beyond the scope of this course. Fortunately, there is a special class of Lie groups, called *matrix Lie groups*, which can be defined much more simply and concretely. Almost all of the Lie groups important in physics are matrix Lie groups, so we will not lose much by restricting to this special case.

4.1 Matrix Lie groups

We begin by recalling a definition already discussed:

> **Definition 4.1.1.** The **general linear** group over the real numbers $\mathrm{GL}(n;\mathbb{R})$ is the group of $n \times n$ invertible matrices with real entries. The general linear group over the complex numbers $\mathrm{GL}(n;\mathbb{C})$ is the group of $n \times n$ matrices with complex entries.

We will denote the set of $n \times n$ complex matrices as $M_n(\mathbb{C})$. Note that $M_n(\mathbb{C})$ does not form a group under matrix multiplication because it has elements with no inverse. To define matrix Lie groups, we will need to define a notion of convergence in $M_n(\mathbb{C})$. A sequence of complex numbers $(z_0, z_1, z_2, \cdots)$ converges to z if for all $\epsilon > 0$ there exists a positive integer k such that $|z_i - z| < \epsilon$ for all $i \geq k$. This essentially tells us that z_i becomes arbitrarily close to z as i becomes large. We define convergence of complex matrices in the same way, element-by-element:

Definition 4.1.2. Let $(A_0, A_1, A_2, \cdots)$ be a sequence of complex matrices in $M_n(\mathbb{C})$. We say that this sequence **converges** to a matrix $A \in M_n(\mathbb{C})$ if each element of A_i converges to the corresponding element of A.

With this notion of convergence in hand we are now ready to define matrix Lie groups:

Definition 4.1.3. A **matrix Lie group** is a subgroup G of $\mathrm{GL}(n;\mathbb{C})$ for some n, with the property that if $(A_0, A_1, A_2, \cdots)$ is a sequence of matrices in G that converges to some $A \in M_n(\mathbb{C})$, then either $A \in G$ or A is not invertible.

The convergence property is a technical requirement necessary for certain general statements about matrix Lie groups to be true. In the language that will be introduced in the topology part of this course, the requirement for convergence is the statement that a matrix Lie group is a *closed* subgroup of $\mathrm{GL}(n;\mathbb{C})$.[1] Note that this is a different concept from closure under group multiplication.

4.2 Examples of matrix Lie groups

In this section we will go through some, but not all, of the most important examples of matrix Lie groups in physics.

4.2.1 General and special linear groups

Definition 4.1.3 makes $\mathrm{GL}(n;\mathbb{C})$ a matrix Lie group, since it is certainly a subgroup of itself, and if a sequence of matrices $A_i \in \mathrm{GL}(n;\mathbb{C})$ converges to some A, then there are two possibilities, either

(i) A is not invertible, or

(ii) A is invertible, in which case by definition $A \in \mathrm{GL}(n;\mathbb{C})$.

So $\mathrm{GL}(n;\mathbb{C})$ is closed. A similar argument shows that $\mathrm{GL}(n;\mathbb{R})$ — the subgroup of $\mathrm{GL}(n;\mathbb{C})$ consisting of real matrices — is also closed and thus $\mathrm{GL}(n;\mathbb{R})$ is a matrix Lie group.

Definition 4.2.1. The **special linear group** $\mathrm{SL}(n;\mathbb{C})$ is the subgroup of $\mathrm{GL}(n;\mathbb{C})$ consisting of $n \times n$ complex matrices with unit determinant. Similarly the special linear group $\mathrm{SL}(n;\mathbb{R})$ is the subgroup of $\mathrm{GL}(n;\mathbb{R})$ consisting of $n \times n$ real matrices with unit determinant. These are both matrix Lie groups.

To see that the special linear groups are matrix Lie groups, note that $\mathrm{SL}(n;\mathbb{C})$ is a subgroup of $\mathrm{GL}(n;\mathbb{C})$ since

(i) it is closed under multiplication since $\det(AB) = \det A \det B = 1$ if $A, B \in \mathrm{SL}(n;\mathbb{C})$,

[1] Compare this terminology to that for interval on the real line. An open interval (a, b) is the set of all numbers x such that $a < x < b$, while a closed interval $[a, b]$ is the set of all numbers x such that $a \leq x \leq b$. Since an open interval does not include its endpoints, there are sequences of numbers in an open interval that converge to a point outside the interval. For instance, the numbers in the sequence $(x_0, x_1, x_2, \cdots)$ with $x_i = 1/(i+1)$ all lie inside the open interval $(0, 2)$. However, the sequence converges to 0 as $m \to \infty$, which is outside the open interval $(0, 2)$ but inside the closed interval $[0, 2]$.

(ii) it contains the identity matrix, and

(iii) if $\det A = 1$ then $\det(A^{-1}) = 1/\det A = 1$, so every element of $\mathrm{SL}(n;\mathbb{C})$ has an inverse that is also in $\mathrm{SL}(n;\mathbb{C})$.

The subgroup is closed because if $(A_0, A_1, A_2, \cdots)$ is a sequence of matrices in $\mathrm{SL}(n;\mathbb{C})$ that converges to some matrix A, then the sequence of numbers $(\det A_0, \det A_1, \det A_2, \cdots)$ must converge to $\det A$. But this sequence is

$$(\det A_0, \det A_1, \det A_2, \cdots) = (1, 1, 1, \cdots), \tag{4.1}$$

which converges to 1. Hence $\det A = 1$ and thus $A \in \mathrm{SL}(n;\mathbb{C})$. All the same arguments apply for $\mathrm{SL}(n;\mathbb{R})$.

4.2.2 Unitary and orthogonal groups

If A and B are a unitary matrices then AB is also a unitary matrix, since $(AB)^\dagger AB = B^\dagger A^\dagger AB = B^\dagger B = \mathbb{1}$. The identity matrix is unitary and the inverse of a unitary matrix is unitary. Thus the set of $n \times n$ unitary matrices forms a subgroup of $\mathrm{GL}(n;\mathbb{C})$. The determinant of any unitary matrix A satisfies

$$|\det A|^2 = (\det A)(\det A)^* = (\det A)(\det A^\dagger) = \det\Big(AA^\dagger\Big) = \det \mathbb{1} = 1, \tag{4.2}$$

in other words the determinant of a unitary matrix is a pure phase. The $n \times n$ unitary matrices A with $\det A = 1$ form a subgroup of the group of $n \times n$ unitary matrices, since this property is preserved under multiplication.

Definition 4.2.2. For each positive integer n, the **unitary group** $\mathrm{U}(n)$ is the subgroup of $\mathrm{GL}(n;\mathbb{C})$ consisting of $n \times n$ unitary matrices. The **special unitary group** $\mathrm{SU}(n)$ is the subgroup of $\mathrm{U}(n)$ consisting of unitary matrices with unit determinant. These are both matrix Lie groups.

To see that $\mathrm{U}(n)$ is closed, and therefore a matrix Lie group, suppose that we have a sequence of matrices $(A_0, A_1, A_2, \cdots)$ in $\mathrm{U}(n)$ that converges to a matrix A. Taking the complex conjugate of every matrix in this sequence we obtain another sequence of matrices $(A_0^\dagger, A_1^\dagger, A_2^\dagger, \cdots)$ that converges to $A^\dagger$. We can then construct a third sequence of matrices $(A_0^\dagger A_0, A_1^\dagger A_1, A_2^\dagger A_2, \cdots)$ converging to $A^\dagger A$. Since $A_i \in \mathrm{U}(n)$ is unitary, we have that $A_i^\dagger A_i = \mathbb{1}$ for all i, so this sequence is

$$(A_0^\dagger A_0,\ A_1^\dagger A_1, A_2^\dagger A_2,\ \cdots) = (\mathbb{1}, \mathbb{1}, \mathbb{1}, \cdots), \tag{4.3}$$

which plainly converges to the identity and therefore $A^\dagger A = \mathbb{1}$. Thus $A \in \mathrm{U}(n)$, so $\mathrm{U}(n)$ is indeed closed. The subgroup $\mathrm{SU}(n)$ is also closed by the same logic that lead to the conclusion that $\mathrm{SL}(n;\mathbb{C})$ is closed.

Similar reasoning tells us that $n \times n$ orthogonal matrices form a closed subgroup of $\mathrm{GL}(n;\mathbb{C})$, thus forming a matrix Lie group. If A is an orthogonal matrix, then its determinant satisfies

$$(\det A)^2 = (\det A^T)(\det A) = \det\Big(A^T A\Big) = \det \mathbb{1} = 1, \tag{4.4}$$

and thus $\det A = \pm 1$. The $n \times n$ orthogonal matrices A with $\det A = 1$ form a subgroup of the group of $n \times n$ orthogonal matrices.

Definition 4.2.3.[2] For each positive integer n, the **orthogonal group** $\mathrm{O}(n)$ is the subgroup of $\mathrm{GL}(n;\mathbb{R})$ consisting of $n \times n$ real orthogonal matrices. The **special orthogonal group** $\mathrm{SO}(n)$ is the subgroup of $\mathrm{O}(n)$ consisting of orthogonal matrices with unit determinant. These are both matrix Lie groups.

The orthogonal group $\mathrm{O}(n)$ shows up in physics as the group of rotational and reflection symmetries in n dimensions. The rotations form the subgroup $\mathrm{SO}(n)$ while the elements of $\mathrm{O}(n)$ with negative determinant include reflections. Meanwhile $\mathrm{U}(n)$ and $\mathrm{SU}(n)$ show up for example in particle physics as gauge symmetries and as flavour symmetries. We will discuss the latter in detail in this course.

4.2.3 Generalised orthogonal group and the Lorentz group

The elements of $\mathrm{O}(n)$ are by definition the $n \times n$ matrices A satisfying

$$A^T \mathbb{1} A = \mathbb{1}. \tag{4.5}$$

We can define other matrix Lie groups by replacing the identity matrix in this equation with another diagonal matrix

$$g_{k,n-k} = \mathrm{diag}(\underbrace{-1,-1,\cdots,-1}_{k \text{ times}},\underbrace{1,1,\cdots 1}_{n-k \text{ times}}). \tag{4.6}$$

We will refer to such a matrix as a **metric**. The $n \times n$ real matrices A satisfying

$$A^T g_{k,n-k} A = g_{k,n-k} \tag{4.7}$$

for a fixed k form a closed subgroup of $\mathrm{GL}(n;\mathbb{C})$. We say that A **preserves** the metric $g_{k,n-k}$. To see that such matrices form a subgroup we note that the identity matrix satisfies equation (4.7) and that if two matrices A and B satisfy equation (4.7) then so does their product, since

$$(AB)^T g_{k,n-k}(AB) = B^T A^T g_{k,n-k} AB = B^T g_{k,n-k} B = g_{k,n-k}. \tag{4.8}$$

Finally, if A satisfies equation (4.7) then so does its inverse, which may be seen by taking the inverse of both sides of equation (4.7). That the group is a closed subgroup of $\mathrm{GL}(n;\mathbb{C})$ follows from similar logic as for unitary groups. Thus this group is a matrix Lie group. Computing the determinant of both sides of equation (4.7) tells us that elements of this group have $\det A = \pm 1$.

Definition 4.2.4. The set of $n \times n$ real matrices satisfying equation (4.7) for some positive integer k is the **generalised orthogonal group** $\mathrm{O}(k, n-k)$. The subgroup of $\mathrm{O}(k, n-k)$ consisting of matrices with unit determinant is denoted $\mathrm{SO}(k, n-k)$.

The most important case for physics is $k = 1$, since $g_{1,n-1} = \mathrm{diag}(-1, 1, 1, \cdots, 1)$ is the Minkowski metric η in n spacetime dimensions, and hence $\mathrm{O}(1, n-1)$ is the symmetry group of special relativity:

[2] The orthogonal and special orthogonal groups are defined to have real entries. One can also obtain matrix Lie groups by allowing the entries to be complex. The complex orthogonal group $\mathrm{O}(n;\mathbb{C})$ is the subgroup of $\mathrm{GL}(n;\mathbb{C})$ consisting of $n \times n$ complex orthogonal matrices, and $\mathrm{SO}(n;\mathbb{C})$ is the subgroup of $\mathrm{O}(n;\mathbb{C})$ consisting of matrices with unit determinant. However, the complex orthogonal groups don't have such clear applications to physics.

Definition 4.2.5. The **Lorentz group** $\mathrm{O}(1, n-1)$ is the group of $n \times n$ real matrices preserving the Minkowski metric. This means that a matrix Λ is part of the Lorentz group if and only if it satisfies

$$\Lambda^T \eta \Lambda = \eta, \tag{4.9}$$

where $\eta = \mathrm{diag}(-1, 1, 1, \cdots, 1)$ is the Minkowski metric. The subgroup $\mathrm{SO}(1, n-1)$ is called the **proper Lorentz group**.

4.2.4 Euclidean and Poincare groups

Definition 4.2.6. The **Euclidean group** $\mathrm{ISO}(n)$ is the group of isometries of the n-dimensional plane $\mathbb{R}^n$. It may be thought of as an $(n+1)$-dimensional matrix Lie group, in the sense described below.

Every element of $\mathrm{ISO}(n)$ may be expressed as a composition of a translation by some vector $\mathbf{a} \in \mathbb{R}^n$ and a rotation/reflection $R \in \mathrm{O}(n)$. We denote the corresponding element of $\mathrm{ISO}(n)$ as $\{\mathbf{a}, R\}$. Acting on a position vector $\mathbf{x}$ in the plane, we have

$$\{\mathbf{a}, R\}\mathbf{x} = R\mathbf{x} + \mathbf{a}. \tag{4.10}$$

We can determine the multiplication law by acting with two elements of $\mathrm{ISO}(n)$ in succession,

$$\{\mathbf{a}_1, R_1\}\{\mathbf{a}_2, R_2\}\mathbf{x} = \{a_1, R_1\}\left(R_2\mathbf{x} + \mathbf{a_2}\right) = R_1R_2\mathbf{x} + R_1\mathbf{a_2} + \mathbf{a_1}, \tag{4.11}$$

so we learn that the multiplication law is

$$\{\mathbf{a}_1, R_1\}\{\mathbf{a}_2, R_2\} = \{\mathbf{a_1} + R_1\mathbf{a_2}, R_1R_2\}. \tag{4.12}$$

The group $\mathrm{ISO}(n)$ is not a subgroup of $\mathrm{GL}(n;\mathbb{C})$, but is isomorphic to the closed subgroup of $\mathrm{GL}(n+1;\mathbb{C})$ consisting of matrices of the block form

$$\begin{pmatrix} R & \mathbf{a} \\ 0 & 1 \end{pmatrix}, \tag{4.13}$$

with $R \in \mathrm{O}(n)$ and $\mathbf{a} \in \mathbb{R}^n$. The 0 in the lower left stands for a row of n zeroes. The matrix form of $\{R, \mathbf{a}\}$ acts on $n+1$ component vectors of the form $(\mathbf{x}, 1)^T$,

$$\begin{pmatrix} R & \mathbf{a} \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \mathbf{x} \\ 1 \end{pmatrix} = \begin{pmatrix} R\mathbf{x} + \mathbf{a} \\ 1 \end{pmatrix}. \tag{4.14}$$

Multiplying two matrices of the form in equation (4.13), we find

$$\begin{pmatrix} R_1 & \mathbf{a}_1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} R_2 & \mathbf{a}_2 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} R_1R_2 & \mathbf{a}_1 + R_1\mathbf{a_2} \\ 0 & 1 \end{pmatrix}, \tag{4.15}$$

which is the expected multiplication law for $\mathrm{ISO}(n)$.

Replacing $\mathrm{O}(n)$ with $\mathrm{O}(1, n-1)$ in the above discussion we obtain the Poincaré group $\mathrm{ISO}(1, n-1)$, the group of isometries of n-dimensional Minkowski space. Hence, the Poincaré group is isomorphic to the closed subgroup of $\mathrm{GL}(n+1;\mathbb{C})$ consisting of matrices of the form

$$\begin{pmatrix} \Lambda & \mathbf{a} \\ 0 & 1 \end{pmatrix}, \tag{4.16}$$

with $\Lambda \in \mathrm{O}(1, n-1)$ a Lorentz transformation and $\mathbf{a} \in \mathbb{R}^n$ a spacetime translation.

4.3 Connectedness

Connectedness is an important topological property of Lie groups:[3]

Definition 4.3.1. Two elements A and B of a matrix Lie group G are **connected** if there exists a continuous path $\gamma(t) \in G$ for $0 \leq t \leq 1$, such that $\gamma(0) = A$ and $\gamma(1) = B$. The set of all elements connected to the identity is called the **identity component** of G, denoted G_0. If every element of G is connected to every other element then we say that G is **connected**.

Theorem 4.3.1. *Let G be a matrix Lie group. The identity component G_0 is a subgroup of G.*

Proof. The identity component always includes the identity element, which is connected to itself by the constant path $\gamma(t) = \mathbb{1}$. We need to verify that G_0 fulfils the other requirements of subgroup: that it is closed under multiplication and that every element of G_0 has an inverse also in G_0.

Closure under group multiplication: Suppose A and B are elements of G_0 connected to the identity by continuous paths $\gamma_A(t)$ and $\gamma_B(t)$. Then $\gamma_A(t)\gamma_B(t)$ is a continuous path that connects the identity to AB, so $AB \in G_0$.

Inverse: Let A be an arbitrary element of G_0 and $\gamma(t)$ the path that connects the identity to A. Then $\gamma(t)^{-1}$ is a continuous path connecting the identity to A^{-1}, and therefore $A^{-1} \in G_0$. □

Example 4.3.1. The special orthogonal group SO(2) is connected. To see this, note that an arbitrary element $A \in \mathrm{SO}(2)$ may be written as

$$A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \tag{4.17}$$

for some real θ. Then A is connected to the identity by the path

$$\gamma_A(t) = \begin{pmatrix} \cos(t\theta) & -\sin(t\theta) \\ \sin(t\theta) & \cos(t\theta) \end{pmatrix}. \tag{4.18}$$

Thus any two elements A and B are connected by a continuous path that runs from A through the identity and finishes at B.

On the other hand O(2) is not connected since matrices in O(2) have determinant ± 1. If A and B are elements of O(2) with $\det A = 1$ and $\det B = -1$, then there is no continuous path connecting A and B, since the determinant would have to jump from 1 to -1 somewhere along the path. The identity component of O(2) consists of all the elements of O(2) with determinant $+1$. In other words the identity component of O(2) is SO(2). There is another connected component of O(2) containing all the orthgonal

[3] What we have defined here is technically called "path connectedness", with "connectedness" usually defined more abstractly. It turns out that both concepts are equivalent for matrix Lie groups.

matrices with negative determinant. This second component of $O(2)$ is not a subgroup since it does not contain the identity element.

Chapter 5

The matrix exponential

In the study of Lie groups and Lie algebras we will often need to take exponentials of matrices. In this chapter we will provide a quick review of the matrix exponential and its key properties.

5.1 Definition

Definition 5.1.1. Given a matrix X, its exponential is defined by the infinite series

$$e^X = \sum_{n=0}^{\infty} \frac{X^n}{n!}, \tag{5.1}$$

where $X^0 = \mathbb{1}$. This series converges for any matrix, just as the the power series for e^x converges for any real or complex number x. An equivalent definition of the matrix exponential that is sometimes useful is

$$e^X = \lim_{m\to\infty} \left(1 + \frac{X}{m}\right)^m. \tag{5.2}$$

We also use the notation $\exp(X) = e^X$.

Equivalence of the power series and the limit formula for the matrix exponential are proven in the same way as for the exponentials of real or complex numbers, for example expanding equation (5.2) for finite m using the binomial theorem, rearranging the sum, and taking the $m \to \infty$ limit.

5.2 Properties

In this section we will derive some of the basic properties of the matrix exponential.

Theorem 5.2.1. *For any matrixes X and Y, the matrix exponential satisfies:*

(i) $e^0 = \mathbb{1}$;

(ii) $(e^X)^* = e^{X^*}$, $(e^X)^T = e^{X^T}$, *and* $(e^X)^\dagger = (e^{X^\dagger})$;

(iii) if C is an invertible matrix then $e^{CXC^{-1}} = Ce^X C^{-1}$;

(iv) If X and Y commute, then $e^X e^Y = e^{X+Y}$; and

(v) $(e^X)^{-1} = e^{-X}$.

Proof. Points (i), (ii), and (iii) follow immediately from the power series definition of the matrix exponential. For point (ii) we use that $(X^n)^* = (X^*)^n$, and similar for the transpose and Hermitian conjugate. Similarly, for point (iii) we use that $CX^nC^{-1} = (CXC^{-1})^n$.

To prove point (iv) we multiply the power series of e^X and e^Y,

$$e^X e^Y = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\frac{X^m Y^n}{m!\,n!} = \sum_{k=0}^{\infty}\sum_{m=0}^{k}\frac{X^m Y^{k-m}}{m!\,(k-m)!}, \tag{5.3}$$

where the second equality is obtained by replacing n with a new summation variable $k = m + n$. We can rearrange the right-hand side to obtain

$$e^X e^Y = \sum_{k=0}^{\infty}\frac{1}{k!}\sum_{m=0}^{k}\frac{k!}{m!\,(k-m)!}X^m Y^{k-m}. \tag{5.4}$$

If X and Y commute then the sum over m on the right-hand side is the binomial expansion of $(X+Y)^k$, so we finally obtain

$$e^X e^Y = \sum_{k=0}^{\infty}\frac{(X+Y)^k}{k!} = e^{X+Y}, \tag{5.5}$$

establishing point (iv).

Finally, point (v) follows from points (i) and (iv). Since X and $-X$ commute,

$$e^X e^{-X} = e^{X-X} = e^0 = \mathbb{1}. \tag{5.6}$$

Therefore $e^{-X} = (e^X)^{-1}$. □

Theorem 5.2.2. *Let X be a matrix and t be a real number. Then*

$$\frac{\mathrm{d}}{\mathrm{d}t}e^{tX} = Xe^{tX} = e^{tX}X. \tag{5.7}$$

In particular,

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}e^{tX}\right|_{t=0} = X. \tag{5.8}$$

Proof. Differentiate the power series of e^{tX} term by term. Note that $Xe^{tX} = e^{tX}X$ since X commutes with itself. □

When the matrices X and Y *do not* commute, it is no longer true that $e^X e^Y$ is the same as e^{X+Y}. As a special case suppose that X and Y do not commute, i.e. $[X, Y] \neq 0$, but that they each commute with their commutator,

$$[X, [X, Y]] = [Y, [X, Y]] = 0. \tag{5.9}$$

The product of e^X and e^Y then takes a relatively simple form:

Theorem 5.2.3. *Given matrices X and Y satisfying equation (5.9), the product of their exponentials is*

$$e^X e^Y = e^{X+Y+\frac{1}{2}[X,Y]}. \tag{5.10}$$

Proof. Let us define

$$A(t) = e^{tX}e^{tY}, \qquad B(t) = e^{t(X+Y)+\frac{t^2}{2}[X,Y]} \tag{5.11}$$

for $t \in \mathbb{R}$. The theorem claims that $A(1) = B(1)$. We will prove this by showing that $A(t)$ and $B(t)$ obey the same differential equation with the same initial condition at $t = 0$.

First, differentiating $A(t)$ we find

$$A'(t) = Xe^{tX}e^{tY} + e^{tX}Ye^{tY}. \tag{5.12}$$

It is straightforward to prove by induction that when X commutes with $[X, Y]$, then $X^nY = YX^n + n[X, Y]X^{n-1}$, using which we can rearrange

$$\begin{aligned} e^{tX}Y &= \sum_{n=0}^{\infty} \frac{1}{n!} t^n X^n Y \\ &= Y \sum_{n=0}^{\infty} \frac{1}{n!} t^n X^n + [X, Y] \sum_{n=1}^{\infty} \frac{1}{(n-1)!} t^n X^{n-1} \\ &= (Y + t[X, Y])\, e^{tX} \end{aligned} \tag{5.13}$$

Hence

$$A'(t) = (X + Y + t[X, Y])\, e^{tX}e^{tY} = (X + Y + t[X, Y])\, A(t). \tag{5.14}$$

Meanwhile, differentiating $B(t)$ we find it obeys the same differential equation,

$$B'(t) = (X + Y + t[X, Y])\, B(t). \tag{5.15}$$

Setting $t = 0$ in equation (5.11) we find that $A(0) = B(0) = \mathbb{1}$, so $A(t)$ and $B(t)$ obey the same differential equation with the same initial condition, and are thus equal. Setting $t = 1$ we recover the result in the theorem. □

Equation (5.10) is a special case of the Baker–Campbell–Hausdorff (BCH) formula, which in general gives an expression for $e^X e^Y$ with no assumptions about the commutators. The full BCH formula is too involved to get into in this course. All we need to know is that writing $e^X e^Y = e^Z$ for some Z, the BCH formula expresses Z in terms of nested commutators, the first few terms are

$$Z = X + Y + \frac{1}{2}[X, Y] + \frac{1}{12}[X, [X, Y]] - \frac{1}{12}[Y, [X, Y]] + \cdots . \tag{5.16}$$

The BCH formula only holds for X and Y such that this series converges. In our applications of it, X and Y will be proportional to some parameter ϵ that we can choose to be small enough that the series will converge and the BCH formula will hold.

5.3 Computing matrix exponentials

The infinite series expression for the matrix exponential is useful for proving various general properties, but for explicit calculations it would be nice to have a way of finding closed form expressions for the exponentials of matrices. In this section we will discuss techniques for finding such expressions.

We begin with a very important example that will arise later,

Example 5.3.1. Consider the exponential $e^{i\theta\sigma_1/2}$, where θ is a complex number and σ_1 is the first Pauli matrix. The square of σ_1 is the two-dimensional identity matrix,

$$(\sigma_1)^2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbb{1}. \tag{5.17}$$

This then tells us that $(\sigma_1)^n = \sigma_1$ when n is odd and $(\sigma_1)^n = \mathbb{1}$ when n is even. The power series definition of the matrix exponential then gives

$$\begin{aligned} e^{i\theta\sigma_1/2} &= \mathbb{1} \sum_{n \text{ even}} \frac{(i\theta/2)^n}{n!} + \sigma_1 \sum_{n \text{ odd}} \frac{(i\theta/2)^n}{n!} \\ &= \mathbb{1} \sum_{k=0}^{\infty} \frac{(-1)^k (\theta/2)^{2k}}{(2k)!} + i\sigma_1 \sum_{k=0}^{\infty} \frac{(-1)^k (\theta/2)^{2k+1}}{(2k+1)!} \end{aligned} \tag{5.18}$$

The two power series on the second line are the power series for the trigonometric functions $\cos(\theta/2)$ and $\sin(\theta/2)$, so we have

$$e^{i\theta\sigma_1/2} = \mathbb{1}\cos(\theta/2) + i\sigma_1 \sin(\theta/2) = \begin{pmatrix} \cos(\theta/2) & i\sin(\theta/2) \\ i\sin(\theta/2) & \cos(\theta/2) \end{pmatrix}. \tag{5.19}$$

Similarly, since $(\sigma_2)^2 = (\sigma_3)^2 = 1$ we have that

$$e^{i\theta\sigma_2/2} = \mathbb{1}\cos(\theta/2) + i\sigma_2 \sin(\theta/2) \tag{5.20}$$

The previous example relied heavily on the special property that $(\sigma_1)^2 = \mathbb{1}$. What about more general matrices? If $D = \text{diag}(D_{11}, D_{22}, D_{33}, \cdots)$ is a diagonal matrix, then its exponential is trivial to compute, it is the diagonal matrix

$$e^D = \begin{pmatrix} e^{D_{11}} & 0 & 0 & \cdots \\ 0 & e^{D_{22}} & 0 & \cdots \\ 0 & 0 & e^{D_{33}} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \tag{5.21}$$

Then, if X is a diagonalisable matrix, $X = CDC^{-1}$ for some invertible matrix C and some diagonal matrix D, then we can use point (iii) of theorem 5.2.1 to find that $e^X = Ce^DC^{-1}$, giving us a way to compute the exponential of any diagonalisable matrix.

Example 5.3.2. Consider the matrix

$$X = \begin{pmatrix} 1 & -2i \\ 2i & 1 \end{pmatrix}. \tag{5.22}$$

Since X is Hermitian, it is diagonalisable by a unitary matrix C with columns consisting of the normalised eigenvectors of X. Concretely

$$X = CDC^{-1}, \qquad C = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}, \qquad D = \begin{pmatrix} 3 & 0 \\ 0 & -1 \end{pmatrix}. \tag{5.23}$$

We then have

$$e^X = Ce^D C^{-1} = \frac{1}{2} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix} \begin{pmatrix} e^3 & 0 \\ 0 & e^{-1} \end{pmatrix} \begin{pmatrix} 1 & -i \\ -i & 1 \end{pmatrix} = \begin{pmatrix} e\cosh 2 & -ie\sinh 2 \\ ie\sinh 2 & e\cosh 2 \end{pmatrix}. \tag{5.24}$$

Many of the matrices we encounter in physics are diagonalisable — for example Hermitian and unitary matrices are diagonalisable — so the above method is widely applicable for the computation of matrix exponentials. However, sometimes we encounter a matrix that is not diagonalisable. Here is an important class:[1]

Definition 5.3.1. Let k be a positive integer. A **nilpotent matrix of degree** k is a matrix N such that $N^k = 0$.

The power series for the exponential of nilpotent matrix terminates after a finite number of terms, so is straightforward to compute.

Example 5.3.3. The matrix

$$N = \begin{pmatrix} 2 & -1 \\ 4 & -2 \end{pmatrix} \tag{5.25}$$

is nilpotent of degree two, since $N^2 = 0$. Therefore

$$e^N = \mathbb{1} + N = \begin{pmatrix} 3 & -1 \\ 4 & -1 \end{pmatrix} \tag{5.26}$$

Not every matrix is diagonalisable or nilpotent. However, there is a result of linear algebra that an arbitrary matrix X may be decomposed as

$$X = S + N, \tag{5.27}$$

where S is diagonalisable, N is nilpotent, and $[S, N] = 0$. Importantly for us, since S and N commute, point (iv) implies that $e^X = e^S e^N$. Since we know how to compute the exponentials of diagonalisable and nilpotent matrices this provides a way of computing e^X for any matrix X. In practice this method is quite involved and since we won't need

[1] To see that a nilpotent matrix is not diagonalisable, assume toward a contradiction that N is a nilpotent matrix of degree k and $N = CDC^{-1}$ for some diagonal matrix D. Then $0 = N^k = CD^kC^{-1}$, which implies that $D^k = 0$. But the diagonal elements of D^k are just the diagonal elements of D raised to the kth power, so $D^k = 0$ implies that $D = 0$, further implying that $N = 0$.

to compute any matrix exponentials using this method, we won't comment further on it. The key point though is that it means that you can use a computer to compute the exponential of any matrix you like in closed form.

5.4 The determinant of the matrix exponential

Here is a result that is used very often in theoretical physics:

Theorem 5.4.1. *The matrix exponential satisfies*

$$\det\left(e^X\right) = e^{\operatorname{tr} X} \tag{5.28}$$

for any matrix X.

Proof. We use the result from linear algebra that any matrix X is similar to an upper triangular matrix J, $X = PJP^{-1}$ for some invertible matrix P. The matrix J has the eigenvalues λ_i of X on the diagonal,

$$J = \begin{pmatrix} \lambda_1 & J_{12} & \cdots & J_{1,n} \\ 0 & \lambda_2 & \cdots & J_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}. \tag{5.29}$$

Taking powers of J we see that J^m is also upper triangular, with λ_i^m on the diagonal. Thus e^J is an upper triangular matrix with e^{λ_i} on the diagonal. We then have

$$\det e^A = \det\left(Pe^J P^{-1}\right) = \det e^J = \prod_{i=1}^{n} e^{\lambda_i} = \exp\left(\sum_{i=1}^{n} \lambda_i\right) = e^{\operatorname{tr} X}. \tag{5.30}$$

where the second equality uses point (iii) of theorem 5.2.1. □

Chapter 6

Lie algebras

6.1 Definitions

A Lie algebra is a vector space equipped with a way of "multiplying" two vectors to get another vector. As for group multiplication, we can define representations of the Lie algebra as matrices obeying the same algebra. We will see in this section that every Lie group is associated to a Lie algebra and that the representations of a Lie group are related to the representations of the associated Lie algebra. In the remainder of this part of the course will discuss systematic ways of constructing the irreducible representations of a certain class of Lie algebras, which will then give us irreducible representations of the corresponding Lie groups.

Definition 6.1.1. A finite-dimensional **Lie algebra** is a finite-dimensional vector space $\mathfrak{g}$, together with a map $\{\cdot,\cdot\} : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$ called the **Lie bracket** satisfying the following properties:

(i) Bilinearity, meaning that if X, Y, and Z are vectors and a and b are scalars,

$$\{X, aY + bZ\} = a\{X, Y\} + b\{X, Z\}, \qquad \{aX + bY, Z\} = a\{X, Z\} + b\{Y, Z\}.$$

(ii) Antisymmetry,

$$\{X, Y\} = -\{Y, X\}.$$

(iii) The Jacobi identity,

$$\{X, \{Y, Z\}\} + \{Y, \{Z, X\}\} + \{Z, \{X, Y\}\} = 0.$$

Two vectors X and Y are said to **commute** if $\{X, Y\} = 0$. If every vector commutes with every other vector, then the Lie algebra is **abelian**. The **dimension** of the Lie algebra is the dimension of the vector space $\mathfrak{g}$.

Example 6.1.1. The space $\mathbb{R}^3$ of real, three-dimensional vectors forms a real Lie algebra with bracket given by the vector product $\{\mathbf{u}, \mathbf{v}\} = \mathbf{u} \times \mathbf{v}$. Bilinearity and antisymmetry

are standard properties of the vector product. To show that the Jacobi identity is satisfied we use the vector triple product formula

$$\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) = (\mathbf{u} \cdot \mathbf{w})\mathbf{v} - (\mathbf{u} \cdot \mathbf{v})\mathbf{w}. \tag{6.1}$$

Using this, we can evaluate the left-hand side of the Jacobi identity.

$$\begin{aligned} &\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) + \mathbf{v} \times (\mathbf{w} \times \mathbf{u}) + \mathbf{w} \times (\mathbf{u} \times \mathbf{v}) \\ &\qquad = (\mathbf{u} \cdot \mathbf{w})\mathbf{v} - (\mathbf{u} \cdot \mathbf{v})\mathbf{w} + (\mathbf{v} \cdot \mathbf{u})\mathbf{w} - (\mathbf{v} \cdot \mathbf{w})\mathbf{u} + (\mathbf{w} \cdot \mathbf{v})\mathbf{u} - (\mathbf{w} \cdot \mathbf{u})\mathbf{v} \\ &\qquad = 0. \end{aligned} \tag{6.2}$$

The Lie algebra associated to a Lie group arises when we consider group elements close to the identity. Consider a continuous path $\gamma(t)$ through a matrix Lie group G passing through the identity at $t = 0$, $\gamma(0) = \mathbb{1}$. This is a continuous family of matrices. We will refer to such a path as **a path through the identity**. We define a new matrix X as the derivative of $\gamma(t)$ at $t = 0$, multiplied by $-i$,[1]

$$X = -i\gamma'(0). \tag{6.3}$$

We define a set of matrices $\mathfrak{g}$ as all the possible matrices that can be obtained from equation (6.3) by considering all possible curves through G that pass through the identity at $t = 0$. It turns out that $\mathfrak{g}$ satisfies all the defining properties of a Lie algebra given definition 6.1.1, with the Lie bracket proportional to the usual commutator of matrices:

Theorem 6.1.1. *The set $\mathfrak{g}$ of all possible X obtained from equation* (6.3) *forms a Lie algebra. The Lie bracket is i times the matrix commutator,*

$$\{X, Y\} = i[X, Y] \equiv i(XY - YX). \tag{6.4}$$

Before proving this theorem, we prove a useful result:

Theorem 6.1.2. *If $X \in \mathfrak{g}$ and $A \in G$, then $AXA^{-1} \in \mathfrak{g}$.*

Proof. Let $\gamma(t)$ be a path in G such that X is obtained as in equation (6.3). Then $\tilde{\gamma}(t) = A\gamma(t)A^{-1}$ is another path through the identity in G. The element of $\mathfrak{g}$ defined by this path is

$$-i\tilde{\gamma}'(0) = A(-i\gamma'(0))A^{-1} = AXA^{-1}, \tag{6.5}$$

so $AXA^{-1} \in \mathfrak{g}$. □

Proof of theorem 6.1.1. We need to prove that all the properties in definition 6.1.1 hold. First we show that $\mathfrak{g}$ is a vector space, with addition of vectors given by usual matrix addition, by showing that (*i*) $\mathfrak{g}$ includes the zero matrix (the additive identity in the vector space); (*ii*) that each element of $\mathfrak{g}$ has an additive inverse also in $\mathfrak{g}$; and (*iii*) that $\mathfrak{g}$ is closed under multiplication by scalars and addition of vectors:

[1] The factor of $-i$ in equation (6.3) is the convention commonly used in physics. In mathematics texts the equivalents of equation (6.3) and (6.4) are $X = \gamma'(0)$ and $\{X, Y\} = [X, Y]$. The extra factors of i that we use in physics may seem annoying, but they turn out the be particularly convenient for physical applications in quantum mechanics.

(i) The zero matrix is obtained from the constant curve $\gamma = \mathbb{1}$.

(ii) If $X \in \mathfrak{g}$ is obtained from a path through the identity $\gamma(t)$, then $-X$ is obtained from the path $\gamma(t)^{-1}$ where for every t, $\gamma(t)^{-1}$ is the inverse of the matrix $\gamma(t)$,

$$-i \left.\frac{\mathrm{d}}{\mathrm{d}t}\gamma(t)^{-1}\right|_{t=0} = i\,\gamma(t)^{-1}\gamma'(t)\gamma(t)^{-1}\Big|_{t=0} = -X, \tag{6.6}$$

where we have used the usual formula for the derivative of the inverse of a matrix.[2]

(iii) Let $\gamma_1(t)$ and $\gamma_2(t)$ be paths through the identity, $X = -i\gamma_1'(0)$ and $Y = -i\gamma_2'(0)$ the corresponding elements of $\mathfrak{g}$, and $a, b \in \mathbb{R}$. We define a new path through the identity $\gamma_3(t) = \gamma_1(at)\gamma_2(bt)$. The element of $\mathfrak{g}$ corresponding to this path is

$$-i\gamma_3'(0) = -ia\gamma_1'(0) - ib\gamma_2'(0) = aX + bY, \tag{6.7}$$

which shows that $aX + bY \in \mathfrak{g}$.

Thus $\mathfrak{g}$ is indeed a vector space.

To show that this vector space forms a Lie algebra, we need to show that $\mathfrak{g}$ is closed under the Lie bracket. Suppose that $X, Y \in \mathfrak{g}$ and that $\gamma(t)$ is a path through the identity in G satisfying and $X = -i\gamma'(0)$. Theorem 6.1.2 tells us that $\gamma(t)Y\gamma(t)^{-1} \in \mathfrak{g}$ for all t. We have already seen that $\mathfrak{g}$ is closed under addition of vectors and multiplication by scalars, so we then have that

$$\frac{\gamma(t)Y\gamma(t)^{-1} - \gamma(0)Y\gamma(0)^{-1}}{t} \in \mathfrak{g}. \tag{6.8}$$

Taking the limit $t \to 0$, this becomes the formula for the derivative of $\gamma(t)Y\gamma(t)^{-1}$ with respect to t at $t = 0$. But

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}(\gamma(t)Y\gamma(t)^{-1})\right|_{t=0} = i(XY - YX) = i[X, Y], \tag{6.9}$$

so we have that $\{X, Y\} = i[X, Y] \in \mathfrak{g}$, meaning that $\mathfrak{g}$ is closed under the Lie bracket. Finally, the matrix commutator satisfies the properties (i)–(iii) in definition 6.1.1 (you will check the Jacobi identity in an exercise). Thus $\mathfrak{g}$ is indeed a Lie algebra. □

Equation (6.3) tells us that for sufficiently small $t = \epsilon$ we can Taylor expand a path $\gamma(\epsilon)$ about the identity to find

$$\gamma(\epsilon) = 1 + i\epsilon X + O(\epsilon^2). \tag{6.10}$$

Let's take $\epsilon = 1/m$ for some very large positive integer m. Since $\gamma(1/m) \in G$, closure of the group multiplication implies that $\gamma(1/m)^m \in G$. From equation (6.10) we have that

$$\gamma(1/m)^m = \left(1 + \frac{iX}{m} + O(m^{-2})\right)^m = e^{iX} + O(m^{-1}), \tag{6.11}$$

[2]The formula for the derivative of the inverse of a matrix γ follows from

$$0 = \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{1} = \frac{\mathrm{d}}{\mathrm{d}t}(\gamma^{-1}\gamma) = \frac{\mathrm{d}\gamma^{-1}}{\mathrm{d}t}\gamma + \gamma^{-1}\frac{\mathrm{d}\gamma}{\mathrm{d}t}.$$

Multiplying from the right by γ^{-1} and rearranging, we obtain

$$\frac{\mathrm{d}\gamma^{-1}}{\mathrm{d}t} = -\gamma^{-1}\frac{\mathrm{d}\gamma}{\mathrm{d}t}\gamma^{-1}.$$

where the final equality follows from the limit formula for the matrix exponential. The sequence of matrices $\gamma(1/m)^m \in G$ converges to e^{iX} as $m \to \infty$. By the closure property of matrix Lie groups, this means that either $e^{iX} \in G$ or e^{iX} is not invertible. But e^{iX} is always invertible; its inverse is e^{-iX}. We conclude that for every $X \in \mathfrak{g}$ it must be that $e^{iX} \in G$. Moreover, if $X \in \mathfrak{g}$ then closure of the vector space $\mathfrak{g}$ under multiplication by scalars implies that $tX \in \mathfrak{g}$ for any $t \in \mathbb{R}$. We thus conclude that $e^{itX} \in G$ for any real number t.

Conversely, suppose that there is a matrix X such that $e^{itX} \in G$ for any $t \in \mathbb{R}$. Then e^{itX} defines a continuous path in G, and differentiating at $t = 0$ we find that $X \in \mathfrak{g}$. Thus we conclude that $X \in \mathfrak{g}$ if and only if $e^{itX} \in G$ for all $t \in \mathbb{R}$. We sum all this up in the following definition:

Definition 6.1.2. Let G be a matrix Lie group. The **Lie algebra** of G, denoted $\mathfrak{g}$ is the set of all matrices X that can be written as in equation (6.3) for some path through the identity. Equivalently, $\mathfrak{g}$ is the set of all matrices X such that $e^{itX} \in G$ for all real numbers t. The Lie bracket is given by i times the commutator of the matrices in $\mathfrak{g}$.

From now on, we will assume that every Lie algebra we discuss is the Lie algebra of a matrix Lie group. We will always use i times the matrix commutator $[\cdot, \cdot]$ in place of the Lie bracket.

Note that it is not necessarily true that every element of a matrix Lie group can be written as e^{iX} for some X in the Lie algebra. However, the following theorem tells us that we can at least in principle reconstruct the whole of the identity component of a Lie group from its Lie algebra,

Theorem 6.1.3. *If G is a matrix Lie group, every element A of the identity component G_0 may be written as*

$$A = e^{iX_1} e^{iX_2} \cdots e^{iX_m}, \tag{6.12}$$

for some set of $X_1, X_2, \cdots X_m$ in $\mathfrak{g}$.

We will not make direct use of this theorem, and since its proof is quite involved we will not give it. If you are interested, it is given as corollary 3.46 in Hall's textbook [2]. Notice that we cannot hope to reconstruct elements of the group outside the identity component from the Lie algebra, since if A is a group element as in equation (6.12), then $e^{itX_1} e^{itX_2} \cdots e^{itX_m}$ is a continuous path connecting the identity to A.

The reason why matrix commutators are important is the BCH formula. If X and Y are Lie algebra elements, then if we multiply group elements e^{iX} and e^{iY} together the BCH formula tells us that the result will be a group element e^{iZ} with Z given by a sum over nested commutators of X and Y. The Lie bracket/matrix commutator thus encodes the group multiplication.

We now determine the Lie algebras of some important Lie groups. In all of the examples below, X denotes an $n \times n$ matrix for some positive integer n, while t is a real number.

Example 6.1.2 (Lie algebras of general and special linear groups)**.** Since the exponential of any matrix is invertible, e^{itX} is an invertible matrix for any X. Thus the Lie algebra $\mathfrak{gl}(n,\mathbb{C})$ of $\mathrm{GL}(n;\mathbb{C})$ is $M_n(\mathbb{C})$, the set of all $n \times n$ complex matrices.

If e^{itX} belongs to $\mathrm{GL}(n;\mathbb{R})$ then it must be a real matrix. For this to be true for all t all the elements of X must be purely imaginary. Therefore the Lie algebra $\mathfrak{gl}(n;\mathbb{R})$ of $\mathrm{GL}(n;\mathbb{R})$ consists of all $n \times n$ purely imaginary matrices.

If e^{itX} belongs to $\mathrm{SL}(n;\mathbb{C})$ or $\mathrm{SL}(n;\mathbb{R})$ then it must have unit determinant,

$$1 = \det\left(e^{itX}\right) = e^{it\,\mathrm{tr}\,X}, \tag{6.13}$$

using theorem 5.4.1, simplying that $\mathrm{tr}\,X = 0$. Therefore the Lie algebra $\mathfrak{sl}(n;\mathbb{C})$ of $\mathrm{SL}(n;\mathbb{C})$ consists of all $n \times n$ traceless matrices, while the Lie algebra $\mathfrak{sl}(n;\mathbb{R})$ of $\mathrm{SL}(n;\mathbb{R})$ consists of all $n \times n$ imaginary, traceless matrices.

Example 6.1.3 (Lie algebras of unitary groups)**.** For e^{itX} to belong to $\mathrm{U}(n)$ we must by definition have $(e^{itX})^{-1} = (e^{itX})^\dagger$. Using the properties of the matrix exponential in theorem 5.2.1 we have that

$$(e^{itX})^{-1} = e^{-itX}, \qquad (e^{itX})^\dagger = e^{-itX^\dagger}. \tag{6.14}$$

These to be equal for all t if and only if $X = X^\dagger$. Therefore the Lie algebra $\mathfrak{u}(n)$ of $\mathrm{U}(n)$ consists of all $n \times n$ Hermitian matrices.

The real dimension of $\mathfrak{u}(n)$ as a vector space is the number of linearly independent $n \times n$ Hermitian matrices, or equivalently the number of real parameters required to specify such a matrix. To do so we would need to specify the n real diagonal elements and the $n(n-1)/2$ complex elements above the diagonal, giving a dimension

$$\dim \mathfrak{u}(n) = n + 2 \times \frac{1}{2}n(n-1) = n^2. \tag{6.15}$$

For $\mathrm{SU}(n)$ we also have that $\det\left(e^{itX}\right) = 1$, implying that $\mathrm{tr}\,X = 0$ using theorem 5.4.1. Therefore the Lie algebra $\mathfrak{su}(n)$ of $\mathrm{SU}(n)$ consists of all $n \times n$ traceless Hermitian matrices. Tracelessness imposes a single linear relation on the elements of matrices in $\mathfrak{su}(n)$, reducing the dimension of $\mathfrak{su}(n)$ by one compared to the dimension of $\mathfrak{u}(n)$,

$$\dim \mathfrak{su}(n) = n^2 - 1. \tag{6.16}$$

Example 6.1.4 (Lie algebras of orthogonal groups)**.** For e^{itX} to belong to $\mathrm{O}(n)$ we must have that $(e^{itX})^{-1} = (e^{itX})^T$ for all t. Using theorem 5.2.1 we find that this becomes $e^{-itX} = e^{itX^T}$ and therefore we must have $X = -X^T$, i.e. X is an antisymmetric matrix. Since the elements of $\mathrm{O}(n)$ are real we must have that the elements X are purely imaginary. Therefore the Lie algebra $\mathfrak{o}(n)$ of $\mathrm{O}(n)$ consists of all $n \times n$ imaginary, antisymmetric matrices. You will compute the dimension of this Lie algebra in an exercise.

An antisymmetric matrix has only zeroes on the diagonal so is automatically traceless. Thus if $X \in \mathfrak{o}(n)$ then $\det\left(e^{itX}\right) = 1$ and so $X \in \mathfrak{so}(n)$, the Lie algebra of $\mathrm{SO}(n)$. Therefore the Lie algebras of $\mathrm{O}(n)$ and $\mathrm{SO}(n)$ are the same.

6.2 Generators and structure constants

Since a Lie algebra is a vector space it is often convenient to use a basis:

Definition 6.2.1. The basis elements are called the **generators** of the Lie algebra and are often denoted by T_a, where the index a runs from one to whatever the dimension of the Lie algebra is.

Consider the commutator $[T_a, T_b]$ of two generators in a Lie algebra $\mathfrak{g}$. The result must be i times an element of $\mathfrak{g}$, so it must be a linear combination of generators since the generators are a basis for $\mathfrak{g}$,

$$[T_a, T_b] = i f_{abc} T_c. \tag{6.17}$$

We will always use the Einstein summation notation on the Lie algebra indices, so the index c is summed over.

Definition 6.2.2. The numbers f_{abc} in equation (6.17) are the **structure constants** of the Lie algebra. Note that the structure constants depend on the choice of generators T_a.

The commutator of any two elements $X, Y \in \mathfrak{g}$ is determined by the structure constants. Decomposing X and Y in the basis of the generators as $X = x_a T_a$ and $Y = y_a T_a$ for some set of coefficients x_a and y_a, using the bilinearity of the commutator we have

$$[X, Y] = i x_a y_b f_{abc} T_c. \tag{6.18}$$

Since the commutator is antisymmetric, $[T_a, T_b] = -[T_b, T_a]$, the structure constants must be antisymmetric in their first two indices

$$f_{abc} = -f_{bac}. \tag{6.19}$$

Applying the Jacobi identity to T_a, T_b, and T_c, we find a another relation that the structure constants must satisfy,

$$f_{ade} f_{bcd} + f_{cad} f_{bde} + f_{abd} f_{cde} = 0. \tag{6.20}$$

Example 6.2.1. The Lie algebra $\mathfrak{so}(3)$ consists of all 3×3 purely imaginary antisymmetric matrices. These are specified by the three elements above the diagonal, so $\mathfrak{so}(3)$ is a three-dimensional Lie algebra. A convenient basis is $T_c = -i\epsilon_{abc}$, explicitly

$$T_1 = i\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \qquad T_2 = i\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \qquad T_3 = i\begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{6.21}$$

Computing commutators of these matrices, we find

$$[T_a, T_b] = i\epsilon_{abc} T_c, \tag{6.22}$$

where ϵ_{abc} is the three-dimensional Levi-Civita symbol. Thus the structure constants in this basis are $f_{abc} = \epsilon_{abc}$.

Let's take a look at the family of elements of SO(3) given by $e^{i\theta T_1}$ for real θ (the reason why we use θ rather than t will be. To compute the matrix exponential, we note that $T_1^2 = M$, where we define

$$M = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{6.23}$$

We then have $T_1^3 = T_1 M = T_1$. This means that for any positive integer n we have $T_1^n = M$ if n is even and $T_1^n = T_1$ if n is odd. The power series expression for the matrix exponential then becomes

$$e^{i\theta T_1} = \sum_{n=0}^{\infty} \frac{(i\theta)^n}{n!} T_1^n = \mathbb{1} + T_1 \sum_{n=0}^{\infty} \frac{(i\theta)^{2n+1}}{(2n+1)!} + M \sum_{n=1}^{\infty} \frac{(i\theta)^{2n}}{(2n)!} \tag{6.24}$$

The sums multiplying T_1 and M are the power series of $i\sin\theta$ and $\cos\theta - 1$, respectively,

$$e^{i\theta T_1} = \mathbb{1} + iT_1 \sin\theta + M(\cos\theta - 1) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{pmatrix}. \tag{6.25}$$

This is the matrix describing a clockwise rotation about the x axis by an angle θ. Similar calculations show that $e^{i\theta T_2}$ and $e^{i\theta T_3}$ are rotations about the y and z axes,

$$e^{i\theta T_2} = \begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix}, \qquad e^{i\theta T_3} = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{6.26}$$

Example 6.2.2. The Lie algebra $\mathfrak{su}(2)$ consists of all 2×2 traceless Hermitian matrices. A 2×2 Hermitian matrix is specified by four real parameters — the two real diagonal elements and the real and imaginary parts of the element above the diagonal. Tracelessness imposes one equation relating the diagonal elements, reducing the number of parameters to three, so $\mathfrak{su}(2)$ is three-dimensional.

We will denote the generators of $\mathfrak{su}(2)$ as τ_a. The conventional basis for $\mathfrak{su}(2)$ is $\tau_a = \frac{1}{2}\sigma_a$, where σ_a are the Pauli matrices. Explicitly,

$$\tau_1 = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \tau_2 = \frac{1}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \tau_3 = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{6.27}$$

Computing commutators of these matrices, we find

$$[\tau_a, \tau_b] = i\epsilon_{abc}\tau_c, \tag{6.28}$$

so the structure constants in this basis are the same as those we found for $\mathfrak{so}(3)$ in the previous example $f_{abc} = \epsilon_{abc}$.

6.3 Lie algebra homomorphisms

Definition 6.3.1. Let $\mathfrak{g}$ and $\mathfrak{h}$ be two Lie algebras. A **Lie algebra homomorphism** is a linear map $\phi : \mathfrak{g} \to \mathfrak{h}$ that preserves the Lie bracket,

$$\phi(i[X, Y]) = i[\phi(X), \phi(Y)]. \tag{6.29}$$

A Lie algebra homomorphism that is bijective is called a **Lie algebra isomorphism**.

Example 6.3.1. The Lie algebras $\mathfrak{su}(2)$ and $\mathfrak{so}(3)$ are isomorphic, $\mathfrak{su}(2) \cong \mathfrak{so}(3)$. The isomorphism $\phi : \mathfrak{su}(2) \to \mathfrak{so}(3)$ is given by $\phi(\tau_a) = T_a$, where τ_a and T_a are the generators in examples 6.2.1 and 6.2.2. The action of ϕ on any other $X \in \mathfrak{su}(2)$ is then determined by linearity; if $X = c_a\tau_a$ for some coefficients c_a then $\phi(X) = c_aT_a$.

As the name suggests, homomorphisms of Lie algebras are related to homomorphisms of groups. Let G and H be matrix Lie groups, with Lie algebras $\mathfrak{g}$ and $\mathfrak{h}$, and let Φ be a group homomorphism $\Phi : G \to H$. For every $X \in \mathfrak{g}$ we can define a path through the identity $\Phi(e^{itX})$ in H. Then differentiating this path at $t = 0$ we obtain an element of $\mathfrak{h}$. Thus Φ induces a map from $\mathfrak{g} \to \mathfrak{h}$. We will call this map ϕ. Explicitly,

$$\phi(X) = -i \left.\frac{\mathrm{d}}{\mathrm{d}t}\left(\Phi(e^{itX})\right)\right|_{t=0}. \tag{6.30}$$

This map is a Lie algebra homomorphism thanks to the following theorem.

Theorem 6.3.1. *The map ϕ in equation (6.30) has the following properties, where X and Y denote arbitrary elements of $\mathfrak{g}$:*

(i) Linearity, $\phi(sX) = s\phi(X)$ and $\phi(X + Y) = \phi(X) + \phi(Y)$ for any $s \in \mathbb{R}$,

(ii) $\phi(AXA^{-1}) = \Phi(A)\phi(X)\Phi(A^{-1})$ for any $A \in G$,

(iii) Preservation of the Lie bracket $\phi(i[X, Y]) = i[\phi(X), \phi(Y)]$,

(iv) $\Phi(e^i X) = e^{i\phi(X)}$.

Proof. Using the chain rule we have

$$\phi(sX) = -i \left.\frac{\mathrm{d}}{\mathrm{d}t}\left(\Phi(e^{itsX})\right)\right|_{t=0} = -i \frac{\mathrm{d}(ts)}{\mathrm{d}t}\frac{\mathrm{d}}{\mathrm{d}(ts)}\left.\left(\Phi(e^{itsX})\right)\right|_{ts=0} = s\phi(X), \tag{6.31}$$

while

$$\begin{aligned}\phi(X+Y) &= -i\,\frac{\mathrm{d}}{\mathrm{d}t}\left(\Phi(e^{it(X+Y)})\right)\Big|_{t=0} \\ &= -i\,\frac{\mathrm{d}}{\mathrm{d}t}\left(\Phi(e^{itX})\Phi(e^{itY})\right)\Big|_{t=0} \\ &= \phi(X)+\phi(Y),\end{aligned} \tag{6.32}$$

where the second line is obtained using that $e^{it(X+Y)} = e^{itX}e^{itY} + O(t^2)$. This establishes point (i).

To prove point (ii), we use theorem 6.1.2 that $AXA^{-1} \in \mathfrak{g}$ for any $X \in \mathfrak{g}$ and $A \in G$. Then we calculate

$$\begin{aligned}\phi(AXA^{-1}) &= -i\,\frac{\mathrm{d}}{\mathrm{d}t}\left(\Phi(e^{itAXA^{-1}})\right)\Big|_{t=0} \\ &= -i\,\frac{\mathrm{d}}{\mathrm{d}t}\left(\Phi(Ae^{itX}A^{-1})\right)\Big|_{t=0} \\ &= -i\Phi(A)\,\frac{\mathrm{d}}{\mathrm{d}t}\left(\Phi(e^{itX})\right)\Big|_{t=0}\Phi(A^{-1}) \\ &= \Phi(A)\phi(X)\Phi(A^{-1}).\end{aligned} \tag{6.33}$$

We now use this to prove point (iii). If $X, Y \in \mathfrak{g}$, then point (ii) implies

$$\phi(e^{itX}Ye^{-itX}) = \Phi(e^{itX})\phi(Y)\Phi(e^{-itX}). \tag{6.34}$$

Differentiating both sides with respect to t and then setting $t = 0$, on the left we obtain $\phi(i[X,Y])$, while on the right we obtain $i[\phi(X), \phi(Y)]$. Thus

$$\phi(i[X,Y]) = i[\phi(X), \phi(Y)]. \tag{6.35}$$

Finally, point (iv) can be established by first considering $\Phi(e^{i\epsilon X})$ for some small ϵ. Taylor expanding, we have

$$\Phi(e^{i\epsilon X}) = \mathbb{1} + i\epsilon\phi(X) + O(\epsilon^2), \tag{6.36}$$

where we have used that for a group homomorphism $\Phi(\mathbb{1}) = \mathbb{1}$. Now we choose $\epsilon = 1/m$ for some large positive integer m. We then use that the group homomorphism preserves group multiplication to find

$$\Phi(e^{iX}) = (\Phi(e^{iX/m}))^m = \left(\mathbb{1} + \frac{i\phi(X)}{m} + O(m^{-2})\right)^m. \tag{6.37}$$

We now take the limit $m \to \infty$, in which the right-hand side becomes the matrix exponential, so we have

$$\Phi(e^{iX}) = e^{i\phi(X)}, \tag{6.38}$$

as required. □

If $\Phi : G \to H$ is an isomorphism, then the corresponding $\phi : \mathfrak{g} \to \mathfrak{h}$ will be a Lie algebra isomorphism. However, the converse is not necessarily true; two non-isomorphic groups can have isomorphic Lie algebras.

Example 6.3.2. There is a two-to-one homomorphism $\Phi : \mathrm{SU}(2) \to \mathrm{SO}(3)$. To construct it, we first note that the elements of SU(2) can be parameterised by

$$U(\alpha, \beta) = \begin{pmatrix} \alpha & -\beta^* \\ \beta & \alpha^* \end{pmatrix}, \tag{6.39}$$

where α and β are general complex numbers satisfying $|\alpha|^2 + |\beta|^2 = 1$. Multiplying two such matrices together one finds $U(\alpha_1, \beta_1)U(\alpha_2, \beta_2) = U(\alpha_1\alpha_2 - \beta_1^*\beta_2, \beta_1\alpha_2 + \alpha_1^*\beta_2)$. The homomorphism is

$$\Phi(\alpha, \beta) = \begin{pmatrix} \mathrm{Re}(\alpha^2 - \beta^2) & \mathrm{Im}(\alpha^2 - \beta^2) & 2\,\mathrm{Re}(\alpha\beta^*) \\ -\mathrm{Im}(\alpha^2 + \beta^2) & \mathrm{Re}(\alpha^2 + \beta^2) & -2\,\mathrm{Im}(\alpha\beta^*) \\ -2\,\mathrm{Re}(\alpha\beta) & -2\,\mathrm{Im}(\alpha\beta) & |\alpha|^2 - |\beta|^2 \end{pmatrix}, \tag{6.40}$$

where we use $\Phi(\alpha, \beta)$ to denote the element of $SO(3)$ that $U(\alpha, \beta)$ maps to. If you like you can check (perhaps using a computer) that $\Phi(\alpha, \beta)$ is an orthogonal matrix with unit determinant, so is an element of SO(3), and that Φ is a homomorphism, $\Phi(\alpha_1, \beta_1)\Phi(\alpha_2, \beta_2) = \Phi(\alpha_1\alpha_2 - \beta_1^*\beta_2, \beta_1\alpha_2 + \alpha_1^*\beta_2)$. To see that every element of SO(3) may be written in this way, we note that there are choices of α and β that give the usual expressions for rotations about the x, y, and z axes, and then we can get any other rotation by composing these,

$$\begin{aligned} \Phi_1(\theta) &\equiv \Phi\left(\cos(\theta/2), i\sin(\theta/2)\right) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{pmatrix}, \\ \Phi_2(\theta) &\equiv \Phi\left(\cos(\theta/2), \sin(\theta/2)\right) = \begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix}, \\ \Phi_3(\theta) &\equiv \Phi\left(e^{i\theta/2}, 0\right) = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}. \end{aligned} \tag{6.41}$$

This homomorphism is two-to-one because $\Phi(\alpha, \beta) = \Phi(-\alpha, -\beta)$. This shows that SU(2) and SO(3) are not isomorphic, even though $\mathfrak{su}(2) \cong \mathfrak{so}(3)$.

The elements Φ_a of SO(3) in equation (6.41) are the images under the group homomorphism of $U_a = e^{i\theta\tau_a}$ where τ_a are the $\mathfrak{su}(2)$ generators, for example

$$e^{i\theta\tau_1} = \begin{pmatrix} \cos(\theta/2) & i\sin(\theta/2) \\ i\sin(\theta/2) & \cos(\theta/2) \end{pmatrix} = U\left(\cos(\theta/2), i\sin(\theta/2)\right). \tag{6.42}$$

The Φ_a themselves are the matrix exponentials $e^{i\theta T_a}$ that we computed in example 6.2.1. For each Φ_a, the element of $\mathfrak{so}(3)$ computed using equation (6.30) (using θ in place of

t) will then be the generator T_a. For example

$$\phi(\tau_1) = -i\Phi_1'(0) = i\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} = T_1. \tag{6.43}$$

Similarly we find $\phi(\tau_2) = T_2$ and $\phi(\tau_3) = T_3$. Thus the Lie algebra homomorphism $\phi : \mathfrak{su}(2) \to \mathfrak{so}(3)$ induced by the group homomorphism (6.40) is the isomorphism from example 6.3.1.

6.4 Subalgebras

Definition 6.4.1. If $\mathfrak{g}$ is a Lie algebra, a **subalgebra** of $\mathfrak{g}$ is a subspace $\mathfrak{h}$, such that $i[Y_1, Y_2] \in \mathfrak{h}$ for all $Y_{1,2} \in \mathfrak{h}$. The subgalgebra $\mathfrak{h}$ is said to be an **ideal** if $i[X, Y] \in \mathfrak{h}$ for all $X \in \mathfrak{g}$ and $Y \in \mathfrak{h}$.

Subalgebras are the Lie algebra analogue of subgroups, a smaller algebra that is closed under the bracket operation. An ideal is then the analogue of a normal subgroup. Every Lie algebra has two trivial ideals. One is the Lie algebra itself, and the other is subalgebra $\{0\}$ consisting only of the zero matrix.

Definition 6.4.2. A Lie algebra $\mathfrak{g}$ is called **simple** if it has no non-trivial ideals and $\dim \mathfrak{g} \geq 2$.[3]

Example 6.4.1. The Lie algebra $\mathfrak{su}(2)$ is simple. To prove this, suppose that $\mathfrak{h}$ is an ideal of $\mathfrak{su}(2)$ and let $Y \in \mathfrak{h}$. We expand Y in the basis τ_b as $Y = y_b\tau_b$ for some coefficients y_b. By definition of an ideal we must have that $\{\tau_a, Y\} \in \mathfrak{h}$ for all a. Using the commutation relation (6.28) we find that

$$[\tau_a, Y] = i\epsilon_{abc}y_b\tau_c. \tag{6.44}$$

But since $i[\tau_a, Y] \in \mathfrak{h}$, we must also have that $[\tau_d, [\tau_a, Y]] \in \mathfrak{h}$. Calculating,

$$\begin{aligned}[\tau_d, [\tau_a, Y]] &= i\epsilon_{abc}y_b[\tau_d, \tau_c] = -\epsilon_{abc}\epsilon_{dce}y_b\tau_e = (\delta_{ad}\delta_{be} - \delta_{ae}\delta_{bd})y_b\tau_e \\ &= \delta_{ad}Y - y_d\tau_a. \end{aligned} \tag{6.45}$$

Since $\mathfrak{h}$ is a vector space and therefore closed under addition, we also have that

$$\delta_{ad}Y - [\tau_d, [\tau_a, Y]] = y_d\tau_a \in \mathfrak{h}, \tag{6.46}$$

for any d and a. There are two possibilities. One possibility is that every element of $\mathfrak{h}$ has coefficients $y_d = 0$ for all d, and $\mathfrak{h} = \{0\}$. The other possibility is that at least one of the y_d are non-zero. Then equation (6.46) implies that $\tau_a \in \mathfrak{h}$ for all a, which implies that $\mathfrak{h} = \mathfrak{su}(2)$ since the τ_a are a basis for $\mathfrak{su}(2)$. Either of these possibilities implies that $\mathfrak{h}$ is a trivial ideal. Thus $\mathfrak{su}(2)$ has no non-trivial ideals, so is a simple Lie algebra.

In fact it turns out that $\mathfrak{su}(n)$ is simple for any $n \geq 2$. It also turns out that $\mathfrak{so}(n)$

[3] If you prefer this phrased without a double negative: A Lie algebra is simple if its only ideals a trivial.

is simple for any $n \geq 3$,[4] with the exception of $\mathfrak{so}(4)$. There is an isomorphism $\mathfrak{so}(4) \cong \mathfrak{su}(2) \oplus \mathfrak{su}(2)$, where elements of two different $\mathfrak{su}(2)$ factors commute with one another. Thus both the $\mathfrak{su}(2)$ factors are non-trivial ideals.

Definition 6.4.3. A Lie algebra $\mathfrak{g}$ is called **semisimple** if it has no non-trivial *abelian* ideals and $\dim \mathfrak{g} \geq 2$.

This definition means that a simple Lie algebra is also semisimple, so $\mathfrak{su}(n)$ is semisimple for any $n \geq 2$ and $\mathfrak{so}(n)$ is semisimple for any $n \geq 3$. This includes the case of $\mathfrak{so}(4)$, since the non-trivial $\mathfrak{su}(2)$ ideals are non-abelian.

[4] The reason $\mathfrak{so}(2)$ is not simple is because it is one-dimensional.

Chapter 7

Representations of Lie groups and Lie algebras

7.1 Representations of Lie algebras

A representation of a group G is a homomorphism from G to the group $\mathrm{GL}(m;\mathbb{C})$ of invertible linear transformations acting on some m-dimensional vector space. If G is a Lie group with Lie algebra $\mathfrak{g}$, then as discussed in section 6.3 there will be an associated Lie algebra homomorphism from $\mathfrak{g}$ to the Lie algebra of $\mathrm{GL}(m;\mathbb{C})$ which we denote $\mathfrak{gl}(m;\mathbb{C})$. We call this a representation of the Lie algebra.

Definition 7.1.1. An n-dimensional **representation** of a Lie algebra $\mathfrak{g}$ on a vector space V is a Lie algebra homomorphism $\pi : \mathfrak{g} \to \mathfrak{gl}(m;\mathbb{C})$, where $\mathfrak{gl}(m;\mathbb{C})$ is the Lie algebra of $\mathrm{GL}(m;\mathbb{C})$.

This is a fancy way of saying that a representation of a Lie algebra is a map from the Lie algebra to a set of matrices, such that the Lie bracket is mapped to i times the matrix commutator.

Note that we have two notions of dimension floating around now. The dimension of a Lie algebra is how many independent generators it has. The dimension of a representation is the size of the matrices in the representation, i.e. it is m in the definition above.

Let G be a Lie group with Lie algebra $\mathfrak{g}$. We exponentiate Lie algebra elements to obtain elements of the group, i.e. for every $X \in \mathfrak{g}$ we have that $e^{iX} \in G$. If Π is a representation of G, then it induces a representation π of $\mathfrak{g}$ through $\Pi(e^{iX}) = e^{i\pi(X)}$. If $\pi(X)$ is a Hermitian operator then $\Pi(X)$ is unitary, motivating the following definition:

Definition 7.1.2. A representation π of a Lie algebra $\mathfrak{g}$ is **unitary** if $\pi(X)$ is Hermitian for every $X \in \mathfrak{g}$.

Suppose that π has an invariant subspace $W \subset \mathbb{C}^m$, meaning that $\pi(X)|w\rangle \in W$ for every $|w\rangle \in W$ and $X \in \mathfrak{g}$. Then using the power series definition of the exponential we see that $e^{i\pi(X)}|w\rangle$ is also in W, so an invariant subspace of the Lie algebra representation is also an invariant subspace of the group representation. Thus if we want to find irreducible

representations of Lie groups, we should look for Lie algebra representations with no invariant subspaces:

Definition 7.1.3. A representation of a Lie algebra is **irreducible** if it has no non-trivial invariant subspaces.

If S is an invertible linear transformation, then defining $\pi'(X) = S^{-1}\pi(X)S$ for all $X \in \mathfrak{g}$ generates a new representation of the Lie algebra. The corresponding group representations are related by $e^{i\pi'(X)} = e^{iS^{-1}\pi(X)S} = S^{-1}e^{i\pi(X)}S$ so are equivalent. It therefore makes sense to define π' and π to be equivalent, too:

Definition 7.1.4. Two representations π and π' of a Lie algebra $\mathfrak{g}$ are **equivalent** if there exists an invertible linear transformation S such that $\pi' = S^{-1}\pi S$.

Here is a representation that exists for all of the Lie groups and Lie algebras that we consider:

Definition 7.1.5. Matrix Lie groups and their Lie algebras are defined in terms of matrices. We thus always have a faithful representation in which we represent the group/Lie algebra elements by themselves. In physics this representation is often called the **fundamental representation**. (In mathematics this representation is instead called the defining representation.)

7.2 The adjoint representation

Definition 7.2.1. Let $\mathfrak{g}$ be an n-dimensional Lie algebra with generators T_a and structure constants f_{abc}. The **adjoint representation** of $\mathfrak{g}$ is an n-dimensional representation π_{adj} such that the matrix elements of the generators are given by the structure constants,

$$[\pi_{\text{adj}}(T_a)]_{bc} = -if_{abc}. \tag{7.1}$$

The adjoint representation of an abritrary element $X = x_aT_a \in \mathfrak{g}$ is then obtained through linearity,

$$[\pi_{\text{adj}}(X)]_{bc} = x_a[\pi_{\text{adj}}(T_a)]_{bc} = -ix_af_{abc}. \tag{7.2}$$

Let's check that this forms a representation. The matrix elements of the commutator are

$$\begin{aligned}[\pi_{\text{adj}}(T_a), \pi_{\text{adj}}(T_b)]_{cd} &= [\pi_{\text{adj}}(T_a)]_{ce}[\pi_{\text{adj}}(T_b)]_{ed} - [\pi_{\text{adj}}(T_b)]_{ce}[\pi_{\text{adj}}(T_a)]_{ed} \\ &= -f_{ace}f_{bed} + f_{bce}f_{aed}. \end{aligned} \tag{7.3}$$

We can simplify the right-hand side using the Jacobi identity, obtaining

$$[\pi_{\text{adj}}(T_a), \pi_{\text{adj}}(T_b)]_{cd} = f_{abe}f_{ecd} = if_{abe}[\pi_{\text{adj}}(T_e)]_{cd}. \tag{7.4}$$

So the matrices $\pi_{\text{adj}}(T_a)$ satisfy the correct commutator algebra, and thus form a representation of $\mathfrak{g}$.

Example 7.2.1. Using the $\mathfrak{su}(2)$ generators τ_a from example 6.2.2, we have that the structure constants are $f_{abc} = \epsilon_{abc}$. The adjoint representation of $\mathfrak{su}(2)$ is then

$$\pi_{\text{adj}}(\tau_1) = i\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \qquad \pi_{\text{adj}}(\tau_2) = i\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \qquad \pi_{\text{adj}}(\tau_3) = i\begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{7.5}$$

These matrices are the same as the generators of $\mathfrak{so}(3)$ in example 6.2.1, which in the language of representations we would say are the generators in the fundamental representation. The structure constants of $\mathfrak{so}(3)$ are also $f_{abc} = \epsilon_{abc}$, so the adjoint representation of $\mathfrak{so}(3)$ is the same as its fundamental representation.

Definition 7.2.2. The **Killing form** is a bilinear form $\mathfrak{g} \times \mathfrak{g} \to \mathbb{R}$ defined by

$$B(X, Y) = \text{tr}(\pi_{\text{adj}}(X)\pi_{\text{adj}}(Y)). \tag{7.6}$$

The Killing form provides an alternative definition of a semisimple Lie algebra:

Theorem 7.2.1. *A Lie algebra $\mathfrak{g}$ is semisimple if and only if its Killing form is non-degenerate, in other words there is no $X \in \mathfrak{g}$ such that $B(X, Y) = 0$ for all $Y \in \mathfrak{g}$.*

We omit the proof of this theorem.

For a given basis of generators T_a, we can consider the matrix B_{ab} formed by evaluating the Killing form on pairs of generators,

$$B_{ab} = B(T_a, T_b) = \text{tr}\left(\pi_{\text{adj}}(T_a)\pi_{\text{adj}}(T_b)\right), \tag{7.7}$$

which is manifestly symmetric in the indices a and b. Suppose that we define a new basis $T'_a = L_{ab}T_b$, where L is some invertible matrix. In the new basis of generators we find

$$B'_{ab} \equiv B(T'_a, T'_b) = L_{ac}L_{bd}B_{cd}. \tag{7.8}$$

In matrix notation this is $B' = LBL^T$. Since B is a symmetric matrix it is always possible to choose L such that B' is a diagonal matrix,[1]

$$B'_{ab} = k_a\delta_{ab} \qquad \text{(no sum over } a\text{)} \tag{7.9}$$

If we like we can then rescale the generators to change the values of the diagonal elements k_a. For example, if for each non-zero k_a we define $T''_a = T'_a/\sqrt{|k_a|}$, then have that each diagonal element of B''_{ab} is ± 1 or zero. Notice that we can't change the sign of the diagonal elements by doing this.

Definition 7.2.3. Lie algebras for which all of the diagonal elements k_a in equation (7.8) are non-negative are called **compact Lie algebras**.

[1] If you want to look up more about this, the diagonalisation of a complex symmetric matrix is called *Autonne–Takagi factorization*.

Theorem 7.2.1 implies that if a Lie algebra is semisimple then none of the k_a can vanish and therefore if the Lie algebra is also compact the k_a must all be strictly positive. Thus, for a compact semisimple Lie algebra we can rescale the generators so that all of the k_a are the same,

$$B(T_a, T_b) = \lambda\, \delta_{ab}, \tag{7.10}$$

for some positive λ that we can choose. The representation theory of compact semisimple Lie algebras turns out to have an elegant structure, which we will explore in the rest of the lectures. We finally note a useful result:

Theorem 7.2.2. *In the basis in which the Killing form is* (7.10)*, the structure constants of a compact semisimple Lie algebra are totally antisymmetric in their indices.*

Proof. We have already discussed that in *any* basis that the structure constants are antisymmetric under interchange of their first two indices. Now notice that

$$\mathrm{tr}\,([T_a, T_b]T_c) = i f_{abd}\,\mathrm{tr}\,(T_d T_c) = i\lambda f_{abc}, \tag{7.11}$$

where we have adopted notation in which we write simply T_a inside the trace, rather that $\pi_{\mathrm{adj}}(T_a)$. Rearranging, we have $f_{abc} = -i\lambda^{-1}\,\mathrm{tr}\,([T_a, T_b]T_c)$, which is totally antisymmetric due to the cyclic property of the trace, for example expanding out the commutator we have

$$\begin{aligned} f_{abc} &= -i\lambda^{-1}\,\mathrm{tr}\,(T_aT_bT_c - T_bT_aT_c) = -i\lambda^{-1}\,\mathrm{tr}\,(T_cT_aT_b - T_aT_cT_b) = i\lambda^{-1}\,\mathrm{tr}\,([T_a, T_c]T_b) \\ &= -f_{acb}, \end{aligned} \tag{7.12}$$

so in this basis the structure constants are antisymmetric under interchange of their last two indices, and are thus totally antisymmetric. □

When dealing with compact semisimple Lie algebras we will always assume that we have chosen a basis in which the structure constants are totally antisymmetric.

In a basis in which the structure constants are totally antisymmetric, $\pi_{\mathrm{adj}}(T_a)$ is an antisymmetric, purely imaginary matrix, and therefore unitary. We therefore see that the adjoint representation of a compact semisimple Lie algebra can be taken to be unitary.

Example 7.2.2. Using the adjoint representation of the $\mathfrak{su}(2)$ generators constructed in example 7.2.1, we find

$$B(\tau_a, \tau_b) = 2\,\delta_{ab} \tag{7.13}$$

From this we learn that $\mathfrak{su}(2)$ is a compact Lie algebra. The structure constants in this basis are $f_{abc} = \epsilon_{abc}$, which are indeed totally antisymmetric.

7.3 Direct sums and tensor products

Often we combine two or more representations of a group via direct sums or tensor products. It will therefore be important to know what these operations mean for the corresponding Lie algebra representations.

Let G be a Lie group and $\mathfrak{g}$ its Lie algebra. Let Π_1 and Π_2 be two representations of G and π_1 and π_2 the associated representations of $\mathfrak{g}$. The Lie algebra representation associated to the direct sum $\Pi_1 \oplus \Pi_2$ is denoted $\pi_1 \oplus \pi_2$. We can determine the matrix representation $(\pi_1 \oplus \pi_2)(X)$ by differentiating the block-diagonal matrix $\Pi_1(e^{itX}) \oplus \Pi_2(e^{itX})$,

$$(\pi_1 \oplus \pi_2)(X) = -i \, \frac{\mathrm{d}}{\mathrm{d}t} \Big(\Pi_1(e^{itX}) \oplus \Pi_2(e^{itX}) \Big) \Big|_{t=0} = \pi_1(X) \oplus \pi_2(X). \tag{7.14}$$

The Lie algebra representation $\pi_1 \otimes \pi_2$ associated to the tensor product $\Pi_1 \otimes \Pi_2$ may also be obtained by differentiation, using the product rule we find

$$(\pi_1 \otimes \pi_2)(X) = -i \, \frac{\mathrm{d}}{\mathrm{d}t} \Big(\Pi_1(e^{itX}) \otimes \Pi_2(e^{itX}) \Big) \Big|_{t=0} = \pi_1(X) \otimes \mathbb{1} + \mathbb{1} \otimes \pi_2(X). \tag{7.15}$$

Chapter 8

Representation theory of $\mathfrak{su}(2)$

8.1 Irreducible representations of $\mathfrak{su}(2)$

In this section we will construct all of the finite-dimensional irreducible representations of $\mathfrak{su}(2)$. This is a calculation you will actually have done already in a quantum mechanics course, under the guise of constructing angular momentum eigenstates.

Let π be a finite-dimensional irreducible representation of $\mathfrak{su}(2)$ acting on a vector space V. If τ_a are the usual generators of $\mathfrak{su}(2)$, then $\pi(\tau_a)$ are three matrices satisfying $[\pi(\tau_a), \pi(\tau_b)] = i\epsilon_{abc}\pi(\tau_c)$. The strategy to construct a representation is to diagonalise $\pi(\tau_3)$. We adopt the notation of angular momentum in quantum mechanics by defining the matrices

$$J_\pm = \frac{\pi(\tau_1) \pm i\pi(\tau_2)}{\sqrt{2}}, \qquad J_3 = \pi(\tau_3). \tag{8.1}$$

Using the commutation relations of the $\pi(T_a)$, one finds that these matrices satisfy

$$[J_3, J_\pm] = \pm J_\pm, \qquad [J_+, J_-] = J_3. \tag{8.2}$$

Let $|v_m\rangle$ be an eigenvector of H, with eigenvalue m. Then

$$J_3 J_\pm |v_m\rangle = (J_\pm J_3 + [J_3, J_\pm])\,|v_m\rangle = (J_\pm J_3 \pm J_\pm)\,|v_m\rangle = (m \pm 1) J_\pm |v_m\rangle. \tag{8.3}$$

So either $J_\pm|v_m\rangle = 0$, or $J_\pm|v_m\rangle$ is also an eigenvector of J_3, with eigenvalue $m \pm 1$. Thus starting from any eigenstate of J_3, we can repeatedly apply J_+ to create more eigenstates with larger eigenvalues. Since J_3 is a finite-dimensional matrix it must have a finite number of eigenvalues, so this process must eventually stop; we must reach some eigenstate $|v_j\rangle$ such that $J_+|v_j\rangle = 0$. This state is called the **highest weight state**.

Now we define the eigenstates of J_3 obtained by repeated application of J_- to $|v_j\rangle$ as

$$|v_{j-k}\rangle = J_-^k |v_j\rangle, \tag{8.4}$$

for non-negative integers k. These satisfy $J_3|v_{j-k}\rangle = (j-k)|v_{j-k}\rangle$ and

$$J_+|v_{j-k}\rangle = \alpha(k)|v_{j-k+1}\rangle, \qquad \alpha(k) = kj - \frac{1}{2}k(k-1), \tag{8.5}$$

which may be proved by induction. Since we cannot generate infinitely many eigenvalues by repeated application of J_-, we must eventually reach some K such that $|v_{j-K}\rangle \neq 0$ but $J_-|v_{j-K}\rangle = 0$. Using the commutation relations of $J_\pm$ and J_3 along with equation (8.5) we find

$$\begin{aligned}(j-K)|v_{j-K}\rangle = J_3|v_{j-K}\rangle = [J_+, J_-]|v_{j-K}\rangle = -J_-J_+|v_{j-K}\rangle \\ = -\alpha(K)|v_{k-K}\rangle.\end{aligned} \tag{8.6}$$

Solving $-\alpha(K) = j - K$ for K we find that $K = 2j$. Since K is a non-negative integer, this implies that j is either a non-negative integer or a half-integer. In physics we usually call j the *spin* of the representation that we are constructing.

We have found that for every finite-dimensional irreducible representation π there is a non-negative integer or half-integer j, and a family of $2j+1$ eigenstates of J_3. Defining $m = j - k$, these eigenstates are $\{|v_m\rangle\}$ satisfying

$$J_3|v_m\rangle = m|v_m\rangle, \qquad m = -j, -j+1, \cdots, j-1, j, \tag{8.7}$$

and

$$\begin{aligned}J_+|v_m\rangle &= \begin{cases}\alpha(j-m)|v_{m+1}\rangle, & m < j,\\ 0, & m = j,\end{cases}\\ J_-|v_m\rangle &= \begin{cases}|v_{m-1}\rangle, & m > -j,\\ 0, & m = -j.\end{cases}\end{aligned} \tag{8.8}$$

The eigenstates $\{|v_m\rangle\}$ are linearly independent since they correspond to distinct eigenvalues of J_3. The subspace of V spanned by $\{|v_m\rangle\}$ is an invariant subspace of the representation π, since acting with J_3 or $J_\pm$ on any state in this subspace produces another state in this subspace. Since π is assumed irreducible, this subspace must be all of V. Thus π is a $(2j+1)$-dimensional representation that acts on linear combinations of the $\{|v_m\rangle\}$ as determined by equations (8.7) and (8.8).

If we like we can construct the explicit matrices of the representation corresponding to a given j. As always, the matrix elements depend on a choice of basis. Using $\{|v_m\rangle\}$ for a basis is not so nice because these vectors do not all have the same norm; using $J_+ = J_-^\dagger$ we find

$$\begin{aligned}\langle v_m|v_m\rangle &= \langle v_{m+1}|J_+J_-|v_{m+1}\rangle\\ &= \langle v_{m+1}|\left(J_-J_+ + [J_+, J_-]\right)|v_{m+1}\rangle\\ &= \left[\alpha(j-m-1) + m + 1\right]\langle v_{m+1}|v_{m+1}\rangle,\end{aligned} \tag{8.9}$$

where we make use of the commutation relation $[J_+, J_-] = J_3$. An orthonormal basis $\{|j,m\rangle\}$ is then obtained by defining $|j,m\rangle = a_m^{-1/2}|v_m\rangle$ where $a_j = 1$ and

$$a_m = \prod_{k=1}^{j-m}\left[\alpha(k+1) + j + 1 - k\right] = \prod_{k=1}^{j-m}\frac{k}{2}(2j+1-k), \tag{8.10}$$

for $m < j$. In this basis, equations (8.7) and (8.8) become

$$J_3|j,m\rangle = m|j,m\rangle, \qquad J_+|j,m\rangle = c(j,m+1)|j,m+1\rangle, \qquad J_-|j,m\rangle = c(j,m)|j,m-1\rangle, \tag{8.11}$$

where

$$c(j,m) = \sqrt{\frac{(j+m)(j-m+1)}{2}}. \tag{8.12}$$

In this basis J_3 is a diagonal matrix, while the matrix elements of J_+ are almost all zero except for just above the diagonal where they are given by $c(j,m)$. Similarly, the matrix elements of J_- are zero except just below the diagonal,

$$J_+ = \begin{pmatrix} 0 & c(j,j) & 0 & \cdots \\ 0 & 0 & c(j,j-1) & \cdots \\ 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \quad J_- = \begin{pmatrix} 0 & 0 & 0 & \cdots \\ c(j,j) & 0 & 0 & \cdots \\ 0 & c(j,j-1) & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \tag{8.13}$$

We then obtain the representation matrices of the $\mathfrak{su}(2)$ generators by inverting equation (8.1), $\pi(\tau_1) = (J_+ + J_-)/\sqrt{2}$, $\pi(\tau_2) = i(J_- - J_+)/\sqrt{2}$, and $\pi(\tau_3) = J_3$.

Example 8.1.1. For $j = 1/2$ we have

$$J_+ = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad J_- = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \qquad J_3 = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{8.14}$$

and so for $j = 1/2$ we find $\pi(\tau_a) = \tau_a$. In other words, the $j = 1/2$ representation is the fundamental representation.

Example 8.1.2. For $j = 1$ we obtain

$$J_+ = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \qquad J_- = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \qquad J_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \tag{8.15}$$

Then the representation matrices are $\pi(\tau_3) = J_3$ and

$$\pi(\tau_1) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \qquad \pi(\tau_2) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{pmatrix}. \tag{8.16}$$

This representation is equivalent to the adjoint representation.

Suppose that π is a unitary representation, meaning that $\pi(X)$ is Hermitian for every $X \in \mathfrak{su}(2)$. The $J_\pm$ are not Hermitian, but rather satisfy $J_\pm^\dagger = J_\mp$. Thus $J_\pm$ are not representation matrices for any element of the Lie algebra. You should think of $J_\pm$ as useful intermediate constructions that allow us to obtain unitary representations. Examples 8.1.1 and 8.1.2 demonstrate this; the $J_\pm$ are not Hermitian, but the $\pi(\tau_a)$ are.

8.2 The quadratic Casimir

A useful operator is the sum of the squares of $\pi(\tau_a)$, which in quantum mechanics is usually called the total angular momentum operator. In the context of Lie algebra theory we give it a different name:

Definition 8.2.1. The **quadratic Casimir operator** is

$$J^2 = \sum_a \pi(\tau_a)^2 = J_+J_- + J_-J_+ + J_3^2. \tag{8.17}$$

The distinguishing feature of this operator is that it commutes with all of the $\pi(\tau_a)$. It can therefore be diagonalised simultaneously with J_3. The state $|j,m\rangle$ is an eigenstate of the quadratic Casimir operator with eigenvalue given by

$$J^2|j,m\rangle = j(j+1)|j,m\rangle. \tag{8.18}$$

We will use this in the next subsection to extract the total spin of states constructed from tensor products. It will be sometimes be convenient to rewrite the quadratic Casimir as

$$J^2 = 2J_-J_+ + J_3^2 + J_3, \tag{8.19}$$

where we have used that $[J_+, J_-] = J_3$.

8.3 Tensor products of $\mathfrak{su}(2)$ representations

Let π_{j_1} and π_{j_2} be the spin-j_1 and spin-j_2 irreducible representations of $\mathfrak{su}(2)$, respectively. The tensor product representation $\pi_{j_1\otimes j_2} = \pi_{j_1} \otimes \pi_{j_2}$ acts on states which are linear combinations of products $|j_1,m_1\rangle|j_2,m_2\rangle$. We would like to know how $\pi_{j_1\otimes j_2}$ decomposes as a direct sum of irreducible representations. You will already have encountered this situation in quantum mechanics when writing wavefunctions for a pair of particles with spin j_1 and j_2 with definite angular momentum, so we will be brief with the details.

The tensor product representation of τ_3, $J_3 \equiv \pi_{j_1\otimes j_2}(\tau_3)$, acts on tensor products of the eigenstates as

$$\begin{aligned} J_3|j_1,m_1\rangle|j_2,m_2\rangle &= (\pi_{j_1}(\tau_3)|j_1,m_1\rangle)|j_2,m_2\rangle + |j_1,m_1\rangle(\pi_{j_2}(\tau_3)|j_2,m_2\rangle), \\ &= (m_1+m_2)|j_1,m_1\rangle|j_2,m_2\rangle. \end{aligned} \tag{8.20}$$

Thus $|j_1,m_1\rangle|j_2,m_2\rangle$ is an eigenstate of J_3 with eigenvalue m_1+m_2. The largest eigenvalue of J_3 is therefore $j = j_1 + j_2$, corresponding to the eigenstate

$$|j,j\rangle = |j_1,j_1\rangle|j_2,j_2\rangle. \tag{8.21}$$

Similarly, the smallest possible J_3 eigenvalue is $-j = -j_1 - j_2$, corresponding to the eigenstate $|j,-j\rangle = |j_1,-j_1\rangle|j_2,-j_2\rangle$.

Applying the lowering operator $J_- = (\pi_{j_1\otimes j_2}(\tau_1) - i\pi_{j_1\otimes j_2}(\tau_2))/\sqrt{2}$ to equation (8.21) generates a new state with J_3 eigenvalue $j-1$,

$$\begin{aligned} |j,j-1\rangle &= c(j,m)^{-1}J_-|j_1,j_1\rangle|j_2,j_2\rangle \\ &= j^{-1/2}\left(J_-|j_1,j_1\rangle\right)|j_2,j_2\rangle + j^{-1/2}|j_1,j_1\rangle\left(J_-|j_2,j_2\rangle\right) \\ &= \sqrt{\frac{j_1}{j}}|j_1,j_1-1\rangle|j_2,j_2\rangle + \sqrt{\frac{j_2}{j}}|j_1,j_1\rangle|j_2,j_2-1\rangle, \end{aligned} \tag{8.22}$$

where we abuse notation on the right-hand side by using J_- to denote the lowering operator in whichever representation is appropriate for the state the operator is acting on. We can

continue applying lowering operators, creating a state $|j, j-2\rangle \propto J_-|j, j-1\rangle$ and so on, until we reach the state $|j, -j\rangle \propto J_-^{2j}|j, j\rangle$. In this way we create a subspace of states, transforming under the spin-j irreducible representation.

The space of J_3 eigenstates with eigenvalue $j-1$ is two-dimensional, spanned by the states

$$|j_1, j_1 - 1\rangle|j_2, j_2\rangle, \qquad |j_1, j_1\rangle|j_2, j_2 - 1\rangle \tag{8.23}$$

The state $|j, j-1\rangle$ in equation (8.22) is a linear combination of these two sates. We can also form the orthogonal linear combination

$$|j-1, j-1\rangle = \sqrt{\frac{j_2}{j}}|j_1, j_1 - 1\rangle|j_2, j_2\rangle - \sqrt{\frac{j_1}{j}}|j_1, j_1\rangle|j_2, j_2 - 1\rangle. \tag{8.24}$$

This is an eigenstate of J_3 with eigenvalue $j-1$ satisfying $J_+|j-1, j-1\rangle = 0$. Applying the quadratic Casimir operator defined in equation (8.17), we find

$$J^2|j-1, j-1\rangle = \left(2J_-J_+ + J_3^2 + J_3\right)|j-1, j-1\rangle = j(j-1)|j-1, j-1\rangle, \tag{8.25}$$

where the first equality is obtained using the commutation relation $[J_+, J_-] = J_3$. Comparing to equation (8.18) we read off that this state has total spin $(j-1)$, so calling the stae "$|j-1, j-1\rangle$" makes sense. We can then repeatedly apply the lowering operator, $|j-1, j-2\rangle \propto |j-1, j-1\rangle$ and so on, to create a subspace of states transforming under the spin-$(j-1)$ irreducible representation.

The space of J_3 eigenstates with eigenvalue $j-2$ is three-dimensional. It is spanned by the states

$$|j_1, j_1 - 2\rangle|j_2, j_2\rangle, \qquad |j_1, j_1 - 1\rangle|j_2, j_2 - 1\rangle, \qquad |j_1, j_1\rangle|j_2, j_2 - 2\rangle. \tag{8.26}$$

Two linear combinations form the states $|j, j-2\rangle$ and $|j-1, j-2\rangle$ that are part of the spin-j and spin-$(j-1)$ subspaces discussed above. The remaining linear combination orthogonal to these two has total spin-$(j-2)$, and using lowering operators we construct the full set of states transforming under the spin-$(j-2)$ irreducible representation. Continuing on in this way, each time we lower the J_3 eigenvalue to some $m \geq |j_1 - j_2|$ we find a new state that's not part of one the subspaces we previously contructed by applying lowering operators to the states with $J_3 > m$.

We stop finding new states below $m = |j_1 - j_2|$. To see why, suppose $j_1 \geq j_2$. The space with J_3 eigenvalue $|j_1 - j_2|$ is spanned by

$$|j_1, j_1 - 2j_2\rangle|j_2, j_2\rangle, \quad |j_1, j_1 - 2j_2 + 1\rangle|j_2, j_2 - 1\rangle, \quad \cdots, \quad |j_1, j_1\rangle|j_2, -j_2\rangle.$$

The space with J_3 eigenvalue $|j_1 - j_2| - 1$ is spanned by

$$|j_1, j_1 - 2j_2 - 1\rangle|j_2, j_2\rangle, \quad |j_1, j_1 - 2j_2\rangle|j_2, j_2 - 1\rangle, \quad \cdots, \quad |j_1, j_1 - 1\rangle|j_2, -j_2\rangle,$$

which is the same number of states. We have not obtained another independent state at this level because we do not have $|j_1, j_1\rangle|j_2, -j_2 - 1\rangle$ as this state vanishes. If $j_2 < j_1$ the argument is the same.

Through the process described above, we find that the tensor product of two irreducible representations of $\mathfrak{su}(2)$ with spins j_1 and j_2 decomposes as a direct sum of one copy of each irreducible representation of spin between $j_1 + j_2$ and $|j_1 - j_2|$,

$$\pi_{j_1} \otimes \pi_{j_2} = \pi_{j_1+j_2} \oplus \pi_{j_1+j_2-1} \oplus \cdots \oplus \pi_{|j_1-j_2|}. \tag{8.27}$$

We have seen that the angular momentum eigenstates in a tensor product representation are linear combinations of product states,

$$|j, m\rangle = \sum_{m_1=-j_1}^{j_1} \sum_{m_2=-j_2}^{j_2} C(j, m; j_1, m_1, j_2, m_2)|j_1, m_1\rangle|j_2, m_2\rangle. \tag{8.28}$$

The coefficients in this expansion are called **Clebsch–Gordan coefficients**. They are given by the overlap

$$C(j, m; j_1, m_1, j_2, m_2) = \langle j_1, m_1|\langle j_2, m_2|j, m\rangle \tag{8.29}$$

The process described above of repeatedly applying lowering operators provides a way to systematically calculate them. For example, from equation (8.22) we read off that

$$C(j, j-1; j_1, j_1-1, j_2, j_2) = \sqrt{\frac{j_2}{j}}, \qquad C(j, j-1; j_1, j_1, j_2, j_2-1) = \sqrt{\frac{j_1}{j}}. \tag{8.30}$$

The great advantage is that they are determined completely by the representation theory, not any underlying physics. This means that you can look up tables of them online and apply them to any physical problem with SU(2) symmetry that you are dealing with.

Applying the identity operator $\mathbb{1} = \sum_{j,m} |j, m\rangle\langle j, m|$ to a product state $|j_1, m_1\rangle|j_2, m_2\rangle$ we find that the Clebsch-Gordan coefficients also appear in the expansion of the product state in the basis of total angular momentum eigenstates,

$$|j_1, m_1\rangle|j_2, m_2\rangle = \sum_{j=0}^{\infty} \sum_{m=-j}^{j} C(j, m; j_1, m_1, j_2, m_2)|j, m\rangle. \tag{8.31}$$

The reason that the Clebsch-Gordan coefficients do not appear complex conjugated in this formula is that they are conventionally taken to be real.

8.4 Tensor operators and the Wigner–Eckart theorem

Many of the operators that are important in quantum mechanics transform under rotations. Examples include the position, momentum, and angular momentum operators, which all transform as vectors under rotations. Such operators are called tensor operators. The matrix elements of tensor operators in different angular momentum eigenstates are related by symmetry through the Wigner–Eckart theorem.

In this section we will use J_a to denote the components of the quantum mechanical angular momentum operator. Angular momentum eigenstates may also carry quantum numbers of other operators that commute with angular momentum, such as energy level.

We represent these additional quantum numbers with an index α, writing $|j, m, \alpha\rangle$. The angular momentum operators act on these states as[1]

$$J_a|j, m, \alpha\rangle = \sum_{m'=-j}^{j} |j, m', \alpha\rangle[\pi_j(\tau_a)]_{m'm}, \tag{8.32}$$

where $[\pi_j(\tau_a)]_{m'm} = \langle j, m', \alpha|J_a|j, m, \alpha\rangle$ are the matrix elements of τ_a in the spin-j irreducible representation.

Definition 8.4.1. An $\mathfrak{su}(2)$ **tensor operator** with spin ℓ is an operator O^ℓ_k, with index k running from $-\ell$ to ℓ, sastisfying

$$[J_a, O^\ell_k] = \sum_{k'=-\ell}^{\ell} O^\ell_{k'} \left[\pi_\ell(\tau_a)\right]_{k'k}, \tag{8.33}$$

where π_ℓ is the spin-ℓ irreducible representation. Spin-one tensor operators are called **vector operators**. In the basis constructed in section 8.1, equation (8.33) is equivalent to the relations

$$[J_+, O^\ell_k] = c(\ell, k+1)\, O^\ell_{k+1}, \qquad [J_-, O^\ell_k] = c(\ell, k)\, O^\ell_{k-1}, \qquad [J_3, O^\ell_k] = kO^\ell_k, \tag{8.34}$$

where as usual $J_\pm = (J_1 \pm iJ_2)/\sqrt{2}$, and $c(\ell, k)$ are the coefficients given in equation (8.12), $c(\ell, k) = \sqrt{(\ell + k)(\ell - k + 1)/2}$

Example 8.4.1. Angular momentum is a vector operator since it satisfies the $\mathfrak{su}(2)$ commutation relations,

$$[J_a, J_b] = i\epsilon_{abc}J_c = -i\epsilon_{acb}J_c = J_c \left[\pi_{\text{adj}}(\tau_a)\right]_{cb}, \tag{8.35}$$

where we use Einstein summation notation, so the index c is summed over. The adjoint representation is equivalent to the spin-one irreducible representation (problem **??**), so we should be able to change basis by forming linear combinations of the J_a that satisfy equation (8.34) with $\ell = 1$. Consider the vector $\mathcal{J}_k$ with components

$$\mathcal{J}_1 = -J_+, \qquad \mathcal{J}_0 = J_3, \qquad \mathcal{J}_{-1} = J_-. \tag{8.36}$$

[1] To understand why the indices go where they do in equation (8.32), we multiply the left-hand side of the equation by the identity operator in the form $\mathbb{1} = \sum_{\alpha', j', m'} |j', m', \alpha'\rangle\langle j', m', \alpha'|$, to find

$$J_a|j, m, \alpha\rangle = \sum_{\alpha', j', m'} |j', m', \alpha'\rangle\langle j', m', \alpha'|J_a|j, m, \alpha\rangle.$$

Since J_a commutes with the total angular momentum operator and any other operators with quantum numbers represented by α, the state $J_a|j, m, \alpha\rangle$ has total angular momentum j and other quantum numbers α. Thus $\langle j', m', \alpha'|J_a|j, m, \alpha\rangle = \delta_{jj'}\delta_{\alpha\alpha'}[\pi_j(\tau_a)]_{m'm}$.

The commutation relations in equation (8.2) are then

$$\begin{aligned} &[J_+, \mathcal{J}_1] = 0, && [J_+, \mathcal{J}_0] = \mathcal{J}_1, && [J_+, \mathcal{J}_{-1}] = \mathcal{J}_0, \\ &[J_-, \mathcal{J}_1] = \mathcal{J}_0, && [J_-, \mathcal{J}_0] = \mathcal{J}_{-1}, && [J-, \mathcal{J}_{-1}] = 0, \\ &[J_3, \mathcal{J}_1] = \mathcal{J}_1, && [J_3, \mathcal{J}_0] = 0, && [J_3, \mathcal{J}_{-1}] = -\mathcal{J}_{-1}. \end{aligned} \tag{8.37}$$

These commutation relations show that $\mathcal{J}_k$ indeed satisfies equation (8.34) with $\ell = 1$.

Example 8.4.2. The position operator r_a is also a vector operator. To see this, we note that the total angular momentum is $J_a = L_a + S_a$, where L_a is orbital angular momentum and S_a is spin. The spin commutes with position, but the orbital angular momentum does not since $L_a = \epsilon_{abc} x_b p_c$ and the momentum p_c does not commute with position. We therefore have

$$[J_a, x_b] = [L_a, x_b] = \epsilon_{acd} x_c [p_d, x_b] = -i\epsilon_{acd} x_c \delta_{db} = -i\epsilon_{acb} x_c = x_c [\pi_{\text{adj}}(\tau_a)]_{cb}, \tag{8.38}$$

so $x_b = (x, y, z)$ is a tensor operator transforming in the adjoint representation. The analysis from the previous example then shows that we can change basis by defining a new vector r_k with

$$r_1 = -\frac{x + iy}{\sqrt{2}}, \qquad r_0 = z, \qquad r_{-1} = \frac{x - iy}{\sqrt{2}}, \tag{8.39}$$

which will satisfy equation (8.34).

Theorem 8.4.1 (Wigner–Eckart theorem)**.** *Let O^ℓ_k be a tensor operator satisfying equation* (8.34)*. Its matrix elements in angular momentum eigenstates satisfy*

$$\langle j, m, \alpha | O^\ell_k | j', m', \beta \rangle = C(j, m; \ell, k, j', m') \, \langle j, \alpha || O^\ell || j', \beta \rangle, \tag{8.40}$$

where $C(j, m; \ell, k, j', m')$ is a Clebsch–Gordan coefficient. The quantity $\langle j, \alpha || O^\ell || j', \beta \rangle$ is called a reduced matrix element.

The Wigner–Eckart theorem does not provide a way to compute the reduced matrix element, but it tells us two useful things:

- The Clebsch–Gordan coefficient $C(j, m; \ell, k, j', m')$, and therefore the matrix element $\langle j, m, \alpha | O^\ell_k | j', m', \beta \rangle$, vanishes unless $|j' - \ell| \leq j \leq j' + \ell$ and $m = k + m'$.
- Since the reduced matrix element $\langle j, \alpha || O^\ell || j', \beta \rangle$ does not depend on any of m, m' or k, for given (j, j', α, β), one only needs to compute the matrix elements for one assignment of (m, k) to determine the rest.

Before proving the Wigner–Eckart theorem let's look at an example application:

Example 8.4.3. The gross structure energy eigenstates of the hydrogen atom are labelled by their orbital angular momentum quantum numbers j, m and the energy level n, $|j, m, n\rangle$, with $0 \leq j \leq n - 1$. The rate of transitions R between two energy eigenstates $|j, m, n\rangle$ and $|j', m', n'\rangle$ is proportional to the sum of the squares of the matrix

elements of the position operator,

$$R \propto |\langle j,m,n|x|j',m',n'\rangle|^2 + |\langle j,m,n|y|j',m',n'\rangle|^2 + |\langle j,m,n|z|j',m',n'\rangle|^2. \tag{8.41}$$

We have seen in example 8.4.2 that z is the $k=0$ component of the spin-one operator r_k, while x and y are linear combinations of the $k=\pm 1$ components. In terms of r_k equation (8.41) reads

$$R \propto \sum_{k=-1}^{1} |\langle j,m,n|r_k|j',m',n'\rangle|^2 \tag{8.42}$$

The Wigner–Eckart theorem tells us that

$$\langle j,m,n|r_k|j',m',n'\rangle = C(j,m;1,k,j',m')\,\langle j,n||r||j',n'\rangle, \tag{8.43}$$

for some reduced matrix element $\langle j,n||r||j',n'\rangle$. The Clebsch–Gordan coefficients will all vanish except for those $|j'-1| \le j \le j'+1$, giving us a selection rule that transitions cannot happen in which j jumps by more than one up or down. This also tells us that we cannot have a transition between two $j=0$ eigenstates. We also have that $C(j,m;1,k,j',m')=0$ unless $m=k+m'$, and since $-1\le k\le 1$ this tells us that m can also jump by at most one quantum up or down.

To illustrate how the Wigner–Eckart theorem can be useful for calculations, suppose we want to study the transition rate between the $|1,m,2\rangle$ energy eigenstate and the $|0,0,1\rangle$ eigenstate (i.e. the transition rate between the 2p and 1s orbitals). In principle there are nine matrix elements of the form $\langle 1,m,2|r_k|0,0,1\rangle$ that we need to compute, since there are three possible values of m and three possible values of k. Thanks to the Wigner–Eckart theorem we only need to compute one of these to determine the reduced matrix element $\langle 1,2||r||0,1\rangle$, and then we can determine the rest as long as we know the Clebsch–Gordan coefficients, which we can look up in a table.

Proof of the Wigner–Eckart theorem. Let $O^{j_1}_{m_1}$ be a spin-j_1 tensor operator that satisfies equation (8.34) (with ℓ and k replaced by j_1 and m_1) and let $|j_2,m_2,\beta\rangle$ be an angular momentum eigenstate, with β any other quantum numbers carried by the state. We have

$$\begin{aligned} J_3 O^{j_1}_{m_1}|j_2,m_2,\beta\rangle &= \Big([J_3,O^{j_1}_{m_1}] + O^{j_1}_{m_1}J_3\Big)\,|j,m,\beta\rangle + O^{j_1}_{m_1}J_3|j_2,m_2,\beta\rangle \\ &= (m_1+m_2)O^{j_1}_{m_1}|j_2,m_2,\beta\rangle. \end{aligned} \tag{8.44}$$

This tells us that the state $O^{j_1}_{m_1}|j,m,\alpha\rangle$ is an eigenstate of J_3, with eigenvalue $k+m$. We also have that

$$\begin{aligned} J_+ O^{j_1}_{m_1}|j_2,m_2,\beta\rangle &= \Big([J_+,O^{j_1}_{m_1}] + O^{j_1}_{m_1}J_+\Big)\,|j_2,m_2,\beta\rangle \\ &= c(j_1,m_1+1)\,O^{j_1}_{m_1+1}|j_2,m_2,\beta\rangle + c(j_2,m_2+1)\,O^{j_1}_{m_1}|j_2,m_2+1,\beta\rangle, \end{aligned} \tag{8.45}$$

and

$$\begin{aligned} J_- O^{j_1}_{m_1}|j_2,m_2,\beta\rangle &= \Big([J_-,O^{j_1}_{m_1}] + O^{j_1}_{k}J_-\Big)\,|j_2,m_2,\beta\rangle \\ &= c(j_1,m_1)\,O^{j_1}_{m_1-1}|j_2,m_2,\beta\rangle + c(j_2,m_2)\,O^{j_1}_{m_1}|j_2,m_2-1,\beta\rangle, \end{aligned} \tag{8.46}$$

from which we see that $J_\pm O^{j_1}_{m_1}|j_2,m_2,\alpha\rangle$ is an eigenstate J_3 with eigenvalue $m_1+m_2\pm 1$.

The way J_3 and $J_\pm$ act on $O^{j_1}_{m_1}|j_2, m_2, \beta\rangle$ is exactly the same as the way they act on tensor product states $|j_1, m_1\rangle|j_2, m_2\rangle$. We can thus repeat the same analysis that we performed in section 8.3 when building tensor product representations, just replacing each instance of $|j_1, m_1\rangle$ with $O^{j_1}_{m_1}$ and keeping track of the additional quantum numbers β. We start with the state with the largest possible eigenvalue of J_3, which is $j = j_1 + j_2$,

$$O^{j_1}_{j_1}|j_2, j_2, \beta\rangle. \tag{8.47}$$

This is a state with total angular momentum and J_3 eigenvalue both equal to j. However, it may not have definite values of the other quantum numbers. For example applying the position operator to an energy eigenstate usually doesn't produce another energy eigenstate since the position operator usually doesn't commute with the Hamiltonian. Instead, the state in equation (8.47) may be a superposition of states with different quantum numbers,

$$O^{j_1}_{j_1}|j_2, j_2, \beta\rangle = \sum_\alpha \langle j, \alpha||O^{j_1}||j_2, \beta\rangle|j, j, \alpha\rangle \tag{8.48}$$

with expansion coefficents given by the matrix elements

$$\langle j, \alpha||O^{j_1}||j_2, \beta\rangle \equiv \langle j, j, \alpha|O^{j_1}_{j_1}|j_2, j_2, \beta\rangle. \tag{8.49}$$

We repeatedly apply J_- to produce states with lower and lower J_3 eigenvalues, all the way down to $-j$. For example, if we apply J_- to either side of equation (8.48) we find

$$\sqrt{\frac{j_1}{j}}\, O^{j_1}_{j_1-1}|j_2, j_2, \beta\rangle + \sqrt{\frac{j_2}{j}}\, O^{j_1}_{j_1}|j_2, j_2 - 1, \beta\rangle = \sum_\alpha \langle j, \alpha||O^{j_1}||j_2, \beta\rangle|j, j-1, \alpha\rangle. \tag{8.50}$$

The left-hand side is the state in equation (8.22), with $|j_1, m_1\rangle$ replaced by $|j_2, m_2\rangle$. Crucially, the expansion coefficients $\langle j, \alpha||O^{j_1}||j_2, \beta\rangle$ are the same matrix elements appearing as in equation (8.21), which is guaranteed by the fact that the algebra of acting with $J_\pm$ on $O^{j_1}_{j_1}|j_2, j_2, \beta\rangle$ is exactly the same as the algebra acting on tensor product states.

We can construct an orthogonal linear combination analogous to (8.24), which has total spin and J_3 eigenvalue both equal to $j - 1$,

$$\sqrt{\frac{j_2}{j}}\, O^{j_1}_{j_1-1}|j_2, j_2, \beta\rangle - \sqrt{\frac{j_1}{j}}\, O^{j_1}_{j_1}|j_2, j_2-1, \beta\rangle = \sum_\alpha \langle j-1, \alpha||O^{j_1}||j_2, \beta\rangle|j-1, j-1, \alpha\rangle, \tag{8.51}$$

with some new expansion coefficents which we denote $\langle j-1, \alpha||O^{j_1}||j_2, \beta\rangle$. We then apply J_- to this state to create new states with lower J_3 eigenvalues.

Since the algebra of all of this is the same as for the tensor product of two states, the coefficients appearing in the states created by applying lowering operators to $O^{j_1}_{m_1}|j_2, m_2, \beta\rangle$ are the same as those obtained by applying lowering operators to $|j_1, m_1\rangle|j_2, m_2\rangle$, namely the Clebsch–Gordan coefficients. We thus have an analogous equation to equation (8.28), the only difference being the need to include a sum over the other quantum numbers,

$$\sum_{m_1=-j_1}^{j_1} \sum_{m_2=-j_2}^{j_2} C(j, m; j_1, m_1, j_2, m_2) O^{j_1}_{m_1}|j_2, m_2, \beta\rangle = \sum_\alpha \langle j, \alpha||O^{j_1}||j_2, \beta\rangle|j, m, \alpha\rangle, \tag{8.52}$$

where j now denotes any possible value of the total angular momentum, not necessarily $j_1 + j_2$. Equation (8.52) implies the following equation, analogous to equation (8.53),

$$O^{j_1}_{m_1}|j_2, m_2, \beta\rangle = \sum_{\alpha'} \sum_{j'=0}^{\infty} \sum_{m'=-j'}^{j'} C(j', m'; j_1, m_1, j_2, m_2)\, \langle j', \alpha'||O^{j_1}||j_2, \beta\rangle\, |j', m', \alpha'\rangle. \tag{8.53}$$

Acting on both sides with $\langle j, m, \alpha|$, we find

$$\langle j, m, \alpha|O^{j_1}_{m_1}|j_2, m_2, \beta\rangle = C(j, m, j_1, m_1, j_2, m_2)\langle j, \alpha||O^{j_1}||j_2, \beta\rangle, \tag{8.54}$$

which is the Wigner–Eckart theorem (8.40). □

Chapter 9

Representation theory of $\mathfrak{su}(3)$

9.1 Introduction

The Lie algebra $\mathfrak{su}(3)$ is eight-dimensional. A conventional basis is $T_a = \frac{1}{2}\lambda_a$, where λ_a are the Gell-Mann matrices,

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad (9.1)$$
$$\lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \quad \lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad \lambda_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}.$$

The generators satisfy the algebra

$$[T_a, T_b] = i f_{abc} T_c, \quad (9.2)$$

with structure constants that are completely antisymmetric in all three indices. The non-zero structure constants are

$$f_{123} = 1, \qquad f_{147} = f_{165} = f_{246} = f_{257} = f_{345} = f_{376} = \frac{1}{2}, \qquad f_{458} = f_{678} = \frac{\sqrt{3}}{2}, \quad (9.3)$$

along with those related to these by the complete antisymmetry of f_{abc}.

To construct the irreducible representations of $\mathfrak{su}(2)$ we diagonalised the representation J_3 of one of the generators and constructed linear combinations $J_\pm$ of the representations of the other two generators that raised and lowered the eigenvalues of J_3. We will adopt a similar strategy for $\mathfrak{su}(3)$, simultaneously diagonalising as many generators as possible and using the rest of the generators to construct raising and lowering operators. This will be more complicated than for $\mathfrak{su}(2)$ simply because we have more generators. However, this basic strategy works for any compact, semisimple Lie algebra, so the methods we develop for $\mathfrak{su}(3)$ in this chapter serve as a prototype for the representation theory of many other algebras. If you are interested in reading more about the application of these methods to other Lie algebras, we recommend Georgi's textbook [3].

Since we want to diagonalise the representations of as many generators as possible, and since matrices need to commute in order to be simultaneously diagonalisable, this motivates the definition:

Definition 9.1.1. A **Cartan subalgebra** $\mathfrak{h}$ of a semisimple Lie algbra $\mathfrak{g}$ is a maximal set of mutually commuting elements of $\mathfrak{g}$. The dimension of any Cartan subalgebra is called the **rank** of $\mathfrak{g}$.

Theorem 9.1.1. *The rank of* $\mathfrak{su}(3)$ *is two. The generators* T_3 *and* T_8 *form the basis of a Cartan subalgebra.*

Proof. The elements of $\mathfrak{su}(3)$ are traceless, Hermitian matrices. Since they are Hermitian they are diagonalisable. Since the elements of a Cartan subalgebra commute with each other, they are mutually diagonalisable. After diagonalisation, the elements of the Cartan subalgebra are traceless, real 3×3 matrices. Such matrices are fully specified by two of their diagonal elements, the third diagonal element being fixed by tracelessness. Therefore the space of such matrices is two-dimensional and so the rank of $\mathfrak{su}(3)$ is two.

Since the generators T_3 and T_8 are diagonal, any linear combination of them will also be diagonal and therefore commute with any other linear combination. Thus T_3 and T_8 form a basis for a mutually commuting two-dimensional subalgebra. Since we have already shown that the rank of $\mathfrak{su}(3)$ is two, we know that this subalgebra is maximal, and is therefore a Cartan subalgebra. □

Our strategy will be to diagonalise the representation matrices of T_3 and T_8. The Cartan subalgebra is not unique, for example there is another Cartan subalgebra spanned by T_1 and T_8. Diagonalising a different subalgebra will give equivalent results, the same way we could just as well have constructed representations of $\mathfrak{su}(2)$ by diagonalising J_1 or J_2 instead of J_3.

9.2 Weights and roots

Let π be a representation of $\mathfrak{su}(3)$. We define

$$H_1 = \pi(T_3), \qquad H_2 = \pi(T_8), \tag{9.4}$$

the matrices we want to simultaneously diagonalise. We will seek unitary representations, so the H_i will be Hermitian matrices.

Definition 9.2.1. For a given representation, a **weight** is a pair $\mu = (\mu_1, \mu_2)$ of eigenvalues of the Cartan generators H_i corresponding to the same eigenvector $|\mu\rangle$,

$$H_i|\mu\rangle = \mu_i|\mu\rangle, \tag{9.5}$$

Because the H_i are Hermitian, the weights are real. We call the eigenvector $|\mu\rangle$ a **weight vector**.

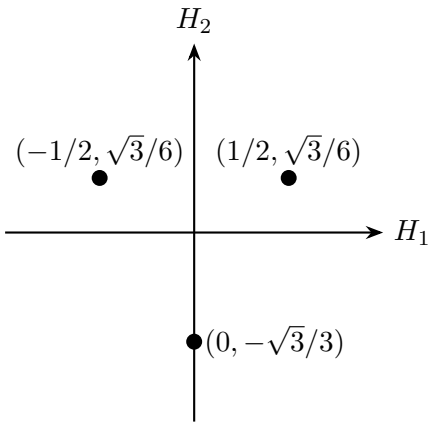
Example 9.2.1. In the fundamental representation we have

$$H_1 = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad H_2 = \frac{\sqrt{3}}{6}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}. \tag{9.6}$$

There are three weights, $\mu^1 = (1/2, \sqrt{3}/6)$, $\mu^0 = (-1/2, \sqrt{3}/6)$, and $\mu^{-1} = (0, -\sqrt{3}/3)$, corresponding to the eigenvectors

$$|\mu^1\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \qquad |\mu^0\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \qquad |\mu^{-1}\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \tag{9.7}$$

We will frequently plot the weights of a representation in a diagram, with the eigenvalue of the H_1 on the horizontal axis and the eigenvalue of H_2 on the vertical axis. The weight diagram for the fundamental representation is:



We see that the weights of the fundamental representation form the corners of an equilateral triangle.

The weights in the adjoint representation turn out to be particularly important. Recall that in the adjoint representation a generator T_a is represented by a matrix with elements proportional to the structure constants,

$$[\pi_{\text{adj}}(T_a)]_{bc} = -if_{abc} \tag{9.8}$$

Because the structure constants of $\mathfrak{su}(3)$ are real and totally antisymmetric, we have that

$$[\pi_{\text{adj}}(T_a)^\dagger]_{bc} = ([\pi_{\text{adj}}(T_a)]_{cb})^* = if_{acb} = -if_{abc} = [\pi_{\text{adj}}(T_a)]_{bc}, \tag{9.9}$$

so the representation matrices in the adjoint representation are Hermitian, which by definition means that the adjoint representation is unitary. For each generator T_a we define a basis state $|T_a\rangle$ for the vector space acted on by the adjoint representation. An arbitrary Lie algebra element $Y = y_a T_a$ corresponds to the state $|Y\rangle = |y_a T_a\rangle = y_a |T_a\rangle$. The action of the adjoint representation on the basis states is

$$\begin{aligned} \pi_{\text{adj}}(T_a)|T_b\rangle &= |T_c\rangle[\pi_{\text{adj}}(T_a)]_{cb} = -if_{acb}|T_c\rangle = if_{abc}|T_c\rangle = |if_{abc}T_c\rangle \\ &= |[T_a, T_b]\rangle. \end{aligned} \tag{9.10}$$

The action of $\pi_{\text{adj}}(X)$ on $|Y\rangle$ for arbitrary $X, Y \in \mathfrak{g}$ is then determined through linearity,

$$\pi_{\text{adj}}(X)|Y\rangle = |[X, Y]\rangle. \tag{9.11}$$

Definition 9.2.2. The non-zero weights in the adjoint representation of a compact semisimple Lie algebra are called **roots** and denoted by α. The eigenstate corresponding to a root α is denoted $|E_\alpha\rangle$:

$$H_i|E_\alpha\rangle = \alpha_i|E_\alpha\rangle \tag{9.12}$$

Since the states transforming with the adjoint representation can be identified with Lie algebra elements, there should be some matrix E_α which according to equation (9.12) must satisfy[1]

$$\alpha_i|E_\alpha\rangle = H_i|E_\alpha\rangle = |[H_i, E_\alpha]\rangle, \tag{9.13}$$

from which we read off an equivalent definition of the roots as the eigenvalues for which we can find solutions to

$$[H_i, E_\alpha] = \alpha_i E_\alpha. \tag{9.14}$$

We call E_α the generator corresponding to α.

Before finding the roots explicitly, we first derive some properties of the roots, that also apply to roots of other compact semisimple Lie algberas.

Theorem 9.2.1. *The roots obey the following properties:*

(i) If α is a root, then $-\alpha$ is also a root. Their generators obey $E_{-\alpha} = E_\alpha^\dagger$.

(ii) In any representation, if μ is a weight and $|\mu\rangle$ its corresponding weight vector, then either $E_\alpha|\mu\rangle$ is weight vector with weight $\mu + \alpha$, or $E_\alpha|\mu\rangle = 0$.

(iii) If α and β are distinct roots, then either $[E_\alpha, E_{-\beta}] \propto E_{\alpha-\beta}$ or $\alpha - \beta$ is not a root in which case $[E_\alpha, E_{-\beta}] = 0$.

Point (ii) tells us that E_α and $E_{-\alpha}$ act to "raise" and "lower" the weights by the root α. They will play a similar role to $J_\pm$ in the representation theory of $\mathfrak{su}(2)$.

Proof. To prove point (i) we take the Hermitian conjugate of equation (9.14), to find

$$[H_i, E_\alpha^\dagger] = -\alpha_i E_\alpha^\dagger. \tag{9.15}$$

This tells us that unless $\alpha = 0$, E_α is not a Hermitian matrix. This is similar to $J_\pm$ in $\mathfrak{su}(2)$, which are not Hermitian but nevertheless allow us to construct unitary representations.

To establish point (ii) we use equation (9.14) to find

$$H_i E_\alpha|\mu\rangle = (E_\alpha H_i + [H_i, E_\alpha])|\mu\rangle = (\mu + \alpha)E_\alpha|\mu\rangle. \tag{9.16}$$

Thus either $E_\alpha|\mu\rangle$ is an eigenvector of H_i with eigenvalue $\mu + \alpha$, or $E_\alpha|\mu\rangle = 0$.

Point (iii) is the special case of point (ii) applied to roots. Applying E_α to $|E_{-\beta}\rangle$ raises the root $-\beta$ by α. In other words, either $E_\alpha|E_{-\beta}\rangle = 0$ if $\alpha - \beta$ is not a root, or otherwise

[1] We are abusing notation slightly here, using H_i and E_α to denote both the Lie algebra elements and their representation matrices in an arbitrary representation.

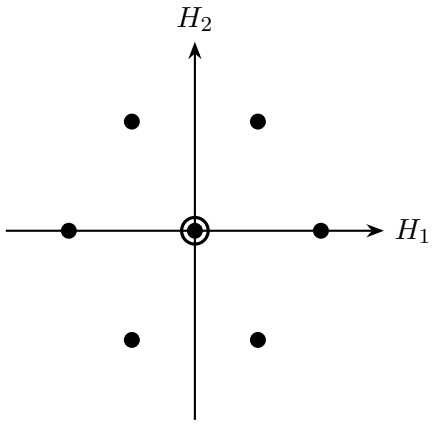
$E_\alpha|E_{-\beta}\rangle \propto |E_{\alpha-\beta}\rangle$. But $E_\alpha|E_{-\beta}\rangle = |[E_\alpha, E_{-\beta}]\rangle$. So we have that either $[E_\alpha, E_{-\beta}] = 0$ if $\alpha - \beta$ is not a root, or $[E_\alpha, E_{-\beta}] \propto E_{\alpha-\beta}$. □

To find the roots of $\mathfrak{su}(3)$ we could explicitly construct H_1 and H_2 in the adjoint representation using the structure constants in equation (9.3) and find their eigenvalues and eigenvectors. However, since the adjoint representation consists of 8×8 matrices this would be a little tedious. Instead we will simply state the roots and the corresponding generators, and then it can be checked that they satisfy equation (9.14). They are

$$\begin{aligned} &\alpha = (\pm 1, 0), && E_{(\pm 1,0)} = \frac{1}{\sqrt{2}}\left(T_1 \pm iT_2\right), \\ &\alpha = (\pm 1/2, \pm\sqrt{3}/2), && E_{(\pm 1/2,\pm\sqrt{3}/2)} = \frac{1}{\sqrt{2}}\left(T_4 \pm iT_5\right), \\ &\alpha = (\mp 1/2, \pm\sqrt{3}/2) && E_{(\mp 1/2,\pm\sqrt{3}/2)} = \frac{1}{\sqrt{2}}(T_6 \pm iT_7), \end{aligned} \tag{9.17}$$

where the $\pm$ signs on each line are correlated.

The roots of $\mathfrak{su}(3)$, form a regular hexagon. Plotted in a weight diagram along with the two vanishing weights, we have:



9.3 Complex conjugate representations

Suppose π is a representation of $\mathfrak{su}(3)$, so that $[\pi(T_a), \pi(T_b)] = if_{abc}\pi(T_c)$. Since the structure constants are real, if we take the complex conjugate of this relation we find

$$[\pi(T_a)^*, \pi(T_b)^*] = -if_{abc}\pi(T_c)^*. \tag{9.18}$$

Defining $\bar{\pi}(T_a) = -\pi(T_a)^*$, we find that

$$[\bar{\pi}(T_a), \bar{\pi}(T_b)] = if_{abc}\bar{\pi}(T_c), \tag{9.19}$$

so we see that $\bar{\pi}$ is also a representation of $\mathfrak{su}(3)$.

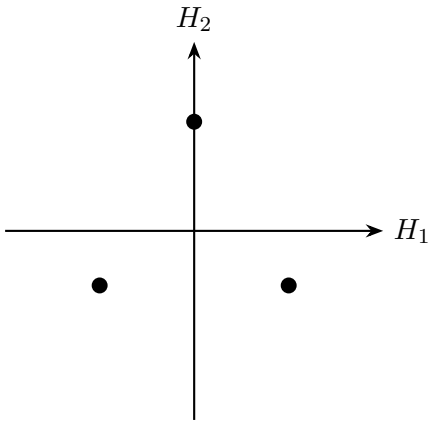
Definition 9.3.1. Given a representation π of $\mathfrak{su}(3)$, the **complex conjugate** representation is defined by $\bar{\pi}(X) = -\pi(X)^*$ for all $X \in \mathfrak{su}(3)$.

In the unitary representations that we are seeking to construct, H_i is Hermitian and therefore has the same eigenvalues as H_i^*. Therefore the weights in a complex conjugate representation $\bar{\pi}$ are minus the weights in the representation π.

Example 9.3.1. The complex conjugate of the fundamental representation is often known in physics as the **anti-fundamental representation**. The representation matrices of the antifundamental are $\bar{T}_a = -\frac{1}{2}\lambda_a^*$. In particular the Cartan generators in this representation are

$$H_1 - \bar{T}_3 = \frac{1}{2}\begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad H_2 = -\bar{T}_8 = \frac{\sqrt{3}}{6}\begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{pmatrix}. \tag{9.20}$$

From this we read off the weights $(-1/2, -\sqrt{3}/6)$, $(1/2, -\sqrt{3}/6)$ and $(0, \sqrt{3}/3)$. The weight diagram for the anti-fundamental representation is:



In physics, we often denote representations of groups by their dimension written in bold. For example, the fundamental representation of $\mathfrak{su}(3)$ is three dimensional and therefore denoted **3**. The anti-fundamental representation is also three-dimensional. It is typically denoted $\bar{\mathbf{3}}$, where the bar is supposed to remind you of conjugation. As another example, the adjoint representation of $\mathfrak{su}(3)$ is eight-dimensional, so is known as the **8**.

9.4 $\mathfrak{su}(2)$ subalgebras and the master formula

For any root α, the generators E_α defined in equation (9.17) satisfy

$$[\alpha \cdot H, E_{\pm\alpha}] = \pm|\alpha|^2 E_\alpha, \qquad [E_\alpha, E_{-\alpha}] = \alpha \cdot H. \tag{9.21}$$

If we define $E_3 = |\alpha|^{-2}\alpha \cdot H$ and $E_\pm = |\alpha|^{-1}E_{\pm\alpha}$, then these relations become

$$[E_3, E_\pm] = \pm E_\pm, \qquad [E_+, E_-] = E_3, \tag{9.22}$$

This is the same as the algebra of $\mathfrak{su}(2)$ raising and lowering operators written in equation (8.2). Thus each root α is associated to an $\mathfrak{su}(2)$ subalgebra. We can thus use our knowledge of the representation theory of $\mathfrak{su}(2)$ to learn things about $\mathfrak{su}(3)$.

Suppose that π is a finite-dimensional representation of $\mathfrak{su}(3)$. Then then the matrices E_3 and $E_\pm$ in this representation form a finite-dimensional representation of $\mathfrak{su}(2)$. If μ is a weight in this $\mathfrak{su}(3)$ representation and $|\mu\rangle$ its corresponding weight vector, then we have that

$$E_3|\mu\rangle = \frac{\alpha \cdot H}{|\alpha|^2}|\mu\rangle = \frac{\alpha \cdot \mu}{|\alpha|^2}|\mu\rangle. \tag{9.23}$$

Since we know from the representation theory of $\mathfrak{su}(2)$ that the eigenvalues of E_3 must be integers or half-integers, we learn that

$$\frac{2\alpha \cdot \mu}{|\alpha|^2} \tag{9.24}$$

is an integer.

We can decompose $|\mu\rangle$ as a linear combination of states with definite spin (i.e. states transforming in irreducible representations of $\mathfrak{su}(2)$). Let j be the largest spin appearing in this decomposition. We can repeatedly apply the raising operator E_+ to $|\mu\rangle$, raising its E_3 eigenvalue by one each time, until we eventually get zero. In other words, there is some non-negative integer p such that

$$E_+^p|\mu\rangle \neq 0, \qquad \text{but} \qquad E_+^{p+1}|\mu\rangle = 0. \tag{9.25}$$

The state $E_+^p|\mu\rangle$ must have the largest possible E_3 eigenvalue, which is j. Hence

$$\frac{\alpha \cdot \mu}{|\alpha|^2} + p = j. \tag{9.26}$$

Similarly we can repeatedly apply E_- to $|\mu\rangle$, lowering the E_3 eigenvalue by one each time, until we reach the smallest possible E_3 eigenvalue, which is $-j$. In other words there is a non-negative integer q such that

$$E_-^q|\mu\rangle \neq 0, \qquad \text{but} \qquad E_-^{q+1}|\mu\rangle = 0, \tag{9.27}$$

such that

$$\frac{\alpha \cdot \mu}{|\alpha|^2} - q = -j. \tag{9.28}$$

Adding up equations (9.26) and (9.28), we obtain the **master formula**,

$$\frac{2\alpha \cdot \mu}{|\alpha|^2} = q - p, \tag{9.29}$$

for two non-negative integers p and q.

9.5 Positive weights

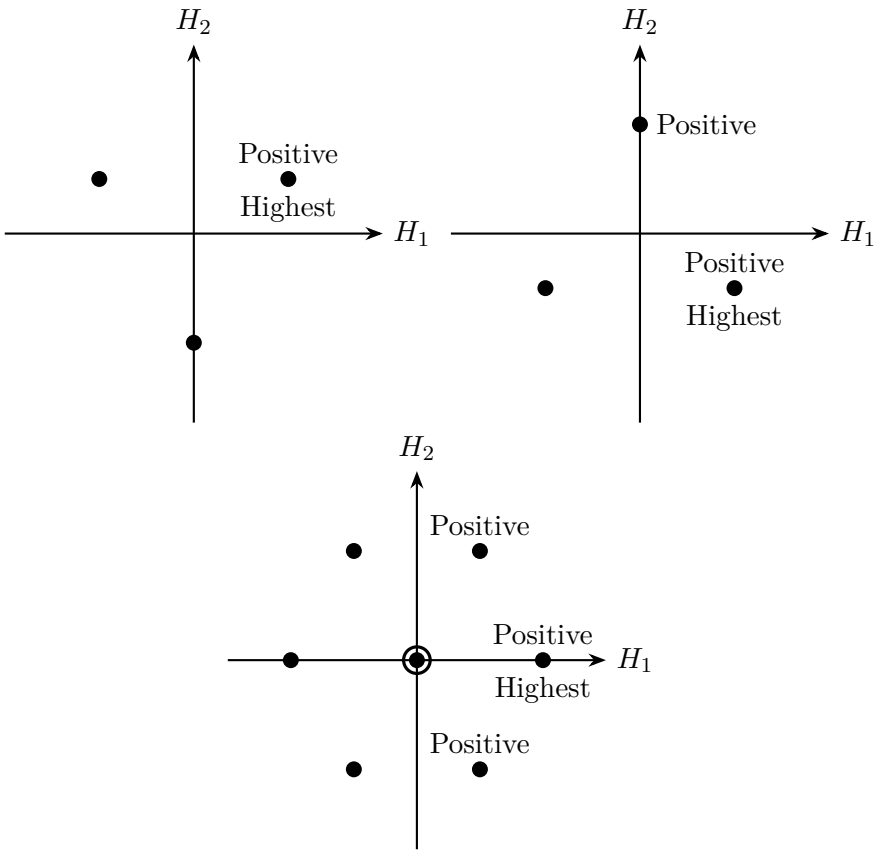
Irreducible representations of $\mathfrak{su}(2)$ are labelled by the highest eigenvalue of J_3. We will end up labelling irreducible representations by their highest eigenvalue of H_i, i.e. by their highest weight. Since the weights lie in a two-dimensional space, we will need to define what we mean by highest.

Definition 9.5.1. We define a weight $\mu = (\mu_1, \mu_2)$ to be **positive** either if $\mu_1 > 0$, or if $\mu_1 = 0$ and $\mu_2 > 0$. A weight that is not positive and non-zero is **negative**.

Example 9.5.1. For the fundamental representation, $(1/2, \sqrt{3}/6)$ is positive, while $(-1/2, \sqrt{3}/6)$ and $(0, -\sqrt{3}/3)$ are negative. Turning to the adjoint representation, the positive roots are $(1, 0)$ and $(1/2, \pm\sqrt{3}/2)$.

Definition 9.5.2. Given two weights μ and ν, we say that μ is **higher** than ν if $\mu - \nu$ is positive. For a given representation, a weight is **highest** if it is higher than all of the other weights in that representation.

The highest weight will therefore be the weight furthest to the right in the weight-diagram, with vertical position used as the tie-breaker. Here are the weight diagrams for the fundamental, anti-fundamental, and adjoint representations with their positive and highest weights marked:

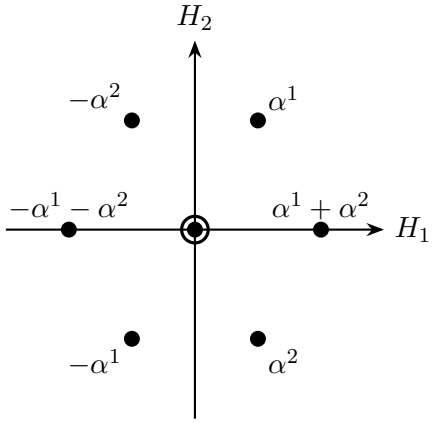


9.6 Simple roots and fundamental weights

Definition 9.6.1. The **simple roots** of $\mathfrak{su}(3)$ are

$$\alpha^1 = (1/2, \sqrt{3}/2), \qquad \alpha^2 = (1/2, -\sqrt{3}/2). \tag{9.30}$$

These are two of the positive roots. The special thing about them is that their sum gives the other positive root, $\alpha^1 + \alpha^2 = (1, 0)$. Therefore every positive root takes the form $\alpha = c_1\alpha^1 + c_2\alpha^2$ with non-negative integers c_i. Similarly every negative root takes the same form with non-positive integers c_i:



A property of the simple roots that will be useful for later:

Theorem 9.6.1. *The simple roots satisfy*

$$[E_{\alpha^1}, E_{-\alpha_2}] = [E_{-\alpha_1}, E_{\alpha^2}] = 0. \tag{9.31}$$

Proof. This follows from point (iii) of theorem 9.2.1, since $\alpha^1 - \alpha^2$ is not a root. □

Definition 9.6.2. The **Cartan matrix** A is the matrix whose elements take the form

$$A_{ji} = \frac{2\alpha^i \cdot \alpha^j}{|\alpha^j|^2}. \tag{9.32}$$

Explicitly, for $\mathfrak{su}(3)$ the Cartan matrix is

$$A = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}. \tag{9.33}$$

Definition 9.6.3. The **fundamental weights** of $\mathfrak{su}(3)$ are the highest weight μ^1 of the fundamental representation and the highest weight μ^2 of the anti-fundamental representation,

$$\mu^1 = (1/2, \sqrt{3}/6), \qquad \mu^2 = (1/2, -\sqrt{3}/6). \tag{9.34}$$

Theorem 9.6.2. *If μ is a weight in any representation of $\mathfrak{su}(3)$, then*

$$\mu = \sum_{i=1}^{2} c_i \mu^i \tag{9.35}$$

with integer coefficients c_i, where μ^i are the fundamental weights.

Proof. First we note that weights are two-component vectors, and so since the two vectors μ^1 and μ^2 are linearly independent we can use them as a basis and write any weight as in equation (9.35) with coefficients c_i. Our task is to prove that the c_i are integers. Explicit computation shows that the simple roots and fundamental weights satisfy

$$\frac{2\alpha^1 \cdot \mu^1}{|\alpha^1|^2} = \frac{2\alpha^2 \cdot \mu^2}{|\alpha^2|^2} = 1, \qquad \frac{2\alpha^1 \cdot \mu^2}{|\alpha^1|^2} = \frac{2\alpha^2 \cdot \mu^1}{|\alpha^2|^2} = 0, \tag{9.36}$$

or more concisely

$$\frac{2\alpha^i \cdot \mu^j}{|\alpha^i|^2} = \delta_{ij}. \tag{9.37}$$

Then if μ has an expansion as in equation (9.35) for some coefficients c_i, this implies that

$$\frac{2\alpha^i \cdot \mu}{|\alpha^i|^2} = c_i. \tag{9.38}$$

The master formula (9.29) tells us that the left-hand side is an integer. Therefore c_i must be an integer. □

9.7 Irreducible representations of $\mathfrak{su}(3)$

We now have all of the ingredients that we need to construct the irreducible representations of $\mathfrak{su}(3)$. Suppose μ is the highest weight of such a representation. Its weight vector must sastisfy $E_\alpha|\mu\rangle = 0$ for any positive root α, as otherwise $\mu + \alpha$ would be another weight that is higher than μ. Therefore we can construct states by repeatedly applying lowering operators,

$$E_{-\alpha^{i_1}} E_{-\alpha^{i_2}} \cdots E_{-\alpha^{i_n}} |\mu\rangle, \tag{9.39}$$

where α^i are positive roots, so that $-\alpha^i$ are negative roots. Such a state has weight

$$\mu - \alpha^{i_1} - \alpha^{i_2} - \cdots - \alpha^{i_n} \tag{9.40}$$

It is sufficient to apply only generators corresponding to negative roots in equation (9.39), because if there was any generator E_β corresponding to some positive root β appearing in the string of generators in equation (9.39) we could use the $\mathfrak{su}(3)$ commutation relations to move it to the right, where it would annihilate the highest weight state $|\mu\rangle$.

It is also sufficient to take the α^i in equation (9.39) to be the simple roots α^1 and α^2. The reason is that the generator of the other positive root satisfies $E_{\alpha^1+\alpha^2} = \sqrt{2}[E_{\alpha^1}, E_{\alpha^2}]$. Taking the Hermitian conjugate of this equation and using $E_\alpha^\dagger = E_{-\alpha}$ we obtain

$$E_{-\alpha^1-\alpha^2} = -\sqrt{2}[E_{-\alpha^1}, E_{-\alpha^2}]. \tag{9.41}$$

Hence, any state of the form in equation (9.39) that contains factors of $E_{-\alpha^1-\alpha^2}$ can be rewritten as a linear combination of states containing only $E_{-\alpha^1}$ and $E_{-\alpha^2}$.

We can therefore construct a representation of $\mathfrak{su}(3)$ with a given highest weight by repeated application of the lowering operators corresponding to the simple roots, $E_{-\alpha^1}$ and $E_{-\alpha^2}$. Any representation we construct in this way will be irreducible since the states will be related to each other by application of raising and lowering operators, so there

will be no invariant subspaces. The main task is to determine which states of the form in equation (9.39) are non-zero. Here's how it works.

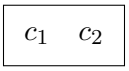
The highest weight of a a given representation must, according to theorem 9.6.2, be of the form $\mu = c_1\mu^1 + c_2\mu^2$ for some integers c_1 and c_2. We denote the representation (c_1, c_2). According to the master formula, the number of times q_i that we can lower the highest weight state $|\mu\rangle$ by the simple root α^i is given by $q_i = c_i$, where we use that $p_i = 0$ for this state since by definition $E_{\alpha^i}|\mu\rangle = 0$. Note that since $q_i \geq 0$, the highest weight state must have coefficients $c_i \geq 0$.

Suppose that q_1 is non-zero. Then we can lower the state by applying $E_{-\alpha^1}$ at least once, creating a new state $E_{-\alpha^1}|\mu\rangle$ with weight $\mu - \alpha^1$. Applying the master formula to $\mu - \alpha^1$ tells us the values of $q_i - p_i$ for this new state. But we know that $p_1 = 1$ since raising by E_{α^1} just brings us back to the highest weight state, while $p_2 = 0$ since $E_{\alpha^2}E_{-\alpha^1}|\mu\rangle = E_{-\alpha^1}E_{\alpha^2}|\mu\rangle = 0$. Therefore we can work out the values of q_i, and if any of them are non-zero we know we can apply a lowering operator to get a new state. We continue on in this way. Each time we apply the lowering operator $E_{-\alpha^i}$ to a state, the master formula tells us that

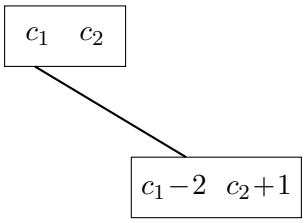
$$q_j - p_j \to q_j - p_j - A_{ji}, \tag{9.42}$$

where A_{ji} are the entries of the Cartan matrix. We know the p_j since we know how the state was constructed by applying lowering operators to the highest weight state, so then we can work out the q_j.

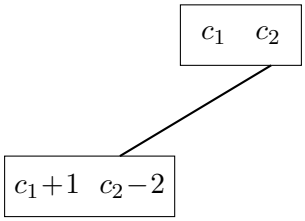
This procedure is best illustrated with examples. We will adopt a diagrammatic notation. A weight of the form $\mu = c_1\mu^1 + c_2\mu^2$ is represented by a box containing the coefficients of the fundamental weights:



Lowering by α^1 is represented by a diagonal line beginning under the first entry in the box and connecting to a box corresponding to the weight $\mu - \alpha^1 = (c_1 - 2)\mu^1 + (c_2 + 1)\mu^2$:

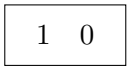


Lowering by α^2 is represented by a diagonal line under the second entry, connecting to a box corresponding to the weight $\mu - \alpha^2 = (c_1 + 1)\mu^1 + (c_2 - 2)\mu^2$:

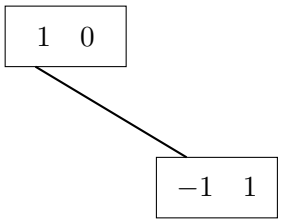


Example 9.7.1. Let's see how this procedure works to reconstruct the representation $(1,0)$. Since the highest weight of the representation is μ^1, we expect that the procedure should recover the fundamental representation.

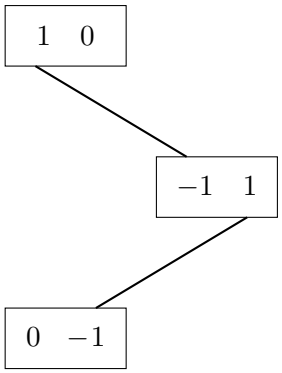
We start with the highest weight:



From this we read off that $q_1 = 1$, so we can apply $E_{-\alpha^1}$ to the highest state once to get a non-zero state, while $q_2 = 0$ so $E_{-\alpha^2}$ annihilates the highest weight state. Lowering by α^1, we obtain:



The new state has $q_1 - p_1 = -1$ and $q_2 - p_2 = 1$. But we also know that $p_1 = 1$ and $p_2 = 0$ since the state is constructed by $E_{-\alpha^1}$ acting on the highest weight state. We therefore have that $q_1 = 0$ and $q_2 = 1$, so we can lower once by α^2:

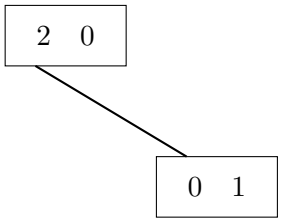


The state at the bottom has $q_1 - p_1 = 0$ and $q_2 - p_2 = -1$. From how it was constructed from the states above it we have that $p_1 = 0$ and $p_2 = 1$. We then read off that $q_1 = q_2 = 0$, so we cannot lower this state any more and we are done. This representation therefore has three weights,

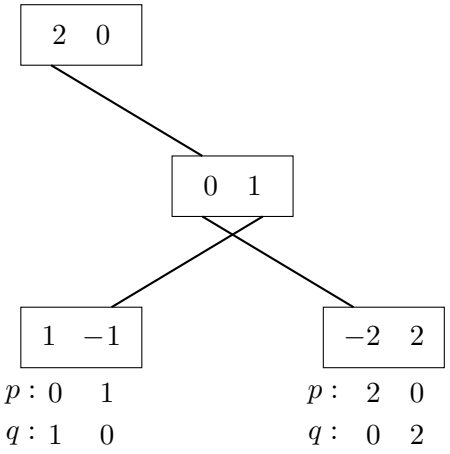
$$\begin{aligned} \mu^1 &= (1/2, \sqrt{3}/6), \\ \mu^2 - \mu^1 &= (0, -\sqrt{3}/3), \\ -\mu^2 &= (-1/2, \sqrt{3}/6). \end{aligned} \tag{9.43}$$

These are the weights of the fundamantal representation from example 9.2.1.

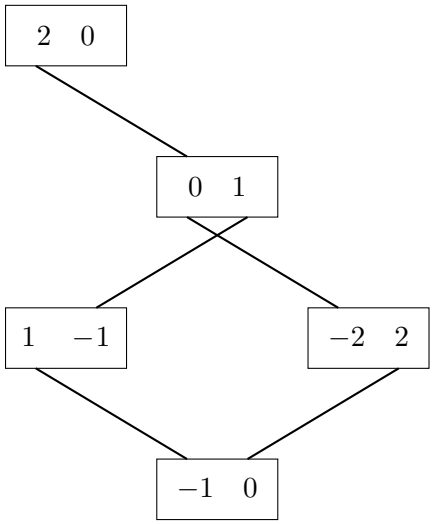
Example 9.7.2. Now for a more complicated example, consider the representation $(2,0)$, with highest weight $2\mu^2$. The highest weight state has $q_1 = 2$, $q_2 = 0$, so we can only lower it by α^1:



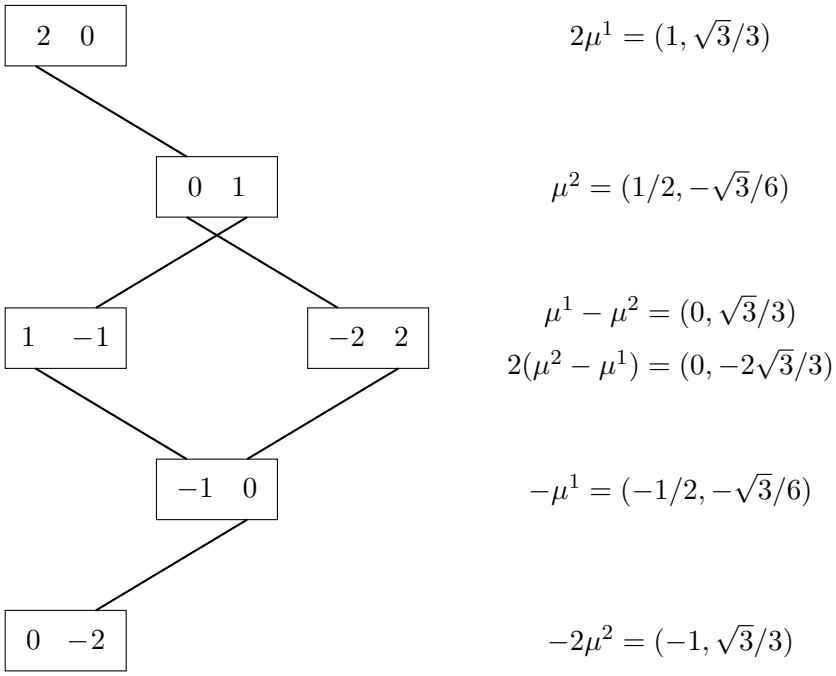
This next state has $p_1 = 1$ and $p_2 = 0$, so we read off that $q_1 = 0 + p_1 = 1$ and $q_2 = 1 + p_2 = 1$, so we can lower either by α^1 or α^2:



The values of p and q for the two new states, determined in the same way, are written below the corresponding boxes. From the values of q we see that the state on the left can be lowered by α^1 and the state on the right can be lowered by α^2. Doing either leads to the same state, with weight $\mu - 2\alpha^1 - \alpha^2 = -\mu^1$:



This state has $p_1 = p_2 = 1$, and therefore $q_1 = 0$ and $q_2 = 1$. We can therefore only lower by α^2, ending up with a state which has $q_1 = q_2 = 0$, completing this representation. The full diagram, with the values of the weights written alongside, is:



The weight $-\mu^1$ appears as though it may be doubly degenerate, since there are two routes to this state from the highest weight state $|2\mu^1\rangle$, corresponding to the states

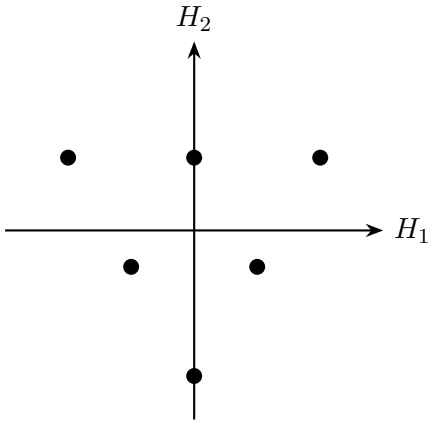
$$E_{-\alpha^2} E_{-\alpha^1} E_{-\alpha^1} |2\mu^1\rangle, \qquad E_{-\alpha^1} E_{-\alpha^2} E_{-\alpha^1} |2\mu^1\rangle. \tag{9.44}$$

However, we can use the $\mathfrak{su}(3)$ commutation relations to show that these states are proportional,

$$E_{-\alpha^2} E_{-\alpha^1} E_{-\alpha^1} |2\mu^1\rangle = 2 E_{-\alpha^1} E_{-\alpha^2} E_{-\alpha^1} |2\mu^1\rangle, \tag{9.45}$$

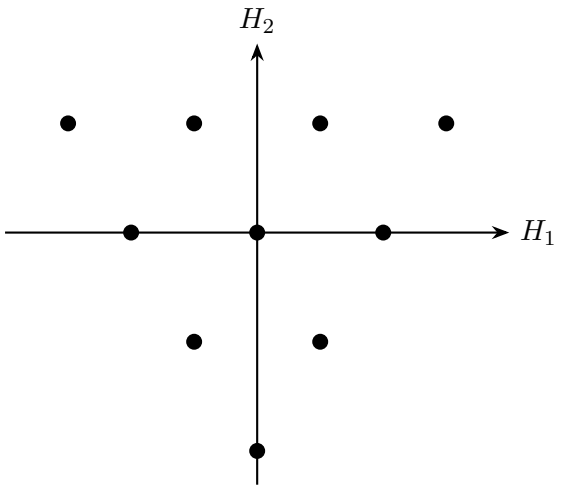
so this weight is not degenerate.

Plotted in a weight diagram, the weights of this representation form an equilateral triangle:



Since it is six-dimensional, this representation is known as the **6** of $\mathfrak{su}(3)$.

Example 9.7.3. In the next chapter we will make use of the (3,0) representation. This turns out to be ten-dimensional, so it is also called the decuplet and denoted **10**. The weight diagram looks like this:



9.8 Other compact semisimple Lie algebras

The methods that we have applied to $\mathfrak{su}(3)$ in this chapter can also be applied to other compact semisimple Lie algebras, for example $\mathfrak{su}(n)$ and $\mathfrak{so}(n)$. One defines the Cartan subalgebra as a maximal set of mutually commuting generators that we wish to diagonalise. We define the weights as the eigenvalues of the Cartan generators, the roots as the weights in the adjoint representation, and use the generators corresponding to the roots as raising and lowering operators. The only difficulty is that there are typically more than two Cartan generators, so the weights live in a larger-dimensional space.

Chapter 10

Application: Flavour symmetries in particle physics

In this chapter we'll give a brief overview of an application of the representation theory of $\mathfrak{su}(2)$ and $\mathfrak{su}(3)$ in particle physics — the spectrum of hadrons. The symmetries we discuss in this chapter relate different flavours of quark, and hence are known as flavour symmetries.

10.1 Isospin

The proton and neutron have very similar masses,

$$\begin{aligned} \text{proton:} \quad & m_p = 938.3 \text{ MeV}/c^2, \\ \text{neutron:} \quad & m_n = 939.6 \text{ MeV}/c^2. \end{aligned} \tag{10.1}$$

Since the mass of the particle is its energy at rest, this tells us that the Hamiltonian governing the strong interaction that binds nucleons together has two almost-equal eigenvalues. Let $|p\rangle$ be the state describing a proton at rest and $|n\rangle$ be the state describing a neutron at rest. We will introduce a column vector notation

$$|p\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad |n\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{10.2}$$

Suppose for now that we neglect the difference in the masses of the proton and neutron, setting equal, $m_p = m_n = m_N$ where N stands for nucleon. Then any superposition of the form

$$|\psi\rangle = a|p\rangle + b|n\rangle = \begin{pmatrix} a \\ b \end{pmatrix} \tag{10.3}$$

would also be an eigenstate of the Hamiltonian, with eigenvalue m_N.[1] We could apply any 2×2 unitary matrix to this state to obtain another such eigenstate. This suggests that the Hamiltonian describing nucleons should have an SU(2) symmetry.[2] This symmetry is

[1] In this discussion we are only considering the strong interaction, neglecting electromagnetism completely, so the fact the the proton and neutron have different electric charges is unimportant.

[2] Isospin is SU(2) and not U(2) because acting with a unitary transformation proportional to the identity just amounts to changing the phase of the wavefunction, which is not physically important.

called **isospin**. The proton and neutron would transform under the spin-half representation of of this symmetry. Conventionally, the "axes" are aligned so that the proton has T_3 eiganvalue $+1/2$. The neutron would then have eigenvalue $-1/2$,

$$T_3|p\rangle = \frac{1}{2}|p\rangle, \qquad T_3|n\rangle = -\frac{1}{2}|n\rangle. \tag{10.4}$$

Since the masses of the nucleons are not quite equal isospin is only an approximate symmetry, in a sense that will be made more precise below.

If this isospin symmetry is a real (approximate) symmetry of the strong interactions, we would expect that other strongly interacting particles should form states in representations of SU(2), also known in this context as SU(2) multiplets, and indeed they do. The lightest strongly interacting particles — the pions — have almost equal masses. Since there are three pions they should form a spin-one multiplet (since the spin-one representation of SU(2) is three dimensional). Here is a table with the pions, their masses, and their T_3 eigenvalues which must be determined by experiment:

Particle	Mass (MeV/c^2)	T_3
π^+	139.6	1
π^0	135.0	0
π^-	139.6	-1

Another example is the four Δ baryons, which all have masses of 1232 MeV/c^2, so would naturally form a spin-3/2 multiplet.

The fact that isospin is only an appoximate symmetry means that we should imagine writing the nucleon Hamiltonian as

$$H = H_0 + \delta H, \tag{10.5}$$

where H_0 has isospin symmetry and δH is a small perturbation that breaks it. Note that electromagnetic interactions break isospin symmetry (since protons and neutrons have different charges) and therefore should be included in δH. We can justify putting the electromagnetic interactions into the small δH because they are so much weaker than the strong interactions.

The idea behind splitting the Hamiltonian as in equation (10.5) is that we can perform approximate calculations by neglecting δH and making use of the isospin symmetry of H_0, without even having to know details of the Hamiltonian H_0. Here's an example:

Example 10.1.1. The deuteron d is a bound state of a proton and neutron. The tensor product of two isospin-half multiplets decomposes as,

$$\frac{1}{2} \otimes \frac{1}{2} = 1 \oplus 0, \tag{10.6}$$

so the deuteron could either have total isospin one or zero. Since the nucleons are fermions, the deuteron state should be antisymmetric under exchange of the proton and neutron. It turns out that this forces the deuteron to have isospin zero, so the deuteron state is

$$|d\rangle = \frac{1}{\sqrt{2}} \left(|p\rangle|n\rangle - |n\rangle|p\rangle \right). \tag{10.7}$$

It will be convenient to adopt notation in which we label the particles by their isospin quantum numbers, the deuteron state becoming

$$|0,0\rangle = \frac{1}{\sqrt{2}}\left(|1/2,1/2\rangle|1/2,-1/2\rangle - |1/2,-1/2\rangle|1/2,1/2\rangle\right). \tag{10.8}$$

Consider the following two processes in which two nucleons scatter to form a deuteron and a pion,

$$p + p \to d + \pi^+, \qquad p + n \to d + \pi^0. \tag{10.9}$$

Expressing the initial and final states in terms of their isospin quantum numbers, we have

$$\begin{aligned} |pp\rangle &= |1/2,1/2\rangle|1/2,1/2\rangle = |1,1\rangle, \\ |pn\rangle &= |1,2,1/2\rangle|1/2,-1/2\rangle = \frac{1}{\sqrt{2}}\left(|1,0\rangle + |0,0\rangle\right), \\ |d\pi^+\rangle &= |0,0\rangle|1,1\rangle = |1,1\rangle, \\ |d\pi^0\rangle &= |0,0\rangle|1,0\rangle = |1,0\rangle. \end{aligned} \tag{10.10}$$

The scattering amplitude to go from an initial state $|i\rangle$ to a final state $|f\rangle$ is $\langle f|S|i\rangle$, where $S = \lim_{t\to\infty} e^{2iHt}$ is an operator called the S-matrix.[3] If we pretend that isospin is an exact symmetry, the isospin generators will all commute with the Hamiltonian, and therefore also with S. This means that S is an isospin-zero tensor operator. The Wigner–Eckart theorem then tells us that the amplitudes for the two scattering processes are

$$\begin{aligned} \langle d\pi^+|S|pp\rangle &= \underbrace{C(1,1;0,0,1,1)}_{=1}\langle 1||S||1\rangle, \\ \langle d\pi^0|S|pn\rangle &= \frac{1}{\sqrt{2}}\underbrace{C(1,0;0,0,1,0)}_{=1}\langle 1||S||1\rangle + \frac{1}{\sqrt{2}}\underbrace{C(1,0;0,0,0,0)}_{=0}\langle 1||S||0\rangle. \end{aligned} \tag{10.11}$$

for some reduced matrix elements $\langle 1||S||1\rangle$ and $\langle 1||S||0\rangle$. The scattering cross section $\sigma_{f\to i}$ is proportional $|\langle f|S|i\rangle|^2$, so we find that the reduced matrix elements drop out of the ratio

$$\frac{\sigma_{pp\to d\pi^+}}{\sigma_{pn\to d\pi^0}} = \frac{|\langle d\pi^+|S|pp\rangle|^2}{|\langle d\pi^0|S|pn\rangle|^2} = 2. \tag{10.12}$$

This prediction works pretty well. The experimental value for the ratio depends on energy, but for example at centre-of-mass energy around 2150 MeV the ratio has been measured to be

$$\frac{\sigma_{pp\to d\pi^+}}{\sigma_{pn\to d\pi^0}} = 2.1 \pm 0.4. \tag{10.13}$$

10.2 Quarks

Isospin symmetry dates back to work by Heisenberg and Wigner in the thirties. In the sixties it was realised that the hadrons, of which nucleons and pions are examples, are

[3] This formula for the S-matrix is not quite right as their are some subtleties to be taken account of, but it captures the important point that the generators of the symmetry will commute with the S-matrix.

composed of smaller particles — the up and down quarks. The masses of these quarks are difficult to measure precisely, but they are approximately

$$m_u \approx 2.2 \text{ MeV}/c^2, \qquad m_d \approx 4.7 \text{ MeV}/c^2 \tag{10.14}$$

We can understand isospin symmetry as arising from the fact that these quarks are so much lighter than any of the hadrons, so we can approximately neglect the difference between m_u and m_d when studying hadronic physics.

The up and down quark form an isospin-two doublet, with the up-quark mass having T_3 eigenvalue $+1/2$ and the down quark having T_3 eigenvalue $-1/2$. Baryons are bound states of three quarks. Since the tensor product of three isospin-half representations is

$$\frac{1}{2} \otimes \frac{1}{2} \otimes \frac{1}{2} = \frac{3}{2} \oplus \frac{1}{2} \oplus \frac{1}{2}, \tag{10.15}$$

baryons composed of up and down quarks can either form doublets (total isospin = 1/2) or quadruplets (total isospin = 3/2). The nucleons are an isospin doublet while the Δ baryons are an isospin quadruplet.

Mesons are bound states of a quark and an antiquark. Since

$$\frac{1}{2} \oplus \frac{1}{2} = 1 \oplus 0, \tag{10.16}$$

mesons composed of up and down quarks can either form isospin singlets or triplets. The pions form a triplet, while there is an η meson that forms a singlet.

10.3 The eightfold way

After the up and the down, the next lightest quark is the strange, with mass

$$m_s \approx 94 \text{ MeV}/c^2. \tag{10.17}$$

This is still lighter than the mass of any hadrons, so we can try to extend the flavour symmetry group to SU(3), consisting of unitary transformations that mix up all three light quarks. Isospin would then be an SU(2) subgroup of the flavour SU(3) symmetry. Since the strange quarks are quite a bit heavier than the up and the down, and not much lighter than pions, flavour SU(3) will not be such a good approximation as isospin.

Historically, a flavour SU(3) symmetry was noticed before the discovery of quarks. As more and more hadrons were discovered in the middle of the previous century, it was observed that they could be assigned a quantum number S, *strangeness*, that was approximately conserved in scattering and decays. These days we think of strangeness as

$$S = (\text{number of antistrange quarks}) - (\text{number of strange quarks}). \tag{10.18}$$

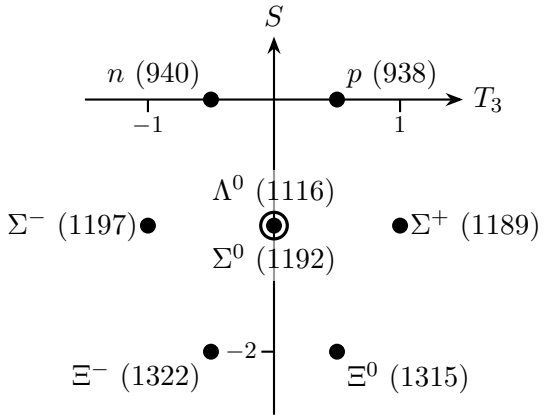
For example, consider the following two baryons, with similar masses but drastically different lifetimes:

Particle	Quark content	Strangeness	Mass (MeV/c^2)	Mean lifetime (s)
Λ^0	uds	-1	1116	3×10^{-10}
Δ^0	udd	0	1232	6×10^{-24}

Both particles typically decay to a nucleon plus a pion, with zero total strangeness. Thus the decay of the Λ^0 violates strangeness conservation, while the decay of the Δ^0 preserves it. The Λ^0 is the lightest baryon with non-zero strangeness, so has no possible decay channel that conserves strangeness. The strangeness-violating decay of the Λ^0 occurs many orders of magnitude slower than the decay of the Δ^0. This is what we mean by strangeness being approximately conserved.

We learned in chapter 9 that SU(3) has two SU(2) subgroups. We have already noted that isospin should be one of the subgroups of flavour SU(3). The approximate conservation of strangeness suggests that S should be related to the quantum numbers of the other SU(2) subgroup. Let's have a look at what happens if we plot the strangeness S versus the third component of isospin T_3 for various hadrons.

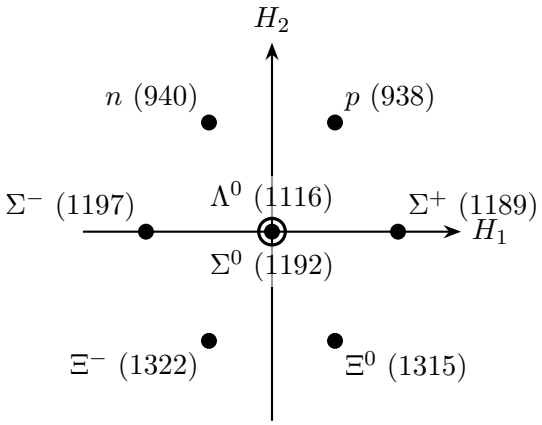
There are eight light spin-half baryons: the proton and neutron, the Λ^0 that we discussed above, three Σ baryons that each contain one strange quark, and two Ξ baryons that each contain two strange quarks. Plotting their strangeness against T_3:



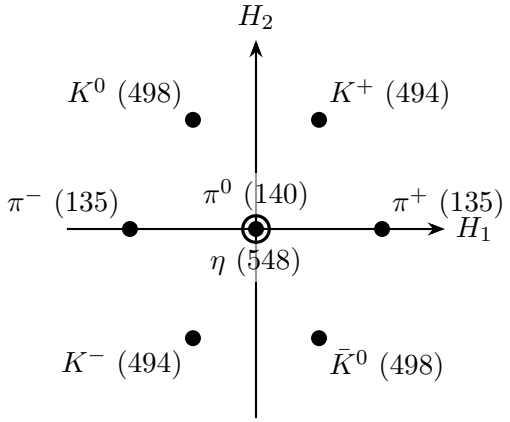
The numbers in brackets are the masses of the baryons in units of MeV/c^2. We see that the masses are indeed relatively similar. For comparison, the next-lightest spin-half baryon is the Λ_c^+, with a mass of 2286 MeV/c^2. This diagram looks a lot like the roots of $\mathfrak{su}(3)$, but the hexagon is irregular and it is centered on $S = -1$. We can fix both of these problems by defining

$$H_1 = T_3, \qquad H_2 = \frac{\sqrt{3}}{2}(S + B), \tag{10.19}$$

where B is the baryon number; $B = 1$ for baryons and $B = 0$ for mesons. The values of H_1 and H_2 for these baryons are then exactly equal to the roots of $\mathfrak{su}(3)$ and the diagram becomes:



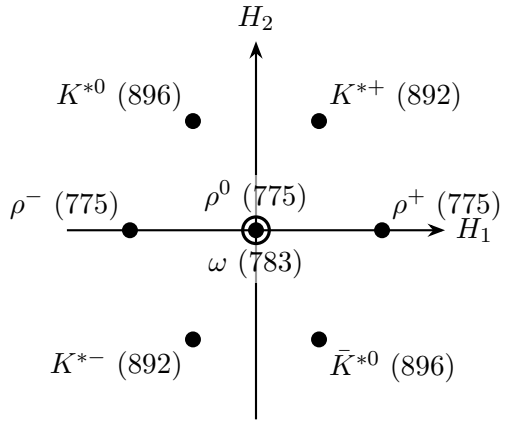
Exactly the same pattern appears for the eight lightest spin-zero mesons:



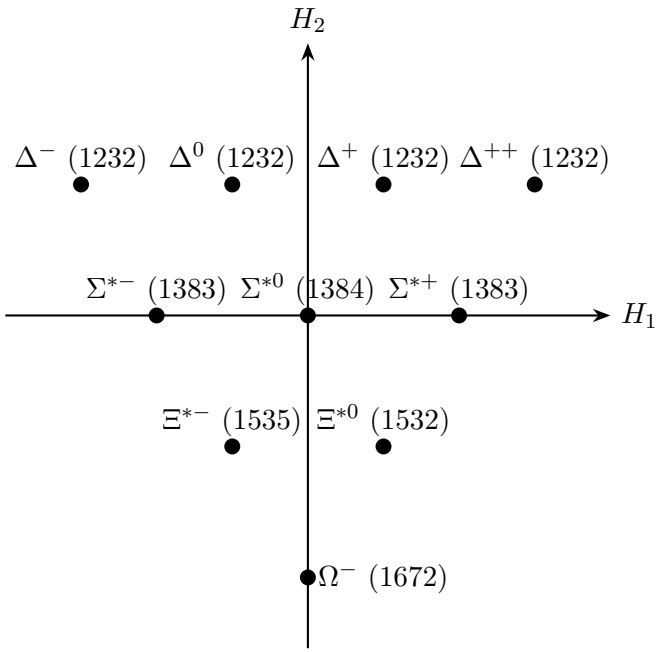
The mass splittings in this meson octet are quite big, so $\mathfrak{su}(3)$ does not work so well for this octet. This isn't too surprising, since the pions are not much more massive than the strange quarks.

The organisation of hadrons into $\mathfrak{su}(3)$ multiplets was proposed by Murray Gell-Mann and Yuval Ne'eman in 1961, three years before the introduction of the quark model. Gell-Mann named this scheme *the eightfold way*.

Continuing the classification of hadrons there are spin-one mesons, called vector mesons, the lightest of which form another octet:



Not everything is an octet though. There are ten light spin-$\frac{3}{2}$ baryons, forming an $\mathfrak{su}(3)$ decuplet:



When the eightfold way was first proposed, the Ω^- had not been discovered. Gell-Mann and Ne'eman were able to predict the existence of the Ω^- based on the fact that without it the spin-$\frac{3}{2}$ baryons do not form a complete SU(3) multiplet. To fit in the right place in the multiplet, the Ω^- had to have zero isospin and strangeness $S = -3$. Indeed, just such a particle was observed a few years later.

Symmetry was also used to predict the mass of the Ω^-. Suppose that the Hamiltonian governing the strong interactions takes the form

$$H = H_0 + \delta H, \tag{10.20}$$

where H_0 has an SU(3) flavour symmetry and δH includes all the effects that break the SU(3) symmetry. Suppose first that we neglect δH completely, so that the SU(3) symmetry is exact. Then all the hadrons in a multiplet will have the same mass m_0. Suppose that $|\psi\rangle$ is the eigenstate of H corresponding to some hadron with mass m_0 at rest, so that $m_0 = \langle\psi|H_0|\psi\rangle$. If we now allow for small but non-zero δH, first-order perturbation theory tells us that the mass of the baryon is now

$$m \approx \langle\psi|(H_0 + \delta H)|\psi\rangle = m_0 + \langle\psi|\delta H|\psi\rangle. \tag{10.21}$$

If we assume that δH is an $\mathfrak{su}(3)$ tensor operator, then there is an $\mathfrak{su}(3)$ version of the Wigner–Eckart theorem that relates the expectation values $\langle\psi|\delta H|\psi\rangle$ for different members of the same hadron multiplet.

Unfortunately we don't have time to go into the details of the Wigner–Eckart theorem for $\mathfrak{su}(3)$, so we will just state some results. Gell-Mann and Susumu Okubo proved that if you assume that δH is the T_8 component of a tensor operator in the adjoint representation, then the masses of the baryons in the octet containing the nucleons are related by

$$2(m_N + m_\Xi) = 3m_\Lambda + m_\Sigma. \tag{10.22}$$

This is the **Gell-Mann–Okubo mass formula**. For instance, if you plug in the approximate masses of the nucleons, Ξs, and Σs,

$$m_N \approx 940 \text{ MeV}/c^2, \qquad m_\Sigma \approx 1190 \text{ MeV}/c^2, \qquad m_\Xi \approx 1320 \text{ MeV}/c^2, \tag{10.23}$$

then the formula gives the following mass of the Λ,

$$m_\Lambda \approx 1110 \text{ MeV}/c^2. \tag{10.24}$$

This is very close to the experimentally measured value of $m_\Lambda = 1116 \text{ MeV}/c^2$, suggesting that the assumptions on δH that led to the Gell-Mann–Okubo mass formula are good.

Applying the $\mathfrak{su}(3)$ Wigner–Eckart theorem to the baryon decuplet leads to the following relations among the masses,

$$m_{\Sigma^*} - m_\Delta = m_{\Xi^*} - m_{\Sigma^*} = m_\Omega - m_{\Xi^*}. \tag{10.25}$$

In other words the masses of the different rows of the decuplet should be equally spaced. Plugging in the experimental results for all but the Ω^-, we have

$$m_{\Sigma^*} - m_\Delta \approx 151 \text{ MeV}/c^2, \qquad m_{\Xi^*} - m_{\Sigma^*} \approx 152 \text{ MeV}/c^2. \tag{10.26}$$

So indeed the masses first three rows are very close to equally spaced. We would then expect

$$m_\Omega \approx m_{\Xi^*} + 150 \text{ MeV}/c^2 \approx 1685 \text{ MeV}/c^2. \tag{10.27}$$

This is very close to the experimental result $m_\Omega = 1672 \text{ MeV}/c^2$. In this way, the mass of the Ω was predicted based on SU(3) symmetry before the particle was ever discovered!

The quark model allows us to understand why it is that the mesons and baryons fall in octets and decuplets of $\mathfrak{su}(3)$, rather than other representations. The up, down, and strange quarks transform in the fundamental representation $\mathbf{3}$ of $\mathfrak{su}(3)$. The antiquarks transform in the antifundamental representation $\bar{\mathbf{3}}$. The mesons, which are bound states of one quark and one antiquark, should then be states in the tensor product representation $\mathbf{3} \otimes \bar{\mathbf{3}}$. It turns out that this representation decomposes in irreducible representations as

$$\mathbf{3} \otimes \bar{\mathbf{3}} = \mathbf{8} \oplus \mathbf{1}, \tag{10.28}$$

where $\mathbf{8}$ is the octet (adjoint) representation and $\mathbf{1}$ is the trivial representation. This is why mesons come in octets and singlets. The baryons are bound states of three quarks, and should therefore transform in the representation

$$\mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3} = \mathbf{10} \oplus \mathbf{8} \oplus \mathbf{8} \oplus \mathbf{1}. \tag{10.29}$$

This is why we find both octets and decuplets of baryons. It looks like it should be possible to also have a baryon singlet, however this turns out to be forbidden by the Pauli exclusion principle.

Bibliography

[1] Wikipedia contributors. *Examples of groups — Wikipedia, The Free Encyclopedia*. [Online; accessed 18-June-2024]. 2023. URL: https://en.wikipedia.org/w/index.php?title=Examples_of_groups&oldid=1175831459.

[2] Brian C. Hall. *Lie Groups, Lie Algebras, and Representations*. Springer Cham, 2015. ISBN: 978-3-319-13467-3.

[3] Howard Georgi. *Lie Algebras in Particle Physics*. ABP, 1999.