MA PH 464: Group theory in physics

Reinier Kramer

April 5, 2023

Contents

0.1 — Introduction

These are the lecture notes for MA PH 464: Group theory in physics, as given at the University of Alberta during the Winter 2023 term. They may be updated during the course.

Group theory is a beautiful subject which on the one hand serves as a natural introduction into abstract algebra, and on the other hand has many important applications, especially in physics. This course will attempt to introduce both of these aspects. As such, there will be a formal aspect – developing the theory from the ground up, via definitions, theorems, and the like – and a more informal part – focusing on examples and applications.

One particular reason symmetries are important for physics, is Noether's theorem, which tells us that continuous symmetries of a system give rise to conserved quantities. At the end of the course, we will see some particular instances of this.

0.1.1 — Acknowledgments

None of this material is new. These lecture notes, an updated version of the ones from Winter 2022, are largely inspired by 2020 lecture noted for the same course, written by Vincent Bouchard, and located at https://sites.ualberta.ca/~vbouchar/MAPH464/notes.html, and by (Dutch) lecture notes by Hendrik Lenstra and Frans Oort, updated by Ben Moonen, and with an appendix by Raf Bocklandt, which can be found at https://staff.fnwi.uva.nl/r.r.j.bocklandt/algebra1.pdf. Other sources include the books *Introduction to Lie Algebras and Representation Theory*, by James E. Humphreys (Springer 1997) and *Linear Representations of Finite Groups*, by Jean-Pierre Serre (Springer 1977).

I would also like to thank Gurkanwal Sedha for sending me numerous typographical corrections on the text.

Chapter 1 — Abstract groups

Group theory is a branch of what is called abstract algebra. This broader subject deals with objects and relations, regardless of setting, and tries to find common structures between them. In a sense, it is the formalisation of the term 'symmetry'.

If the previous sentence sounds too abstract for you, do not worry. Groups may be part of this broad setting, but they are relatively accessible, and in particular, you already know quite a few examples! This chapter will study in more (or different) detail some groups that should already be familiar to you, and will also introduce new examples. Furthermore, it will give you tools for dealing with them.

1.1 — Groups: definition and examples

When talking about group theory (or any mathematical theory really), it is of the utmost importance to specify exactly what you talk about. Hence, we should start with the main definition of the course.

Definition 1.1.1. A *group* consists of a set G, together with a binary operation, often called *multiplication* or *product*, $m\colon G \times G \to G\colon (g,h) \mapsto g \cdot h$, a *unit* element $e \in G$, and a unary operation called *inverse* $i\colon G \to G\colon g \mapsto g^{-1}$. These are required to satisfy the following axioms:

G1 *Associativity:* for any three elements $g, h, k \in G$, we have $(g \cdot h) \cdot k = g \cdot (h \cdot k)$;

G2 *Unit:* for any element $g \in G$, we have $e \cdot g = g = g \cdot e$;

G3 *Inverse:* for any element $g \in G$, we have $g \cdot g^{-1} = e = g^{-1} \cdot g$.

Often we denote the group with the same symbol as the underlying set. We also often drop the dot from the notation of multiplication, i.e. we will write ab in stead of $a \cdot b$.

Definition 1.1.2. A group G is a *abelian* if the following extra axiom holds:

G4 *Commutativity:* for any $g, h \in G$, we have $g \cdot h = h \cdot g$.

For an abelian group, we often (but not always) use additive notation, i.e. we write $+$, 0, and $-$ in stead of $\cdot$, e, and $()^{-1}$ for the operations. The language will then also reflect this, e.g. we would use 'term' in stead of 'factor'.

Example 1.1.3. The trivial group, $G = \{e\}$, is an abelian group under the only possible multiplication. The empty set is *not* a group: it has no unit. ◇

Example 1.1.4. The integers, $\mathbb{Z}$, are an abelian group under addition. The unit is 0, and the inverse of an integer n is $-n$.

It is *not* a group under multiplication: while G1 and G2 hold, with unit 1, not every element has an inverse. E.g., there is no integer n such that $n \cdot 0 = 1$, so 0 has no inverse. ◇

Example 1.1.5. In a similar way, the rational numbers $\mathbb{Q}$, the real numbers $\mathbb{R}$, and the complex numbers $\mathbb{C}$, are all abelian groups under addition. In fact, any field, even any ring is, if you happen to have seen these notions. ◇

EXAMPLE 1.1.6. The natural numbers, $\mathbb{N} = \{0, 1, 2, \dots\}$, are not a group under addition. Indeed, while G1 and G2 are satisfied (with 0 being the unit), there is no inverse of, say, 1. ◇

EXAMPLE 1.1.7. The *cyclic group* $\mathbb{Z}/n\mathbb{Z}$ is the set of integers modulo n. Its elements are $\overline{m} = m \pmod{n} = \{m + kn \mid k \in \mathbb{Z}\}$, so in particular $\overline{m} = \overline{m+n}$, and we may choose representatives $\overline{0}, \dots, \overline{n-1}$. Then $\overline{m} + \overline{l} = \overline{m+l}$ gives a well-defined group structure on this set, with unit $\overline{0}$, and the inverse of $\overline{m}$ is $\overline{-m}$. These groups are all abelian.

These groups are also denoted by C_n, $\mathbb{Z}/n$, $\mathbb{Z}/(n)$, or $\mathbb{Z}_n$. The last notation is ambiguous, as it is also used for other groups, so we will refrain from using it.

For the case $n = 2$, $\mathbb{Z}/2\mathbb{Z}$ has two elements: $\bar{0} = \{\dots, -2, 0, 2, 4, \dots\}$ is the set of even numbers, and $\bar{1} = \{\dots, -3, -1, 1, 3, 5, \dots\}$ the set of odd numbers. We know that the sum of two even numbers is even, the sum of two odd numbers is even, and the sum of an even and an odd number is odd. This is the addition of $\mathbb{Z}/2\mathbb{Z}$. ◇

EXAMPLE 1.1.8. Any vector space (over any field) is an abelian group with respect to the addition. This is part of the definition of a vector space. ◇

EXAMPLE 1.1.9. Given a vector space V, the set of invertible linear transformations $f\colon V \to V$ is a group under composition of transformations, called the *general linear group*, and denoted $\mathrm{GL}(V)$. The unit is the identity map $\mathrm{Id}_V : v \mapsto v$.

If $V = \mathbb{R}^n$, this group may be identified with the set of $n \times n$ real matrices with non-vanishing determinant. In this case, we write $\mathrm{GL}(n, \mathbb{R})$. Similarly, if $V = \mathbb{C}^n$, we write $\mathrm{GL}(n, \mathbb{C})$. The unit is the identity matrix. For $n = 1$, we see $\mathrm{GL}(1, \mathbb{R}) = \mathbb{R}^\times = \mathbb{R} \setminus \{0\}$ (with multiplication), and similarly for $\mathbb{C}$. If $n \geq 2$, this group is not abelian. ◇

EXAMPLE 1.1.10. The unit circle in $\mathbb{C}$, i.e. the set $\{z \in \mathbb{C} \mid \|z\| = 1\}$, is also a group under multiplication. The unit is 1, and multiplication and inverse are the same as for the complex numbers. Any element of this group can be written $e^{i\vartheta}$ for some $\vartheta \in \mathbb{R}$, and $e^{i\vartheta}e^{i\varphi} = e^{i(\vartheta+\varphi)}$. The inverse of $e^{i\vartheta}$ is $e^{-i\vartheta}$. ◇

EXAMPLE 1.1.11. There exists a particular four-dimensional $\mathbb{R}$-vector space which also has a multiplication (just like the two-dimensional $\mathbb{R}$-vector space $\mathbb{C}$). It is called the space of *quaternions* and denoted $\mathbb{H}$ (the letter is chosen because they were discovered by Hamilton, also known for the Hamiltonian in physics).

The quaternions have a basis $\{1, i, j, k\}$, with 1 as a unit, and further multiplication rules

$$\begin{aligned} i^2 = j^2 = k^2 &= -1\,, \\ ij = k \quad ji &= -k\,, \\ jk = i \quad kj &= -i \\ ki = j \quad ik &= -j\,. \end{aligned} \tag{1.1}$$

You can think of the vectors i, j, k as the basis vectors of Euclidean three-space, and 1 as a scalar. Then the above multiplication encodes the scalar product, the cross product, and the dot product!

For general elements, the multiplication is

$$\begin{aligned} (a + bi + cj + dk)(a' + b'i + c'j + d'k) = &(aa' - bb' - cc' - dd') \\ &+ (ab' + ba' + cd' - dc')i \\ &+ (ac' + ca' + db' - bd')j \\ &+ (ad' + da' + bc' - cb')k\,. \end{aligned}$$

This multiplication is associative. Try to check this yourself! It helps to write $a + bi + cj + dk = \alpha + \beta j$, where $\alpha = a + bi$ and $\beta = c + di$ are complex. Then

$$(\alpha + \beta j)(\gamma + \delta j) = (\alpha\gamma - \beta\bar{\delta}) + (\alpha\delta + \beta\bar{\gamma})j\,.$$

Moreover, any element $a + bi + cj + dk \neq 0$ has an inverse (again, check this!)

$$\frac{a}{n} - \frac{b}{n}i - \frac{c}{n}j - \frac{d}{n}k\,, \qquad \text{where } n = a^2 + b^2 + c^2 + d^2 \neq 0\,.$$

With this multiplication, $\mathbb{H}^\times = \mathbb{H} \setminus \{0\}$ is a non-commutative group. ⋄

The way we defined groups, we assume there is such a thing as a unit and an inverse. But could there be another one of either? The next proposition shows this is not the case. Furthermore, it gives some rules on using the inverse.

Proposition 1.1.12. *The unit of a group is unique: there is no other element satisfying G2.*
The inverse of an element is unique: there is no other element satisfying G3.
For any elements $g, h \in G$, we have that $(g^{-1})^{-1} = g$ and $(gh)^{-1} = h^{-1}g^{-1}$.

Proof. Suppose e' is another unit. Then $e = e \cdot e' = e'$.
Let $g \in G$, and suppose it has two inverses, g' and g^*. Then

$$g' = g' \cdot e = g' \cdot (g \cdot g^*) = (g' \cdot g) \cdot g^* = e \cdot g^* = g^*\,.$$

The element g satisfies the definition of being an inverse of g^{-1}, because $g \cdot g^{-1} = e = g^{-1} \cdot g$. By uniqueness of the inverse for g^{-1}, it must be g. For the final point,

$$(gh) \cdot (h^{-1}g^{-1}) = g \cdot (h \cdot (h^{-1}g^{-1})) = g \cdot ((hh^{-1}) \cdot g^{-1}) = g \cdot (e \cdot g^{-1}) = gg^{-1} = e\,,$$

and similarly for the other relation. So by uniqueness of the inverse for gh, it must be $h^{-1}g^{-1}$. □

As you may have noticed in the proof above, writing out identities becomes lengthy quite quickly. However, using associativity, one may prove by induction (try to do this carefully yourself!) that any way of inserting parentheses inside the expression $a_1 a_2 \cdots a_n$, without shuffling the factors, will give the same result. Hence, we will not write these parentheses from now on.

Let us also introduce some more shorthand notation: for $g \in G$ and $n \in \mathbb{Z}$, we mean by g^n the element

$$g^n = \begin{cases} e & n = 0\,; \\ \underbrace{g \cdot g \cdots g}_{n \text{ times}} & n > 0\,; \\ \underbrace{g^{-1} \cdots g^{-1}}_{-n \text{ times}} & n < 0\,. \end{cases} \tag{1.2}$$

For an additively written group, this becomes

$$n \cdot g = \begin{cases} 0 & n = 0\,; \\ \underbrace{g + g + \cdots + g}_{n \text{ times}} & n > 0\,; \\ \underbrace{(-g) + \cdots + (-g)}_{-n \text{ times}} & n < 0\,. \end{cases} \tag{1.3}$$

Lemma 1.1.13. *For any group G, any $g \in G$, and any $n, m \in \mathbb{Z}$, we have $g^n \cdot g^m = g^{n+m}$.*

Proof. If $n = 0$, then

$$g^n \cdot g^m = e \cdot g^m = g^m = g^{0+m}\,.$$

If $n = 1$, then

$$g^n \cdot g^m = \begin{cases} g \cdot e = g = g^{1+m} & m = 0\,; \\ \underbrace{g \cdot g \cdots g}_{1+m \text{ times}} = g^{1+m} & m > 0\,; \\ g \cdot \underbrace{g^{-1} \cdots g^{-1}}_{-m \text{ times}} = \underbrace{g^{-1} \cdots g^{-1}}_{-m-1 \text{ times}} = g^{1+m} & m < 0\,. \end{cases}$$

By induction, the assertion hold for all $n > 0$. The case $n < 0$ is similar. □

Looking at the examples we gave, you may notice that most of them have one of two flavours: they are either *discrete*, meaning that all elements are somehow isolated, or *continuous*. Some discrete groups are $\mathbb{Z}$ or $\mathbb{Z}/n\mathbb{Z}$, while $\mathrm{GL}(n, \mathbb{R})$ is continuous. This is not quite a mathematically formal distinction, but it is useful nonetheless. In the first two parts of this course, we will mostly focus on discrete groups, while the third part deals with certain continuous groups called *Lie groups*. Note that e.g. $\mathbb{Q}$ is neither discrete nor continuous.

Definition 1.1.14. A group is *finite* if its underlying set is. We call the number of elements of a finite group G its *order*, and denote it $|G|$ or $\#G$. If G is not finite, we may write $|G| = \infty$.

The *order* of an element $g \in G$ is the smallest positive integer n such that $g^n = e$. If no such n exists, the order is infinity.

Lemma 1.1.15. *In a finite group, all elements have finite order, bounded by the order of the group.*

Proof. Let G be a finite group and $g \in G$. Because G is finite, not all g^n for $n > 0$ can be distinct. So let us say $g^n = g^m$, for some $n < m$. Then $g^{m-n} = g^m g^{-n} = g^n g^{-n} = g^{n-n} = g^0 = e$. Hence, the order of g is at most $m - n$.

If we take m to be the smallest positive integer such that g^m equals some g^n for $0 \leq n < m$, then the elements $g^{n+1}, g^{n+2}, \ldots, g^m$, must all be distinct, and therefore their number, $m - n$ is bounded by the number of distinct elements in G, which is its order. □

Example 1.1.16. The order of $\mathbb{Z}/n\mathbb{Z}$ is n. The order of an element $\overline{m} \in \mathbb{Z}/n\mathbb{Z}$ need not be n, it may be smaller. In fact, it is $n/\gcd(n, m)$.

For example, in $\mathbb{Z}/4\mathbb{Z}$, the order of $\bar{2}$ is 2, because $2 \cdot \bar{2} = \bar{4} = \bar{0}$ (recall that this group uses additive notation, so $2 \cdot \bar{2}$ means $\bar{2} + \bar{2}$). ⋄

Example 1.1.17. In the group $\mathrm{GL}(n, \mathbb{R})$, the element

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

has order 2, as it squares to the identity.

On the other hand,

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^n = \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}$$

will never be the identity for $n > 0$, so

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

has infinite order. ⋄

EXAMPLE 1.1.18. The *symmetric group* S_n is the set of permutations (i.e. bijections to itself) of a set of n elements, usually chosen as $\{1, \ldots, n\}$. The multiplication is composition of maps (read from right to left as usual). The first two are abelian, but for $n \geq 3$, S_n is not abelian.

The order of S_n is $n!$: to specify an element $\sigma \in S_n$, we have n choices for $\sigma(1)$, then $n-1$ choices for $\sigma(2)$, and so on, until $\sigma(n)$ is the unique element that has not been picked yet.

These groups are important, so let us give some more details.

One way of writing an element $\sigma \in S_n$ is as an $n \times 2$ matrix, where the top row records all elements $1, \ldots n$, and the bottom row their images, so

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

In this notation, one can permute columns and keep the same permutation, which is useful for calculating the composition: $\sigma\tau$ is found by stacking the matrix of τ on top of that of σ such that the two middle rows agree. Then the composition is given by the top and bottom row.

A specific kind of permutations are called *cycles*. A cycle $\sigma \in S_n$ is a permutation such that there exists some elements $1 \leq a_1, \ldots a_k \leq n$ such that $\sigma(a_i) = a_{i+1}$, $\sigma(a_k) = a_1$, and $\sigma(x) = x$ if x is not among the a_i. We write such a cycle as $\sigma = (a_1\, a_2 \cdots a_k)$.

For example, the cycle $(2\,5\,4) \in S_5$ is given by

$$(2\,5\,4) = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 5 & 3 & 2 & 4 \end{pmatrix} .$$

Cycles of length 2 are called *transpositions*.

Any permutation σ is a product of disjoint cycles: starting from 1, we obtain a cycle $(1\, \sigma(1)\, \sigma(\sigma(1)) \cdots)$, which eventually returns to 1. Then, picking any x not in this cycle, we record its iterated images under σ in another cycle. Continuing until there are no elements left, we have a number of cycles that are disjoint, and σ is the product of these.

This way of writing is unique up to reordering of the cycles, cyclic reordering withing the cycles, and cycles of order one. For example, $(1\,5\,4)(2\,6) \in S_6$ is equal to $(6\,2)(3)(5\,4\,1)$. In particular, the unit can be represented by any number of order one cycles, although we may also write $e = ()$.

Now we have another way to write elements of S_n, how do we multiply them? Let us try to multiply two cycles in S_4, namely $(2\,4\,3)$ and $(2\,3)$. So we want to calculate $(2\,4\,3)(3\,4)$ and write it as a product of disjoint cycles. Note that we are talking about maps acting on the set $\{1, 2, 3, 4\}$, so we first act with the right-most element.

Let us start writing down the answer (even if we do not know it yet). We know that 1 will be in some cycle, so our first start is

$$(1$$

I have not closed the parenthesis, as we do not know the rest of the cycle yet, but we will find that now: First we act by $(3\,4)$. This leaves 1 alone. Similarly for $(2\,4\,3)$. So our first cycle is

$$(1)$$

Now we start a new cycle, by 2:

$$(1)(2$$

Again, no closing of the parenthesis. We find that 2 gets mapped to itself by $(3\,4)$ and then to 4:

$$(1)(2\,4$$

Now we cannot yet close the parenthesis; we need to continue with 4. This gets mapped to 3 and then to 2, so it closes the cycle:

$$(1)(2\,4)$$

There is no space left for 3 to go, and indeed, it gets mapped to 4 and back, so in the end we find that

$$(2\,4\,3)(3\,4) = (1)(2\,4)(3) = (2\,4)\,. \qquad \diamond$$

For any group element, we can consider what happens if we multiply by that element. The theorem below tells us that this is quite a well-behaved operation – in fact we will use this a lot later.

THEOREM 1.1.19 (REARRANGEMENT THEOREM). *Let G be a group, and $g, h \in G$. Then there exists exactly one solution $x \in G$ to $gx = h$, namely $x = g^{-1}h$. Similarly, there is one solution y to $yg = h$, namely $y = hg^{-1}$.*

Hence, for any $g \in G$, the maps $\lambda_g \colon G \to G \colon h \mapsto gh$ and $\rho_g \colon G \to G \colon h \mapsto hg$ are bijections. We call these maps left multiplication by g *and* right multiplication by g*, respectively.*

Proof. We first see that $x = g^{-1}h$ is a solution, as

$$g(g^{-1}h) = e \cdot h = h\,.$$

On the other hand, from $gx = h$, we find that

$$x = e \cdot x = g^{-1}gx = g^{-1}h\,.$$

The proof for y is similar. □

1.2 — Subgroups, cosets, direct products

1.2.1 — Subgroups

Often, groups have natural subsets which are groups themselves.

DEFINITION 1.2.1. Let G be a group. A subset $H \subseteq G$ is called a *subgroup* if it contains the unit of G and for all $g, h \in H$, both gh and g^{-1} lie in H.

This definition means exactly that H is a group itself with the same operations.

DEFINITION 1.2.2. Every group has two *trivial subgroups*, namely $\{e\}$ and the group itself. Any other subgroup is called a *proper subgroup*.

EXAMPLE 1.2.3. The chain $\mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C} \subset \mathbb{H}$ is a chain of (additive) subgroups.

Similarly, $\mathbb{Q}^\times = \mathbb{Q} \setminus \{0\} \subset \mathbb{R}^\times \subset \mathbb{C}^\times \subset \mathbb{H}^\times$ is a chain of multiplicative subgroups. ⋄

EXAMPLE 1.2.4. The inclusion $\mathbb{R}^\times \subset \mathbb{R}$ is not a subgroup, even though both sets are groups. However, they are groups under different group structures: multiplication for $\mathbb{R}^\times$ and addition for $\mathbb{R}$. ⋄

EXAMPLE 1.2.5. In $\mathrm{GL}(n, \mathbb{R})$ there are a number of well-known subgroups:

The *special linear group* $\mathrm{SL}(n, \mathbb{R})$ is the subgroup of matrices M such that $\det M = 1$.

The *orthogonal group* $\mathrm{O}(n)$ is the subgroup of matrices M such that $M^T M = I$.

We may intersect these to obtain the *special orthogonal group* $\mathrm{SO}(n)$ of matrices M such that $\det M = 1$ and $M^T M = I$.

If $n = 2m$ is even, we may define a non-singular skew-symmetric matrix, e.g.

$$\Omega = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}\,, \tag{1.4}$$

and consider the *symplectic group* $\mathrm{Sp}(n, \mathbb{R})$ to be the subgroup of matrices such that $M^T \Omega M = \Omega$. The symplectic group is a subgroup of the special linear group, as all symplectic matrices have determinant 1.⋄

Example 1.2.6. Similarly, in $\mathrm{GL}(n,\mathbb{C})$ there are a number of well-known subgroups:

The *special linear group* $\mathrm{SL}(n,\mathbb{C})$ is again the subgroup of matrices M such that $\det M = 1$.

The *unitary group* $\mathrm{U}(n)$ is the subgroup of matrices M such that $M^\dagger M = I$. Here by $M^\dagger$, we mean $(M^T)^*$, i.e. the complex conjugate transpose, also called *adjoint*.

Specifically, if $n = 1$, $\mathrm{U}(1)$ is the subgroup of 1×1 matrices, i.e. complex numbers, such that $z^\dagger = z^{-1}$. These are exactly the complex numbers of unit norm, cf. example 1.1.10.

Again, if $n = 2m$ is even, the *symplectic group* $\mathrm{Sp}(n,\mathbb{C})$ is the subgroup of matrices such that $M^T\Omega M = \Omega$. We may intersect these to obtain the *special unitary group* $\mathrm{SU}(n)$ of matrices M such that $\det M = 1$ and $M^\dagger M = I$ or the *compact symplectic group* $\mathrm{Sp}(n) = \mathrm{Sp}(n,\mathbb{C}) \cap \mathrm{U}(n)$. ◇

The previous two examples will be studied more in chapter 3, as they are examples of Lie groups.

Example 1.2.7. Consider a regular n-gon in the complex plane, with vertices $\{e^{2\pi ik/n} \mid k = 0,\dots,n-1\}$. E.g., for $n = 3$ this is an equilateral triangle, for $n = 4$ it is a square, &c. If we consider $\mathbb{C}$ as a two-dimensional vector space over $\mathbb{R}$, we can consider the subgroup of $\mathrm{GL}(2,\mathbb{R})$ that preserves this n-gon. This group, D_{2n}, is called the *dihedral group* and has $2n$ elements, of two kinds. There are pure rotations, which can be considered as multiplication by $e^{2\pi ik/n}$ for some k, and reflections about axes of symmetry, of which one is the conjugation $\sigma\colon z \mapsto \bar{z}$. If we write ρ for the rotation by $2\pi i/n$, then

$$D_{2n} = \{\rho^k \mid 0 \le k < n\} \cup \{\rho^k\sigma \mid 0 \le k < n\}\,, \tag{1.5}$$

and multiplication is given by the rules

$$\begin{aligned} \rho^n &= e \\ \sigma^2 &= e \\ \sigma\rho^k &= \rho^{n-k}\sigma\,. \end{aligned} \tag{1.6}$$

For $n > 2$, this group is not abelian. ◇

Example 1.2.8. The set $Q = \{1,-1,i,-i,j,-j,k,-k\} \subset \mathbb{H}^\times$ is a subgroup, as follows from equation (1.1). We will call this the *quaternion group*. It is still not abelian. ◇

There are several ways to construct subgroups of a given group.

Definition 1.2.9. Let G be a group and $S \subseteq G$ be a subset. The *subgroup of G generated by S*, written $\langle S\rangle$, is the smallest subgroup of G that contains S. It contains all products of elements of S and their inverses.

If $S = \{s\}$, then $\langle s\rangle = \langle S\rangle$ is the *cyclic subgroup generated by s*.

We also say a group is *cyclic* if it can be generated (as a subgroup of itself) by a single element.

This definition may sound a bit vague: it does not give an explicit way to construct $\langle S\rangle$. However, we can actually construct it.

First of all, $\langle S\rangle$ contains S by definition. But, because it is a sub*group*, it must contain the unit and products and inverses of all elements in S as well. So we may construct $\langle S\rangle$ by starting from S, taking all products and inverses and adding them, along with the unit, to obtain some set S_1, adding all products and inverses of elements of S_1 to obtain some S_2, &c. All of these S_i must be subsets of $\langle S\rangle$, and hence $\bigcup_{i=1}^\infty S_1 \subseteq \langle S\rangle$. But this union is closed under multiplication and inverses, hence is a group containing S. Therefore, it must equal $\langle S\rangle$.

Definition 1.2.10. Given a group G and a subgroup $H \subset G$, we define its *centraliser* $C_G(H) = \{g \in G \mid gh = hg \text{ for all } h \in H\}$, so the group of all elements of G which commute with all elements of H.

In case $H = G$, we write $Z(G) = C_G(G)$ and call it the *centre* of G.

Definition 1.2.11. Let G be a group and $g, h \in G$. The *commutator* of g and h is $[g, h] = ghg^{-1}h^{-1}$.
The *commutator subgroup* of G, written $[G, G]$ is the subgroup generated by all $[g, h]$.

The commutator of two elements is equal to the identity if and only if the two elements commute. This is where the name comes from.

The commutator subgroup $[G, G] = \{e\}$ if and only if G is abelian if and only if $G = Z(G)$.

Example 1.2.12. The centre of $\mathrm{GL}(n, \mathbb{R})$ is $Z(\mathrm{GL}(n, \mathbb{R})) = \mathbb{R}^\times \cdot \mathrm{Id}$, the subgroup of non-zero multiples of the identity. ⋄

Example 1.2.13. The centre of the quaternion group Q of example 1.2.8 is $\{1, -1\}$. ⋄

1.2.2 — Cosets

Now we know that groups may have subgroups, we can try and think about the relation between the two. For this, the notion of cosets is very useful.

Definition 1.2.14. Let G be a group with a subgroup $H \subset G$ and an element $g \in G$. Then

$$gH = \{gh \mid h \in H\} \tag{1.7}$$

is a *left coset* of H, while

$$Hg = \{hg \mid h \in H\} \tag{1.8}$$

is a *right coset* of H.

Example 1.2.15. Consider the group $G = S_3$ with subgroup $H = \{e, (1\,2)\}$. Then its left and right cosets are

$$\begin{aligned} eH &= (1\,2)H = \{e, (1\,2)\} & He &= H(1\,2) = \{e, (1\,2)\} \\ (1\,3)H &= (1\,2\,3)H = \{(1\,3), (1\,2\,3)\} & H(1\,3) &= H(1\,3\,2) = \{(1\,3), (1\,3\,2)\} \\ (2\,3)H &= (1\,3\,2)H = \{(2\,3), (1\,3\,2)\} & H(2\,3) &= H(1\,2\,3) = \{(2\,3), (1\,2\,3)\} \end{aligned}$$

Note any two left (or two right) cosets are either identical or disjoint, and they partition the entire group. This will turn out to be a general phenomenon.

The collections of left cosets and right cosets are different. This is not always the case – for an abelian group, they will clearly be identical, but even for a non-abelian group, they may be. ⋄

The non-trivial cosets of a subgroup, i.e. those that are not the subgroup itself, are not groups, but only sets. However, they are still useful in studying the subgroup. First, let us prove an observation we made in the previous example.

Proposition 1.2.16. *Let G be a group and $H \subseteq G$ a subgroup.*

1. *If $a, b \in G$, then $aH = bH$ if and only if $a^{-1}b \in H$. Similarly, $Ha = Hb$ if and only if $ba^{-1} \in H$.*
2. *Two left cosets of H are either disjoint or identical, and similar for right cosets.*
3. *Every element $g \in G$ is contained in exactly one left and one right coset of H.*

Proof. We give the proof only for left cosets.

1. If $aH = bH$, then $b = be \in bH = aH$, so $ah = b$ for some $h \in H$. Hence, $a^{-1}b = a^{-1}ah = h \in H$. Vice versa, if $a^{-1}b = h \in H$, then for any $h' \in H$, $bh' = aa^{-1}bh' = ahh' \in aH$ and similarly $ah' = b(a^{-1}b)^{-1}h' = bh^{-1}h' \in bH$.

2. Suppose aH and bH are not disjoint. Then they have a common element x, so $x = ah_1 = bh_2$ for some $h_1, h_2, \in H$. But then $a^{-1}b = h_1 h_2^{-1} \in H$, so by part 1., $aH = bH$.

3. Certainly $g \in gH$. As any other coset must be disjoint from gH, it cannot contain g. □

What this proposition tells us, is that cosets give an equivalence relation:

Definition 1.2.17. An *equivalence relation* on a set S (hence also on a group) is a relation $a \sim b$ between some of the elements of S which is

- *Reflexive*: for each $a \in S$, $a \sim a$;
- *Symmetric*: if $a \sim b$, then $b \sim a$;
- *Transitive*: if $a \sim b$ and $b \sim c$, then $a \sim c$.

Given an equivalence relation on S, it partitions S into disjoint subsets, called *equivalence classes*, of elements that are all equivalent.

By proposition 1.2.16, we can define an equivalence relation on G by $g \sim h$ if and only if they lie in the same (left) H-coset. The cosets are then the equivalence classes.

Using proposition 1.2.16, we find the following very useful theorem.

Theorem 1.2.18 (Lagrange's theorem). *Let G be a group and $H \subseteq G$ a subgroup. Then, for any $g \in G$, $|gH| = |H|$. In particular, if G is a finite group, $|H|$ divides $|G|$.*

Proof. The map $f\colon H \to gH\colon h \mapsto gh$ is a bijection with inverse $f^{-1}\colon gH \to H\colon gh \mapsto g^{-1}gh = h$. So certainly $|H| = |gH|$.

Now, if G is finite, by the previous proposition we can find some $g_1, \dots g_k \in G$ such that $G = g_1H \sqcup g_2H \sqcup \dots \sqcup g_kH$. Then $|G| = \sum_{i=1}^{k} |g_iH| = \sum_{i=1}^{k} |H| = k|H|$. □

The collection g_i in the above proof is called a *system of representatives* for the left cosets of H in G.

Definition 1.2.19. Let G be a group and $H \subseteq G$ a subgroup. Then the *index* of H in G, written as $[G : H]$, is the number of left (or right) cosets of H in G.

If G is of finite order, $[G : H] = |G|/|H|$. But the index may be finite even if $|G| = \infty$.

Corollary 1.2.20. *Let G be a group of prime order. Then G is cyclic and has no proper subgroups.*

Proof. By Lagrange's theorem, any subgroup of G must have an order dividing the order of G. But this is prime, so the only possible divisors are 1 and $|G|$, corresponding to the trivial subgroups.

Pick a non-identity element $g \in G$ (this is possible as 1 is not prime). Then the subgroup $\langle g \rangle$ is cyclic and not trivial, so it must be G itself. Hence, $G = \langle g \rangle$ is cyclic. □

1.2.3 — Direct products

Both subgroups and cosets are ways to make smaller groups or sets out of a given group. We could also combine several groups into one to make a larger group.

Definition 1.2.21. Let G and H be two groups. Then the *direct product* of G and H is given as a set by the Cartesian product

$$G \times H = \{(g, h) \mid g \in G, h \in H\}\,. \tag{1.9}$$

Its product is given *component-wise*:

$$(g_1, h_1) \cdot (g_2, h_2) = (g_1 g_2, h_1 h_2)\,. \tag{1.10}$$

Similarly, its unit is (e_G, e_H), where e_G is the unit of G and e_H the unit of H, and the inverse is given by

$$(g, h)^{-1} = (g^{-1}, h^{-1})\,. \tag{1.11}$$

EXAMPLE 1.2.22. The vector space $\mathbb{R}^n$, viewed as a group under addition, is the n-fold direct product of the group $\mathbb{R}$ with itself. So $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$, $\mathbb{R}^3 = \mathbb{R} \times \mathbb{R} \times \mathbb{R}$, &c. ◇

EXAMPLE 1.2.23. The Klein four-group is the simplest non-cyclic group. It is $\mathbb{Z}/2 \times \mathbb{Z}/2$, so its elements are $(\bar{0}, \bar{0}), (\bar{1}, \bar{0}), (\bar{0}, \bar{1}), (\bar{1}, \bar{1})$. Its unit is $(\bar{0}, \bar{0})$, and all three other elements are of order two.

This shows it is not cyclic, as any cyclic group has an element whose order is the order of the group.

Let us use this example to give a different way of giving the multiplication of a group: by a *multiplication table*. If we write $\{e, a, b, c\}$ for the elements of the Klein four-group (it does not matter how you label them, as long as e is the unit), then the multiplication table is given by

$\cdot$	e	a	b	c
e	e	a	b	c
a	a	e	c	b
b	b	c	e	a
c	c	b	a	e

In other words, it is a table where both rows and columns are indexed by elements of G and the (g, h)-th entry is the product gh. Needless to say, this becomes very bothersome for groups of order larger than, say, 10. ◇

EXAMPLE 1.2.24. By contrast, $\mathbb{Z}/2 \times \mathbb{Z}/3$ is cyclic: the element $(\bar{1}, \bar{1})$ generates it. (Note here that this element may be written $(1 \bmod 2, 1 \bmod 3)$.) Its powers – or rather multiples in additive terminology – are

$$\begin{aligned} 0(\bar{1},\bar{1}) &= (\bar{0},\bar{0}) & 3(\bar{1},\bar{1}) &= (\bar{1},\bar{0}) \\ 1(\bar{1},\bar{1}) &= (\bar{1},\bar{1}) & 4(\bar{1},\bar{1}) &= (\bar{0},\bar{1}) \\ 2(\bar{1},\bar{1}) &= (\bar{0},\bar{2}) & 5(\bar{1},\bar{1}) &= (\bar{1},\bar{2})\,. \end{aligned}$$ ◇

1.3 — HOMOMORPHISMS

In the previous section, we saw a way of relating two groups: one may be a subgroup of the other. There is a more general way of doing this, by mapping one to the other. However, not every map (of sets) between groups is meaningful, we would like the map to say something about the actual groups. Think of the same situation in linear algebra: we do not want to consider all maps from one vector space to another, but only the linear transformations. In mathematical terms, we would like our maps to "preserve the group structure". This is formalised by the following definition.

DEFINITION 1.3.1. Let G, H be two groups. A *group homomorphism* from G to H is a map $f\colon G \to H$ such that for any $g, g' \in G$,

$$f(gg') = f(g)f(g')\,. \tag{1.12}$$

Here the product on the left-hand side is taken in G and the product on the right-hand side in H.

An *endomorphism* of G is a homomorphism of G to itself. An *automorphism* of G is an isomorphism of G to itself.

EXAMPLE 1.3.2. If G is a group and $H \subseteq G$ is a subgroup, then the inclusion $i\colon H \to G\colon h \mapsto h$ is a homomorphism. ◇

EXAMPLE 1.3.3. Any linear transformation is a homomorphism of the underlying abelian groups. ◇

EXAMPLE 1.3.4. The determinant is a group homomorphism $\det\colon \mathrm{GL}(n, \mathbb{R}) \to \mathbb{R}^\times$, because $\det(AB) = \det(A)\det(B)$. Of course, this also holds over $\mathbb{C}$. ◇

PROPOSITION 1.3.5. *Let G and H be groups and $f: G \to H$ a homomorphism. Then $f(e_G) = e_H$, where e_G is the unit of G and e_H the unit of H.*

For any $g \in G$, $f(g^{-1}) = f(g)^{-1}$.

Proof. We have $f(e_G)f(e_G) = f(e_G \cdot e_G) = f(e_G)$, but also $e_H f(e_G) = f(e_G)$. But by theorem 1.1.19, the equation $xf(e_G) = f(e_G)$ has a unique solution, so $f(e_G) = e_H$.

Similarly, we find that both $f(g^{-1})$ and $f(g)^{-1}$ solve $xf(g) = e_H$, so they must be equal as well. □

DEFINITION 1.3.6. Let G and H be groups and $f: G \to H$ a homomorphism. The *kernel* of f, written $\mathrm{Ker}(f)$ is the set

$$\mathrm{Ker}(f) = \{g \in G \mid f(g) = e_H\}. \tag{1.13}$$

The *image* of f, written $\mathrm{Im}(f)$ is the set

$$\mathrm{Im}(f) = \{f(g) \mid g \in G\}. \tag{1.14}$$

EXAMPLE 1.3.7. There is a well-known homomorphism from $\mathbb{R}$ (with addition) to $\mathbb{R}^\times$ (with multiplication). It is given by the exponential function $\exp: \mathbb{R} \to \mathbb{R}^\times : x \mapsto e^x$. This is a homomorphism, because of the rule $e^{x+y} = e^x e^y$. Its image is $\mathbb{R}_{>0}$, the subgroup of positive reals.

In fact, the map $\exp: \mathbb{R} \to \mathbb{R}_{>0}$ is an isomorphism, with inverse given by the logarithm. ◇

EXAMPLE 1.3.8. For the group $\mathbb{Z}$, multiplication by any $n \in \mathbb{Z}$ is an endomorphism f_n, because $n(x+y) = nx + ny$ for any $n, x, y \in \mathbb{Z}$. If $n = \pm 1$, this is an automorphism, but otherwise the image is a strict subgroup – in case $n = 0$, it is even the trivial group. We write $n\mathbb{Z}$ for this subgroup.

On the other hand, the kernel of f_n is $\{0\}$ unless $n = 0$, in which case it is $\mathbb{Z}$. ◇

PROPOSITION 1.3.9. *Let G and H be groups and $f: G \to H$ a homomorphism. Then* $\mathrm{Ker}(f)$ *is a subgroup of G.*

Proof. We have already seen that $e_G \in \mathrm{Ker}(f)$, so we still have to check that if $a, b \in \mathrm{Ker}(f)$, then both ab and a^{-1} are in $\mathrm{Ker}(f)$ as well. For the first, we see

$$f(ab) = f(a)f(b) = e_H \cdot e_H = e_H ,$$

so this is indeed in the kernel. For the second,

$$f(a^{-1}) = f(a)^{-1} = e_H^{-1} = e_H ,$$

so this is in the kernel as well. □

PROPOSITION 1.3.10. *Let G and H be groups and $f: G \to H$ a homomorphism. Then f is injective if and only if* $\mathrm{Ker}(f) = \{e_G\}$.

Proof. If f is injective, there can be at most one element mapping to e_H. We already know that $f(e_G) = e_H$, so this must be the only element.

If $\mathrm{Ker}(f) = \{e_G\}$, let $a, b \in G$ be such that $f(a) = f(b)$. Then

$$f(ab^{-1}) = f(a)f(b)^{-1} = f(a)f(a)^{-1} = e_H ,$$

so $ab^{-1} = e_G$. But then $a = b$ by proposition 1.1.12. □

PROPOSITION 1.3.11. *Let G and H be groups and $f: G \to H$ a homomorphism. Then* $\mathrm{Im}(f)$ *is a subgroup of H.*

Proof. We have seen that $e_H \in \operatorname{Im}(f)$, so we have to check that if $x, y \in \operatorname{Im}(f)$, then both xy and x^{-1} are in $\operatorname{Im}(f)$ as well. Let $x = f(a)$ and $y = f(b)$ for some $a, b \in G$. Then

$$xy = f(a)f(b) = f(ab) \in \operatorname{Im}(f)$$

and

$$x^{-1} = f(a)^{-1} = f(a^{-1}) \in \operatorname{Im}(f)\,. \qquad \square$$

PROPOSITION 1.3.12. *Let G, H, K be groups and $f\colon G \to H$ and $g\colon H \to K$ group homomorphisms. Then their composition $g \circ f\colon G \to K\colon a \mapsto g(f(a))$ is a group homomorphism as well. If both f and g are isomorphisms, so is $g \circ f$.*

Proof. Let $a, b \in G$. Then

$$(g \circ f)(ab) = g\big(f(ab)\big) = g\big(f(a)f(b)\big) = g\big(f(a)\big) \cdot g\big(f(b)\big) = (g \circ f)(a)(g \circ f)(b)\,,$$

and this proves that $g \circ f$ is a homomorphism.

As we know that a composition of bijections is a bijection, the statement for isomorphisms follows. $\square$

PROPOSITION 1.3.13. *Let $f\colon G \to H$ be a group isomorphism. Then $f^{-1}\colon H \to G$ exists and is also a group isomorphism.*

Proof. The inverse of f exists (and is bijective) as a map of sets because f is bijective. So we still need to prove that f^{-1} is a homomorphism. For this, let $a, b \in H$. Then

$$f\big(f^{-1}(ab)\big) = ab = f\big(f^{-1}(a)\big) \cdot f\big(f^{-1}(b)\big) = f\big(f^{-1}(a) \cdot f^{-1}(b)\big)\,,$$

because f is a group homomorphism.

Because f is injective, if $f(x) = f(y)$, then $x = y$. Applying this, we get $f^{-1}(ab) = f^{-1}(a) \cdot f^{-1}(b)$, which is what we had to prove. $\square$

EXAMPLE 1.3.14. We have seen in example 1.2.24 that $\mathbb{Z}/2 \times \mathbb{Z}/3$ is cyclic, generated by $(\bar{1}, \bar{1})$. This means that there is an isomorphism $f\colon \mathbb{Z}/6 \to \mathbb{Z}/2 \times \mathbb{Z}/3\colon \overline{n} \mapsto n(\bar{1}, \bar{1})$. From example 1.2.24 we see that this map is well-defined, as $f(\overline{n+6}) = f(\overline{n})$.

We also see that $f(\overline{n} + \overline{m}) = (n+m)(\bar{1}, \bar{1}) = n(\bar{1}, \bar{1}) + m(\bar{1}, \bar{1}) = f(\overline{n}) + f(\overline{m})$, so f is a homomorphism.

Finally, it is straightforward to check that f is a bijection: it is a map between two groups of order six, and is surjective by example 1.2.24. ◇

EXAMPLE 1.3.15. Let us now consider the groups S_3 and D_6, the symmetric group and dihedral group. Both are of order six, as $|S_n| = n!$ and $|D_{2n}| = 2n$. They are also both non-abelian, so they cannot be isomorphic to $\mathbb{Z}/6 \cong \mathbb{Z}/2 \times \mathbb{Z}/3$ from the previous example. However, they are isomorphic to each other.

We can construct a map $f\colon D_6 \to S_3$ as follows. Recall that D_3 is the group of symmetries of a regular triangle. Therefore, any element of D_6 will map the vertices of this triangle to vertices, giving a permutation of these three vertices. If we label these vertices by $1, 2, 3$, such a permutation is an element of S_3. This gives the map f. Check yourself that this is an isomorphism. ◇

Before, in section 1.2.3, we have seen how to construct, from two groups G and H, their direct product $G \times H$. Now, we could wonder if there are any maps between this product and the two individual factors. It turns out there are some natural maps, in both directions, as is shown in the following proposition.

PROPOSITION 1.3.16. *Let G and H be two groups and $G \times H$ their direct product. Then there are natural inclusion and projection homomorphisms, which are given as follows:*

$$\iota_1\colon G \to G \times H\colon g \mapsto (g, e_H)\,, \qquad \iota_2\colon H \to G \times H\colon h \mapsto (e_G, h)\,, \tag{1.15}$$

$$\pi_1\colon G \times H \to G\colon (g, h) \mapsto g\,, \qquad \pi_2\colon G \times H \to H\colon (g, h) \mapsto h\,. \tag{1.16}$$

Furthermore, there is a natural isomorphism

$$\tau\colon G \times H \to H \times G\colon (g, h) \mapsto (h, g)\,. \tag{1.17}$$

Proof. Verify this for yourself. $\square$

1.4 — Actions

Recall that we introduced the symmetric groups S_n in example 1.1.18. We will now see why these groups are so important: they give rise to actions. In particular, this paves the way to Cayley's theorem.

Definition 1.4.1. Let G be a group and X a set. Write S_X for the symmetric group on X, i.e. the group of bijections $X \to X$. An *action* of G on X is a homomorphism $\rho\colon G \to S_X$. Equivalently, it is a map

$$\bar{\rho}\colon G \times X \to X\colon (g, x) \mapsto g.x \tag{1.18}$$

such that for all $g, h \in G$ and $x \in X$,

$$(gh).x = g.(h.x)\,. \tag{1.19}$$

In some sense actions are the reason that groups are interesting at all: groups encode symmetries, and these symmetries are given exactly by the action of the group on whatever is symmetric.

Example 1.4.2. The dihedral groups are defined via an action on a regular n-gon. Example 1.3.15 is an example of the restricted action of D_3 on the vertices of the regular triangle, and can be extended to an action of D_n on the vertices of the n-gon. ⋄

One specific kind of action is called a representation. We will not define these here, but they will be the main focus point of the second part of the course.

Definition 1.4.3. Let G be a group, X a set, and $\rho\colon G \to S_X$ an action. For an element $x \in X$, its *G-orbit* is

$$Gx = \{g.x \mid g \in G\}\,. \tag{1.20}$$

Proposition 1.4.4. *Let G be a group, X a set, and $\rho\colon G \to S_X$ an action. If $x, y \in X$, then either $Gx = Gy$ or Gx and Gy are disjoint.*

The proof is similar to that of proposition 1.2.16 – in fact that is a special case.

Proof. Suppose $Gx \cap Gy \neq \emptyset$. Then there is a $z \in Gx \cap Gy$, so there are $g, h \in G$ with $z = g.x$ and $z = h.y$. But then, $x = g^{-1}.z = (g^{-1}h).y$, and similarly $y = (h^{-1}g).x$. So if $w = g'.x \in Gx$, then $w = (g'g^{-1}h).y \in Gy$ and so $Gx \subseteq Gy$. The other inclusion is the same. □

Corollary 1.4.5. *The orbits of an action give equivalence classes for an equivalence relation (cf. definition 1.2.17) written as $x \sim_G y$. So $x \sim_G y$ if $x \in Gy$.*

Example 1.4.6. Let $X = \mathbb{R}^2$ and $G = \mathrm{O}(2) \subseteq \mathrm{GL}(2, \mathbb{R})$. There is a natural action of G on X – this is the way G is defined. As G consists of rotations and reflections, the orbits are circles centred at the origin. ⋄

Example 1.4.7. Let G be a group and $H \subseteq G$ a subgroup. Then H acts on G by $H \times G \to G\colon (h, g) \mapsto h.g = hg$, i.e. the action is the multiplication in G. The orbits under this action are exactly the right cosets, and this is convenient, as both would be written as Hg anyway. ⋄

We will use actions in two main examples in this part of the course: for Cayley's theorem and for the conjugation action, see section 1.4.1.

Theorem 1.4.8 (Cayley's theorem). *Let G be a group of finite order n. Then G is isomorphic to a subgroup of S_n.*

Proof. The idea of the proof is that we let G act on itself via the left multiplication maps λ_g given in the rearrangement theorem 1.1.19. These maps were given by

$$\lambda_g \colon G \to G \colon h \mapsto gh\,.$$

They permute all the elements of G by theorem 1.1.19, so we can interpret them as elements of S_G, the symmetric group on the *set* G. So we get a map

$$\Lambda \colon G \to S_G \colon g \mapsto \lambda_g\,.$$

This is an action: $\Lambda(gh)k = \lambda_{gh}(k) = ghk = \lambda_g \circ \lambda_h(k) = \Lambda(g)\Lambda(h)k$ for any $g, h, k \in G$. It is injective, because $\lambda_g(e) = e$ implies $g = e$. Therefore, it is a bijection on its image, hence an isomorphism.

As $S_X \cong S_Y$ for any two sets which are in bijection, we also see $S_G \cong S_n = S_{\{1,\dots,n\}}$. □

EXAMPLE 1.4.9. From Cayley's theorem, we can see that symmetric groups are quite important, so let us study them a bit more. We have seen that any permutation $\sigma \in S_n$ can be decomposed into disjoint cycles. But there are more ways to decompose permutations.

LEMMA 1.4.10. *Any cycle (and hence any permutation) is a product of transpositions.*

Proof. Indeed, $(i_1\, i_2\, i_3 \cdots i_r) = (i_1\, i_r)(i_1\, i_{r-1}) \cdots (i_1\, i_2)$, as can be checked by seeing how both act. □

This decomposition is very far from unique: for example, one could add the square of any transposition anywhere. However, the parity (even or odd) of the number of transpositions in such a decomposition is unique (a fact we will not prove).

DEFINITION 1.4.11. The *parity* of a permutation is the parity of the length of any decomposition into transpositions. This gives a homomorphism $\varepsilon \colon S_n \to \mathbb{Z}/2\mathbb{Z}$.

The kernel of ε is called the *alternating group*: $A_n = \mathrm{Ker}(\varepsilon)$. It is a subgroup by proposition 1.3.9.

The alternating group is generated by three-cycles: any even permutation can be decomposed into pairs of transpositions, and for these there are three cases:

$$\begin{aligned}
&(a\,b)(a\,b) = ()\,,\\
&(a\,b)(a\,c) = (a\,c\,b)\,,\\
&(a\,b)(c\,d) = (a\,b)(a\,c)(a\,c)(c\,d) = (a\,c\,b)(a\,c\,d)\,.
\end{aligned}$$

In general, the parity of an r-cycle is $\overline{r-1} \in \mathbb{Z}/2\mathbb{Z}$, by counting the transpositions in the formula in the proof of lemma 1.4.10. ◇

1.4.1 — CONJUGACY CLASSES

There is an interesting action of groups on themselves, given by conjugation.

DEFINITION 1.4.12. Let G be a group and $g, h \in G$. Then the *conjugation* of h by g is ghg^{-1}.

We may this of this conjugation from two perspectives: from the viewpoint of g or that of h. Both are useful, so let us take a look at them.

PROPOSITION 1.4.13. *Let G be a group and $g \in G$. Then conjugation by g defines an isomorphism*

$$c_g \colon G \to G \colon h \mapsto ghg^{-1}\,, \tag{1.21}$$

Hence $c_g \in S_G$, and in fact $c \colon G \to S_G \colon g \mapsto c_g$ is an action.

Proof. The map c_g is a homomorphism because

$$c_g(hk) = g(hk)g^{-1} = gh \cdot e \cdot kg^{-1} = gh \cdot g^{-1}g \cdot kg^{-1} = c_g(h)c_g(k)\,.$$

It is a bijection, because it has an inverse: $c_{g^{-1}}$. So indeed $c_g \in S_G$.

To prove $c\colon S \to S_G$ is an action, we need to show that $c_{gh} = c_g \circ c_h$. So let us have this act on a third element k:

$$c_{gh}(k) = (gh)k(gh)^{-1} = ghkh^{-1}g^{-1} = c_g(hkh^{-1}) = c_g(c_h(k)) = c_g \circ c_h(k)\,. \qquad \square$$

Definition 1.4.14. Two elements h, k of G are called *conjugate* if there exists a $g \in G$ such that $ghg^{-1} = k$. This is an equivalence relation, and the equivalence classes (the orbits for the conjugation action) are called *conjugacy classes*.

Conjugate elements behave similarly from a group-theoretic perspective. We will see some examples here, but many more will follow later, when talking about representation theory.

Lemma 1.4.15. *All elements of a conjugacy class have the same order.*

Proof. Let h and k be conjugate, say $c_g(h) = k$. By proposition 1.4.13, c_g is a homomorphism, so if $h^n = e$ for some n, then

$$k^n = c_g(h)^n = c_g(h^n) = c_g(e) = e$$

as well. Vice versa, if $k^m = e$ for some m, then

$$h^m = c_{g^{-1}}(k)^m = c_{g^{-1}}(k^m) = c_{g^{-1}}(e) = e$$

as well. $\square$

Lemma 1.4.16. *Any element in the centre of a group forms a one-element conjugacy class. In particular, the unit is always a one-element class, and in an abelian group, all conjugacy classes have one element.*

Proof. If $c \in Z(G)$, then for any $g \in G$, $gc = cg$ by definition. Therefore, $gcg^{-1} = cgg^{-1} = c$. $\square$

Example 1.4.17. The quaternion group Q from example 1.2.8 has five conjugacy classes: the centre is $\{1, -1\}$, by example 1.2.13, so this gives two classes $\{1\}$ and $\{-1\}$. The other three classes are $\{i, -i\}$, $\{j, -j\}$, and $\{k, -k\}$. This can be seen because any two elements of Q commute up to a sign (by the defining relations) and e.g. $c_j(i) = jij^{-1} = ji(-j) = -k \cdot -j = kj = -i$. ◇

1.4.2 — Conjugacy classes in the symmetric group

Lemma 1.4.18. *Let $\sigma, \tau \in S_n$. Then $c_\sigma(\tau)$ can be described as follows: when writing τ as a product of disjoint cycles, replace each number i by $\sigma(i)$.*

As an example, if $\tau = (1\,4\,3)(2\,5)$, then $c_\sigma(\tau) = \big(\sigma(1)\,\sigma(4)\,\sigma(3)\big)\big(\sigma(2)\,\sigma(5)\big)$.

Proof. By definition, $c_\sigma(\tau) = \sigma\tau\sigma^{-1}$ sends $\sigma(i)$ to $\sigma(\tau(i))$. So, either $\tau(i) = i$, in which case (i) is a one-cycle in τ and $(\sigma(i))$ is a one-cycle in $c_\sigma(\tau)$, or $i\,\tau(i)$ is part of a cycle of τ and $\sigma(i)\,\sigma(\tau(i))$ is part of a cycle of $c_\sigma(\tau)$. $\square$

Theorem 1.4.19. *The conjugacy classes of S_n are given exactly by all elements of the same cycle type, i.e. the same number of cycles of any length.*

Proof. By lemma 1.4.18, two conjugate elements will have the same cycle type.

On the other hand, if two elements have the same cycle type, we can match all of these cycles, and this gives us a possible element to conjugate by, with the same lemma. $\square$

We will go into this example a bit deeper in exercises.

1.5 — Normal subgroups and quotients

In this section, we will return to the idea of cosets we introduced in section 1.2.2. We found there that all cosets of a subgroup have the same size, and that the collection of cosets partitions the original group. We will now take a better look at this collection of cosets: it turns out that under certain conditions this is a group itself.

The first question you may now ask yourself is: how do we multiply two cosets? In fact, we may multiply any two subsets.

Definition 1.5.1. Let G be a group and let $S, T \subseteq G$ be two subsets. Then we define their *product* to be

$$ST = \{st \mid s \in S, t \in T\} . \tag{1.22}$$

Remark 1.5.2. This product is in general not the same as the cartesian product of sets (or the direct product of groups if both S and T happen to be groups). For example, $|ST| \leq |S||T|$, but there is no guarantee that $|ST| = |S||T|$: if we take $S = T = G$, then $ST = G$ as well.

This notion of product subsumes certain notation we have used before. For example, if $S = \{g\}$ contains a single element and $T = H$ is a subgroup, then $ST = \{g\}H = gH$ is a coset. If $S = \{g\}$ and $T = \{h\}$ both contain a single element, then $ST = \{g\}\{h\} = \{gh\}$, so this is just the product of the group.

Lemma 1.5.3. *Let G be a group and $H \subseteq G$ a subgroup. Then $H \cdot H = H$.*

Proof. Because H is a group, for any $h, k \in H$, $hk \in H$. This proves $H \cdot H \subseteq H$. On the other hand, for each h, $h = e \cdot h \in H \cdot H$, so $H \cdot H \supseteq H$. This proves equality. □

Lemma 1.5.4. *The multiplication of sets is associative: for any three subsets $S, T, U \subseteq G$, $(ST)U = S(TU)$.*

What do we mean by these triple products? Well, if S and T are subsets of G, then ST is as well. So we can take the product of ST with a third subset U to obtain $(ST)U$.

Proof. The set $(ST)U$ is defined by

$$(ST)U = \{au \mid a \in ST, u \in U\} .$$

If we now plug in the definition of ST, we find that

$$(ST)U = \{(st)u \mid s \in S, t \in T, u \in U\} .$$

We know we can get rid of parentheses for products of elements, so doing this similarly on the other side, we find

$$(ST)U = \{stu \mid s \in S, t \in T, u \in U\} = S(TU) .$$

□

Hence, we may write $STU = S(TU)$ and similarly for more factors.

1.5.1 — Normal subgroups

Even though we have now defined a product of subsets of a group, and hence also of cosets, this is not quite good enough. The problem is that we do not know if the product of two cosets is again a coset. Indeed, this is generally not the case, unless the left cosets and right cosets of the subgroup agree.

Definition 1.5.5. Let G be a group. A subgroup $N \subseteq G$ is *normal* if for any $g \in G$, $gN = Ng$, or equivalently $gNg^{-1} = N$.

A subgroup $N \subseteq G$ is normal exactly if all conjugates gng^{-1} of elements $n \in N$ are again in N. This last statement is $gNg^{-1} \subseteq N$, but this is sufficient, as we require it for all $g \in G$.

PROPOSITION 1.5.6. *A subgroup $N \subseteq G$ is normal if and only if it is the union of conjugacy classes.*

Proof. If N is normal, then all conjugates of elements of N are in N, so it is the union of all the conjugacy classes of its elements.

If N is a union of conjugacy classes, then for any $g \in G$, $gNg^{-1} = N$, as this holds by definition for any conjugacy class. □

EXAMPLE 1.5.7. Any subgroup of the centre of a group is normal, as all elements of the centre are one-element conjugacy classes by lemma 1.4.16. In particular, the centre itself is normal. Also, all subgroups of abelian groups are normal, as for abelian groups the centre is the entire group. ⋄

EXAMPLE 1.5.8. If G and H are two groups, the subgroups $G \times \{e_H\}$ and $\{e_G\} \times H$ of $G \times H$, which are the kernels of the maps π_2 and π_1 (or the images of the maps ι_1 and ι_2) of proposition 1.3.16, are normal. ⋄

EXAMPLE 1.5.9. The subgroup R of the dihedral group D_{2n} given by rotations (so no reflections) is a normal subgroup, isomorphic to $\mathbb{Z}/n\mathbb{Z}$. It must be normal, as it has index 2, and hence both its left and right cosets must be R and $D_{2n} \setminus R$. ⋄

PROPOSITION 1.5.10. *Let $f: G \to H$ be a homomorphism. Then* $\operatorname{Ker} f$ *is a normal subgroup of G.*

Proof. Let $h \in \operatorname{Ker} f$ and $g \in G$. We need to prove that $ghg^{-1} \in \operatorname{Ker} f$. This follows from the following calculation:

$$f(ghg^{-1}) = f(g)f(h)f(g^{-1}) = f(g)e_H f(g)^{-1} = e_H\,. \qquad \square$$

EXAMPLE 1.5.11. The alternating group $A_n \subset S_n$ is normal: it is the kernel of the parity homomorphism ε. ⋄

EXAMPLE 1.5.12. Consider the Klein four-group from example 1.2.23. By Cayley's theorem 1.4.8, it can be seen as a subgroup of S_4, and in fact it is realised as the subgroup $V_4 = \{(), (1\,2)(3\,4), (1\,3)(2\,4), (1\,4)(2\,3)\}$. It follows from theorem 1.4.19 that this subgroup is the union of two conjugacy classes: the unit one, and the class with two cycles of length two. Then it is a normal subgroup by proposition 1.5.6. ⋄

EXAMPLE 1.5.13. The special linear group $\mathrm{SL}(n, \mathbb{R})$ is a normal subgroup of $\mathrm{GL}(n, \mathbb{R})$: it is the kernel of the determinant $\det: \mathrm{GL}(n, \mathbb{R}) \to \mathbb{R}^\times$. ⋄

1.5.2 — QUOTIENTS

We claimed before that we introduced normal subgroups in order to have the collection of cosets be a group. This is the content of the next theorem.

THEOREM 1.5.14. *Let G be a group and $N \subseteq G$ a normal subgroup. Then the set of cosets $G/N = \{gN \mid g \in G\}$ is a group under multiplication of subsets from definition 1.5.1, called the* quotient group*. Its order is* $[G : N]$.

Proof. First, we need to show that the multiplication is well-defined, e.g. the product of two cosets is again a coset. Let $g, h \in G$. Then, using that N is normal, we have

$$gN \cdot hN = \{g\}N\{h\}N = \{g\}\{h\}N \cdot N = \{gh\}N = ghN\,,$$

so this is again a coset. We have also found that the multiplication can be written $gN \cdot hN = ghN$. From this, it follows that the unit of G/N is $N = eN$, the inverse of gN is $g^{-1}N$, and that all axioms hold. We write out the case of associativity:

$$(gN \cdot hN) \cdot kN = ghN \cdot kN = ghkN = gN \cdot hkN = gN \cdot (hN \cdot kN)\,. \qquad \square$$

There is a natural *quotient map* $q\colon G \to G/N\colon g \mapsto gN$, and this is a surjective homomorphism.

PROPOSITION 1.5.15. *Let G be a group and $N \subseteq G$ a normal subgroup. If $q\colon G \to G/N$ is the quotient map, then* $\operatorname{Ker} q = N$.

This proposition is the main 'point' of the quotient construction. We know from proposition 1.5.10 that all kernels are normal subgroups, and this gives the converse: any normal subgroup is a kernel – namely, it is the kernel of the associated quotient map.

Proof. We have to prove two things: $\operatorname{Ker} q \subseteq N$ and $\operatorname{Ker} q \supseteq N$. We will start with the second.

Let $g \in \operatorname{Ker} q$. Then, by definition of the kernel, $q(g) = gN$ is the unit in G/N, which is $eN = N$. But $gN = N$ exactly if $g \in N$, because otherwise $g \in gN$ but $g \notin N$. So $g \in N$, and hence $\operatorname{Ker} q \subseteq N$.

Conversely, pick $n \in N$. Then $q(n) = nN = N = eN$, by the rearrangement theorem 1.1.19. So $n \in \operatorname{Ker} q$, and hence $N \subseteq \operatorname{Ker} q$. ☐

On a heuristic meta level, this proposition is 'obvious' (although it really is not first time you see it, as you first have to learn to think on such a level – this is partially what this course will do). Namely, the kernel of q has to be some normal subgroup of G, and apart from the trivial normal subgroups $\{e\}$ and G, we are given exactly one other normal subgroup by the assumption of the proposition. So this 'must' be the answer, as it is the only 'natural' choice.

On a more concrete level, how should you think about the quotient group and the quotient map? Well, in proposition 1.2.16, we have seen that cosets (left or right, it does not matter as we are looking at a normal subgroup) form equivalence classes, and the definition of equivalence relations, definition 1.2.17, was set up in such a way that it mimics equality. The quotient group is the result of considering the equivalence relation are an actual equality. And the quotient map sends an element g of G to its equivalence class. As these partition G, this is a nicely well-defined map. You can thin of the quotient map of 'forgetting' the element g and only remembering in what class it lies.

EXAMPLE 1.5.16. The two trivial subgroups of a group G are always normal, and $G/\{e\} \cong G$ and $G/G \cong \{e\}$. ◇

EXAMPLE 1.5.17. In example 1.3.8 we found subgroups of $\mathbb{Z}$ of the form $n\mathbb{Z} = \{nk \mid k \in \mathbb{Z}\}$. As $\mathbb{Z}$ is abelian, these are automatically normal. Because this group uses additive notation, the cosets look like $m + n\mathbb{Z} = \{m + nk \mid k \in \mathbb{Z}\}$. But we have seen these before, back in example 1.1.7, where we denoted them $\overline{m}$ or $m \pmod n$. So we see that the cyclic groups, which we already denoted $\mathbb{Z}/n\mathbb{Z}$, are in fact quotients of $\mathbb{Z}$ with respect to the subgroups $n\mathbb{Z}$.

In the particular case of $n = 2$, the quotient map is the map which for every element $m \in \mathbb{Z}$ only remembers whether it is even or odd. This is not really different from other cases, except that English has words for the cosets of $2\mathbb{Z}$ and not for the cosets of $n\mathbb{Z}$ with $m > 2$. ◇

EXAMPLE 1.5.18. Another interesting example is the subgroup $\mathbb{Z} \subset \mathbb{R}$. Again, this is normal as the groups are abelian. One abstract way of think about the quotient group $\mathbb{R}/\mathbb{Z}$ is to only look 'after the period', but there is a better way.

We can define the map $f\colon \mathbb{R}/\mathbb{Z} \to \mathrm{U}(1)\colon x+\mathbb{Z} \mapsto e^{2\pi i x}$ (recall that $\mathrm{U}(1)$ is the group of complex numbers of unit norm, cf. example 1.2.6), which is well-defined as $e^{2\pi i k} = 1$ for any integer k. Furthermore, this map is bijective, and even an isomorphism, as

$$f(x + y + \mathbb{Z}) = e^{2\pi i(x+y)} = e^{2\pi i x} e^{2\pi i y} = f(x + \mathbb{Z}) f(y + \mathbb{Z}) \,.$$

Cf. also example 1.3.7, where we used the real exponent.

We can also interpret these complex numbers – or really complex 1×1 matrices – as real 2×2 matrices. This gives a map

$$g\colon \mathrm{U}(1) \to \mathrm{GL}(2, \mathbb{R})\colon e^{2\pi i x} \mapsto \begin{pmatrix} \cos(2\pi x) & -\sin(2\pi x) \\ \sin(2\pi x) & \cos(2\pi x) \end{pmatrix} ,$$

whose image is exactly $\mathrm{SO}(2)$, the group of orthogonal 2×2 matrices with unit determinant. As g is injective, $\bar{g}\colon \mathrm{U}(1) \to \mathrm{SO}(2)$ is an isomorphism.

So we have found that $\mathbb{R}/\mathbb{Z} \cong \mathrm{U}(1) \cong \mathrm{SO}(2)$. ⋄

Example 1.5.19. In example 1.5.12, we found a normal subgroup V_4 of order 4 in S_4. Its index is $[S_4 : V_4] = \frac{|S_4|}{|V_4|} = \frac{24}{4} = 6$, so that is also the order of the quotient group S_4/V_4. What is this group?

Well, there is another subgroup (not normal this time) in S_4, which is just S_3 seen as acting on $\{1, 2, 3, 4\}$ by leaving 4 invariant. We claim that each coset of V_4 contains exactly one element of this S_3.

We see this as follows: from the definition of V_4, each element sends 4 to a different element of $\{1, 2, 3, 4\}$. Namely,

$$\begin{aligned} ()4 &= 4 & \qquad (1\,2)(3\,4)4 &= 3 \\ (1\,3)(2\,4)4 &= 2 & \qquad (1\,4)(2\,3)4 &= 1\,. \end{aligned}$$

So for each coset $V_4\sigma$, there is exactly one element leaving 4 invariant, namely the element $\tau\sigma$ such that $\tau \in V_4$ and $\tau\big(\sigma(4)\big) = 4$. And this then is the unique element of S_3 in $V_4\sigma$.

Because of this, we find that the map

$$i\colon S_3 \to S_4/V_4\colon \sigma \mapsto V_4\sigma = \sigma V_4$$

is a bijection, and as it is also a homomorphism by construction (it is the composition of a subgroup inclusion and a quotient map), it is an isomorphism. ⋄

Example 1.5.20. We have seen in example 1.2.13 that the centre of the quaternion group Q is $\{\pm 1\}$. As the centre is a always a normal subgroup, we may calculate the quotient $Q/\{\pm 1\}$. The cosets here are $\{\pm 1\}$, $\{\pm i\}$, $\{\pm j\}$, and $\{\pm k\}$, and every non-trivial coset squares to the identity. This is enough to conclude that $Q/\{\pm 1\}$ is isomorphic to the Klein four-group of example 1.2.23. ⋄

Example 1.5.21. The commutator subgroup $[G, G]$ of a group G, defined in definition 1.2.11, is normal. Recall that $[G, G]$ is the group generated by all commutators $[g, h] = ghg^{-1}h^{-1}$ in G. If k is a third element in G, then

$$k[g, h]k^{-1} = kghg^{-1}h^{-1}k^{-1} = (kgk^{-1})(khk^{-1})(kgk^{-1})^{-1}(khk^{-1})^{-1} = [kgk^{-1}, khk^{-1}]\,.$$

Because $[G, G]$ is generated by commutators, this is not quite enough, but the argument can be completed by using that conjugation is a homomorphism (try to do this yourself).

The quotient $G/[G, G]$ is abelian, and in a specific sense the largest quotient of G which is abelian. Hence, it is called the *abelianisation* of G and written G^{ab}. (We will not go into this deeper.) ⋄

Quotients of groups give us a useful way of encoding the data of an abstract group in what is called a presentation. A presentation should be thought of as analogous to a basis of a vector space in linear algebra: it gives a couple of elements that generate the entire vector space. This choice is far from unique, but it is very useful for doing actual calculations.

Definition 1.5.22. Let G be a group. A subset $S \subseteq G$ is said to *generate* G if $\langle S \rangle = G$, i.e. if the subgroup generated by S is equal to all of G.

A *presentation* of G is a pair $\langle S \mid R \rangle$ of a generating set S and a generating set of relations R between elements of S.

This definition may seem quite abstract, so let us look at a few examples.

Example 1.5.23. The integers $\mathbb{Z}$ can be generated by a single element 1, with no relation, so $\mathbb{Z} = \langle 1 \mid \rangle$. ⋄

EXAMPLE 1.5.24. If $R = \emptyset$, we call $\langle S \mid \rangle$ the *free group* on S. These groups are very large in general: they contain all 'words' $s_1^{n_1} s_2^{n_2} s_3^{n_3} \cdots s_k^{n_k}$, for $s_i \in S$ and $n_i \in \mathbb{Z}$, only with the condition that $s_i \neq s_{i+1}$.

In general, we can view $\langle S \mid R \rangle$ as the quotient of the free group $\langle S \rangle$ by the normal subgroup generated by R. You do not have to know this, but it may give you some more insight. ◇

EXAMPLE 1.5.25. The cyclic group $\mathbb{Z}/n\mathbb{Z}$ can also be generated by a single element $\bar{1}$, but this has the relation $n \cdot \bar{1} = 0$. We would write this as $\mathbb{Z}/n\mathbb{Z} = \langle \bar{1} \mid n \cdot \bar{1} = 0 \rangle$.

More abstractly, and using multiplicative notation, we may write $\mathbb{Z}/n\mathbb{Z} \cong \langle g \mid g^n = e \rangle$. ◇

Often, when writing relations, we write them in the form $f(g, h, \dots) = e$. In that case we usually only write $f(g, h, \dots)$, and forget the $= e$ part in the notation, so $\mathbb{Z}/n\mathbb{Z} \cong \langle g \mid g^n \rangle$. Think of this as analogous to talking about a root of a polynomial $P(x)$ in stead of talking about a root of the equation $P(x) = 0$.

EXAMPLE 1.5.26. When introducing the dihedral group in example 1.2.7, we essentially saw a presentation for it, namely $D_n = \langle \rho, \sigma \mid \rho^n, \sigma^2, (\sigma\rho)^2 \rangle$. This last relation may be rewritten as $\sigma\rho = \rho^{-1}\sigma$ (using the second relation), and this tells us that we may write every element as $\rho^k \sigma^l$: every time we have a σ to the left of a ρ, we may commute it past. Then the other two relations tell us that we may restrict to $0 \leq k < n$ and $0 \leq l < 2$. ◇

EXAMPLE 1.5.27. An extreme case of a presentation would be to take $S = G$, i.e. take all elements of the group as generators, and let the relations be the entire multiplication table, i.e. for each $g, h \in G$, take the relation $g \cdot h = gh$. This is of course very inefficient. ◇

EXAMPLE 1.5.28. The *modular group* is the group

$$\mathrm{PSL}(2, \mathbb{Z}) = \mathrm{SL}(2, \mathbb{Z})/\{\pm \mathrm{Id}\} = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \,\middle|\, a, b, c, d \in \mathbb{Z}, ad - bc = 1 \right\} / \{\pm \mathrm{Id}\} .$$

This group has a faithful action on the upper half-plane $\mathbb{H} = \{z \in \mathbb{C} \mid \Im(z) > 0\}$ by Möbius transformations:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} .z = \frac{az + b}{cz + d} .$$

This group can be generated by two elements

$$S = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} , \qquad T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} ,$$

which act as $S.z = -\frac{1}{z}$, and $T.z = z + 1$, and in fact $\mathrm{PSL}(2, \mathbb{Z}) = \langle S, T \mid S^2, (ST)^3 \rangle$. ◇

One reason that presentations are useful is because it is easy to give a homomorphism from a presented group. This following proposition is not part of the course material, but may help your understanding.

PROPOSITION 1.5.29. *Let G be a group, presented as $G = \langle S \mid R \rangle$, and H be another group. Giving a homomorphism $f\colon G \to H$ is equivalent to giving a collection of elements $\{f(s) \mid s \in S\}$ such that $f(r) = e_H$ for all $r \in R$.*

If this looks very abstract to you, think of vector spaces. To give a linear map from a vector space V to a vector space W, it is sufficient to give the images of a basis of V, and extend linearly. If we have a spanning set of V (not necessarily linearly independent), to give a map to W, we must give the images of all elements of the spanning set, and check that the linear relations still hold.

Proof. If we have a homomorphism $f\colon G \to H$, then we know $f(g)$ for all $g \in G$, so certainly we know $f(s)$ for $s \in S$. Furthermore, $r = e_G$ in G, so $f(r) = f(e_G) = e_H$ in H.

Vice versa, if we have all $f(s)$ for $s \in S$, we can construct $f(g)$, because g is a word in the elements of S and their inverses, and we can use $f(st) = f(s)f(t)$ and $f(s^{-1}) = f(s)^{-1}$. However, g may be a word in elements of S in different ways. If $f(r) = e$ for all $r \in R$, this cannot happen, as R generated all relations in G.

More abstractly, giving $f(s)$ for all $s \in S$ gives a map $\tilde{f}\colon \langle S \mid \rangle \to H$, and enforcing $f(r) = e_H$ for all $r \in R$, ensures that $R \subseteq \operatorname{Ker} \tilde{f}$, so $\tilde{f}$ descends to $f\colon \langle S \mid \rangle / \langle R \rangle = \langle S \mid R \rangle \to H$. □

While a presentation gives us all information to determine a group in general, it is not easy to actually see what kind of group you are dealing with just from the presentation. It is even not easy to see if the group is finite or not.

Furthermore, given two presentations of groups, it is in general hard to find out whether the groups are isomorphic or not.

1.6 — The isomorphism theorems

A cornerstone of group theory – in fact of all abstract algebraic theories – are the isomorphism theorems. There are several of them (three by most counts), but we will mostly focus on the most important one here: the first isomorphism theorem.

Theorem 1.6.1 (First isomorphism theorem). *Let $f\colon G \to H$ be a group homomorphism. Then* $\operatorname{Im}(f) \cong G/\operatorname{Ker}(f)$.

Proof. Let us write $K = \operatorname{Ker}(f)$. By propositions 1.3.9 and 1.3.11, we know that both $K \subseteq G$ and $\operatorname{Im}(f) \subseteq H$ are subgroups. By proposition 1.5.10, K is a normal subgroup.

Now, we define the map

$$\bar{f}\colon G/K \to \operatorname{Im}(f)\colon gK \mapsto f(g)\,.$$

This map is well-defined, because if $gK = g'K$, then $g = g'n$ for some $n \in K$, and hence

$$\bar{f}(gK) = f(g) = f(g'n) = f(g')f(n) = \bar{f}(g'K)\,.$$

It is a homomorphism, because

$$\bar{f}(gg'K) = f(gg') = f(g)f(g') = \bar{f}(gK)\bar{f}(g'K)\,.$$

It is injective, because if $\bar{f}(gK) = e_H$, then $f(g) = e_H$, so $g \in K$, and $gK = K$ is the unit of G/K. It is surjective by constuction, hence bijective. This proves it is an isomorphism. □

This theorem can be used to give easier proofs of results we found before.

Example 1.6.2. Let us look back at example 1.5.18. With the first isomorphism theorem, we can look at the homomorphism $f\colon \mathbb{R} \to \mathbb{C}^\times : x \mapsto e^{2\pi i x}$. Then $\operatorname{Ker}(f) = \mathbb{Z}$ and $\operatorname{Im}(f) = \mathrm{U}(1)$, so we recover $\mathbb{R}/\mathbb{Z} \cong \mathrm{U}(1)$. ⋄

Example 1.6.3. Let G be a group and $g \in G$. Then we can define a homomorphism $f\colon \mathbb{Z} \to G\colon n \mapsto g^n$. The image of this map is $\langle g \rangle$, and the kernel depends on the order of g:

- If the order of g is infinite, then $\operatorname{Ker}(f) = \{e\}$, so we find $\langle g \rangle \cong \mathbb{Z}$.
- If the order of g is finite, say n, then $\operatorname{Ker}(f) = n\mathbb{Z}$, so $\langle g \rangle \cong \mathbb{Z}/n\mathbb{Z}$.

This explains the name cyclic subgroup for $\langle g \rangle$. $\diamond$

For completeness, let us also give the other two isomorphism theorems, without proof. (You can try yourself if you like, or find them somewhere.) They are non-examinable.

THEOREM 1.6.4 (SECOND ISOMORPHISM THEOREM). *Let G be a group, $H \subseteq G$ a subgroup, and $N \subseteq G$ a normal subgroup. Then*

1. *$H \cap N$ is a normal subgroup of H;*
2. *HN is a subgroup of G;*
3. *$H/(H \cap N) \cong HN/N$.*

Note that example 1.5.19 is a particular example of the second isomorphism theorem, with $G = S_4$, $N = V_4$, and $H = S_3$. In this case $H \cap N = \{e\}$, so the statement simplifies.

THEOREM 1.6.5 (THIRD ISOMORPHISM THEOREM). *Let G be a group, and $N, M \subseteq G$ two normal subgroups of G such that $N \subseteq M$. Then M/N is a normal subgroup of G/N. Conversely, any normal subgroup of G/N is of the form P/N for $P \subseteq G$ a normal subgroup containing N. Moreover, $(G/N)/(M/N) \cong G/M$.*

Chapter 2 — Representation theory

In the first part of this course, we introduced group theory, and gave many examples of groups and ways of constructing new groups from old. We also saw a way groups could 'interact' with non-groups, via actions. In this second part, we will focus on a very specific kind of action, namely by linear transformations on a vector space. These actions are called representations.

Why whould we do this? There are two main reasons, one mathematical and one physical. The mathematical reason is that the representations of a group actually tell us a lot about the group itself, so we can use them to study groups. Besides, linear transformations are just matrices, which we know very well. So we can use something we know well to study something we do not know quite as well.

The physical reason is quantum mechanics. Quantum mechanics is essentially a linear theory, as the Schrödinger equation is linear in the wave function. Therefore, symmetries of quantum-mechanical systems should also be linear. In fact, in the last part of the course, we will actually see that particles are representations of something called Lie groups, and that their properties (such as spin) can be derived from representation theory.

However, in this part of the course, we will focus on finite groups, as this is the easiest way to get to know the theory. Lie groups – which are essentially matrix groups – will be introduced later.

2.1 — Definitions and examples

2.1.1 — Review: vector spaces

We will work with complex vector spaces in this part of the course. Recall that abstractly a complex vector space can given by the following definition (which is different from what you have seen before).

Definition 2.1.1. A *complex vector space* is an additively written abelian group V with a scalar multiplication $\lambda\colon \mathbb{C} \times V \to V : (z, v) \mapsto zv$ such that

1. *Associativity:* for any $z, w \in \mathbb{C}$ and $v \in V$, we have $(zw)v = z(wv)$;
2. *Identity:* for any $v \in V$, we have $1v = v$;
3. *Left distributivity:* for any $z, w \in \mathbb{C}$ and $v \in V$, we have $(z + w)v = zv + wv$;
4. *Right distributivity:* for any $z \in \mathbb{C}$ and $u, v \in V$, we have $z(u + v) = zu + zv$.

Elements of a vector space are called *vectors*.

An important tool for talking about vector spaces is a basis.

Definition 2.1.2. Let V be a complex vector space. A *basis* is a set of *basis vectors* $\{e_i \in V \mid i \in I\}$ for some set I such that any vector $v \in V$ can be written as $v = \sum_{i\in I} c_i e_i$ in a unique way, where $c_i \in \mathbb{C}$.

If V has a finite basis, we call its size the *dimension* of V.

From linear algebra, we know that every vector space has a basis, and that all bases of a vector space have the same size, so the dimension is well-defined. However, it should be noted that a vector space generally has no distinguished basis. So when given a vector space, there is no given basis which is natural to work in.

However, we can go the other way. Given a finite set S, we may define a vector space which has S as its basis; it is by definition $\{\sum_{s\in S} c_s s \mid c_s \in \mathbb{C} \text{ for all } s\}$.

This sounds very abstract, but it is not. The space $\mathbb{C}^n$ is constructed this way: we take a set $\{e_1, \ldots, e_n\}$ and let $\mathbb{C}^n$ be the vector space on this set.

The second important notion from linear algebra is a linear transformation.

Definition 2.1.3. Let V and W be vector spaces. A *linear transformation* from V to W is a map $L\colon V \to W$ which

1. is a homomorphism of abelian groups;
2. satisfies $L(zv) = zL(v)$ for all $z \in \mathbb{C}$ and $v \in V$.

A linear transformation is called *invertible* if it is bijective. Then the inverse map is also a linear transformation.

Given linearly ordered bases $\{e_j\}$ of V and $\{f_k\}$ of W, we may represent a linear transformation T as a matrix, namely a block of matrix coefficients T_{jk} such that $T(e_j) = \sum_{k=1}^{\dim W} T_{jk} f_k$. As any finite-dimensional vector space (which is what we are mostly interested in) has a basis, and we can always just pick some order, any linear transformation can be given as a matrix. However, this depends on the basis.

In what follows, we will often have a distinguished basis for our vector spaces, or we will just choose one, but be aware that these are actual choices.

2.1.2 — Actual definitions and examples

As usual, let us start with the main definition of this part of the course.

Definition 2.1.4. Let G be a group. A *(complex, linear) representation* of G is a complex vector space V along with a homomorphism $\rho\colon G \to \mathrm{GL}(V)$. In other words, it is an action of G on V by linear transformations.

If V is finite-dimensional (which we will always assume, unless explictly stated otherwise), the dimension of V is called the *dimension of the representation*.

A representation is called *faithful* if ρ is injective. It is called *unfaithful* otherwise.

Choosing a basis, we may consider a representation as giving a collection of $n \times n$ matrices, one for each $g \in G$.

Remark 2.1.5. Quite often, a representation is given by only naming the vector space V. This is unfortunate, as a group may have several different representations on one vector space, but it is common practice. I will try to always be explicit about the actual map ρ.

Of course, we could define a representation over any base field in stead of $\mathbb{C}$, but we will restrict to this case. One reason for this is that this is the natural setting for quantum mechanics. The other is that we need the base field to be algebraically closed and/or of characteristic zero for some of the nice results to hold. (If you do not know what algebraically closed or characteristic zero means, do not worry, this will not be important.)

Example 2.1.6. For any group G, the *trivial representation* is given by $V = \mathbb{C}$, and $\rho(g) = 1$ for all $g \in G$. This is unfaithful, unless $G = \{e\}$. ◇

EXAMPLE 2.1.7. Consider the symmetric group S_n, and let V be an n-dimensional vector space with basis $\{e_1, \dots, e_n\}$. The *permutation representation* of S_n on V is given by $\pi(\sigma)e_i = e_{\sigma(i)}$. For an explicit example, take $n = 3$, and let $\sigma = (1\,2\,3)$. Then $\pi(\sigma)e_1 = e_2$, $\pi(\sigma)e_2 = e_3$, and $\pi(\sigma)e_3 = e_1$, so in this basis,

$$\pi(\sigma) = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} .$$

It is a faithful representation.

Matrices with only zeroes and ones and a single one in each row and column are called *permutation matrices* because of this. ◇

EXAMPLE 2.1.8. Recall that in the proof of Cayley's theorem 1.4.8, we found a homomorphism $\Lambda\colon G \to S_G$ for any finite group G. We may compose this with the permutation representation of example 2.1.7 to get what is called the *regular representation* $r_G = \pi \circ \Lambda$. Explicitly, we construct a vector space V with basis $\{e_g \mid g \in G\}$, and we define the representation by $r_G(h)e_g = e_{hg}$, extending linearly. Again, this representation is faithful.

When choosing an order of a basis elements, we see that this representation is also given by permutation matrices.

This representation is the most important one from a mathematical point of view; we will see why later. Intuitively, it contains all information of the group – the elements are the basis vectors, and the mutliplication is encoded in the representation – and it gives us the extra tools from linear algebra to study it. ◇

EXAMPLE 2.1.9. The dihedral groups D_{2n} are defined in example 1.2.7 as a subgroup of $\mathrm{GL}(2, \mathbb{R})$. As $\mathrm{GL}(2, \mathbb{R})$ is naturally a subgroup of $\mathrm{GL}(2, \mathbb{C})$, this gives a complex 2-dimensional representation $\delta\colon D_{2n} \to \mathrm{GL}(2, \mathbb{C})$, called the *defining representation*. Explicitly, it is given on generators by

$$\rho \mapsto \begin{pmatrix} \cos(2\pi/n) & -\sin(2\pi/n) \\ \sin(2\pi/n) & \cos(2\pi/n) \end{pmatrix} , \qquad \sigma \mapsto \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} .$$

Because this representation defines D_{2n} as a subgroup of $\mathrm{GL}(2, \mathbb{R})$, it is faithful.

Note that, although we actually described the dihedral groups as acting on $\mathbb{C}$ by $\rho\colon z \mapsto e^{2\pi i/n}z$ and $\sigma\colon z \mapsto \bar{z}$, this is *not* a one-dimensional complex representation, as conjugation is not $\mathbb{C}$-linear. ◇

EXAMPLE 2.1.10. Consider the cyclic groups $\mathbb{Z}/n\mathbb{Z}$. Because these are generated by one element $\bar{1}$, any representation ρ is determined by $\rho(\bar{1})$, as $\rho(\bar{k}) = \rho(k \cdot \bar{1}) = \rho(\bar{1})^k$. Moreover, we will need that $\rho(\bar{1})^n = \rho(\bar{n}) = \rho(\bar{0}) = \mathrm{Id}$.

Let us use this to find the one-dimensional representations of $\mathbb{Z}/n\mathbb{Z}$. We need to find elements $\rho(\bar{1})$ of $\mathrm{GL}(1, \mathbb{C}) = \mathbb{C}^\times$ whose n-th power is 1. But these are the n-th roots of unity, which we know are $e^{2\pi i j/n}$ for $j = 0, \dots, n-1$.

So we get n one-dimensional representations $\rho_j\colon \mathbb{Z}/n\mathbb{Z} \to \mathrm{GL}(1, \mathbb{C})\colon \bar{k} \mapsto e^{2\pi i jk/n}$.

When are these faithful? Exactly if $e^{2\pi i jk/n} \neq 1$ for $k = 1, \dots, n-1$, so if $jk \notin n\mathbb{Z}$ for $k = 1, \dots, n-1$. This is the case if and only if j and n are coprime, i.e. $\gcd(j, n) = 1$.

For example, ρ_0 is the trivial representation, and will not be faithful for $n > 1$.

On the other hand, ρ_1 is always faithful.

For $n = 3$, ρ_1 and ρ_2 are faithful.

For $n = 4$, ρ_1 and ρ_3 are faithful, but ρ_2 is not: $\rho_2(\bar{2}) = e^{2\pi i \cdot 2 \cdot 2/4} = 1$. ◇

EXAMPLE 2.1.11. We can combine the previous example with the parity homomorphism $\varepsilon\colon S_n \to \mathbb{Z}/2\mathbb{Z}$. Taking the non-trivial representation ρ_1 for $\mathbb{Z}/2\mathbb{Z}$, we then get a one-dimensional representation for S_n which is called the *sign representation*. It sends all elements of the alternating group A_n to 1 and all other elements to -1. ◇

In view of proposition 1.5.29, if we have a presentation of a group, $G = \langle S \mid R \rangle$, then a representation of G is given by a matrix for each element in S, subject to the relations in R.

Example 2.1.12. Let us now look at the group of integers, $\mathbb{Z}$. Because we have quotient maps $\mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$, we get a lot of one-dimensional representations from the previous examples. In fact, these are only the tip of the iceberg: for *every* $z \in \mathbb{C}^\times$, we get a representation $\rho^z : \mathbb{Z} \to \mathrm{GL}(1, \mathbb{C}) \colon k \mapsto z^k$.

There is nothing special about dimension one here. In fact, for any vector space V, and any element $A \in \mathrm{GL}(V)$, we get a representation $\rho^A \colon \mathbb{Z} \to \mathrm{GL}(V) \colon k \mapsto A^k$. ◇

This example of the integers tells us that $\mathbb{Z}$ has a lot of representations, far more than we found for $\mathbb{Z}/n\mathbb{Z}$. This is a general issue with infinite (or rather non-compact) groups. By contrast, for finite groups, representations are quite well-behaved. From now on, we will focus on representations of finite groups, to show that this theory is already very rich and useful.

2.2 — First properties of representations

As stated at the end of the previous section, we will restrict our attention to representation theory of finite groups for this part of the course.

2.2.1 — Transformations of representations

In order to understand the representation theory of a given finite group G, we would like to be able to, for example, list all of its different representations. But what do we mean by 'different' here? Or rather, when can we call two representations 'the same'? For groups, we had the notion of isomorphism, which told us when two groups are 'the same'. Now, if we have two representations (V, ρ) and (W, π) of the same dimension, and they have bases in which the matrices of the representations are equal, should we not want to call these the same?

More generally, in group theory, we had the notion of homomorphism, which were maps that are compatible with the group structure. We could try to find something similar for representations, and then later restrict to the case of isomorphisms.

Let us try to make this more precise. In the case of groups, a homomorphism $f \colon G \to H$ was a map such that $f(gh) = f(g)f(h)$. In other words, one may either first multiply in G and then use f, or first use f and then multiply in H.

For two representations (V, ρ) and (W, π) of G, and a transformation T between them, we want to do something similar: either first use the representation ρ, and then T, or first use T, and then the representation π. We will need T to be a linear transformation from $T \colon V \to W$ as vector spaces. Then let us pick a $g \in G$, and consider the following diagram

$$\begin{array}{ccc} V & \xrightarrow{T} & W \\ {\scriptstyle \rho(g)}\downarrow & & \downarrow{\scriptstyle \pi(g)} \\ V & \xrightarrow[T]{} & W \end{array}$$

What does this mean? Coming from the upper left corner, we can either first use $\rho(g)$ (go down) and then use T (go right). Or we can first use T (go right) and then $\pi(g)$ (go down). We want these to be the same. This gives the following definition.

Definition 2.2.1. Let G ba a finite group, and (V, ρ) and (W, π) two representations of G. A *transformation* or *intertwiner* from (V, ρ) to (W, π) is a linear transformation $T \colon V \to W$ such that for all $g \in G$, $T \circ \rho(g) = \pi(g) \circ T$ as linear maps $V \to W$.

If T is an invertible linear transformation, it is called an *equivalence*. In this case, we have $\rho(g) = T^{-1}\pi(g)T$.

If such an equivalence T exists, then (V, ρ) and (W, π) are called *equivalent*.

Another way of thinking about equivalences: if we have a linear transformation $\alpha\colon V \to V$, but expressed in two different bases E and F, the associated matrices α_E and α_F will in general be different. However, they are related by the change of basis matrix B: $\alpha_E = B^{-1}\alpha_F B$. The two matrices are then similar (or conjugate, in terms of section 1.4.1). Then equivalence of representations means that we have one change of basis that transforms *all* matrices $\rho(g)$ to the matrices $\pi(g)$ at the same time.

Example 2.2.2. Let G be a finite group, and take $V = \mathbb{C}$. Then two representations ρ, π of G on V are equivalent if and only if they are equal: because $\mathrm{GL}(V) = \mathrm{GL}(1, \mathbb{C}) = \mathbb{C}^\times$ is abelian, $T^{-1}\pi(g)T = \pi(g)$ for all g.

So in particular, we can already see that the representations ρ_j we found in example 2.1.10 are not equivalent for different j. ◇

Example 2.2.3. We could restrict the defining representation of the dihedral group example 2.1.9 to the cyclic subgroup $\langle\rho\rangle \cong \mathbb{Z}/n\mathbb{Z}$. Writing $j\colon \mathbb{Z}/n\mathbb{Z} \to D_{2n}\colon \bar{k} \mapsto \rho^k$, this is given by

$$\delta \circ j(\bar{k}) = \begin{pmatrix} \cos(2\pi k/n) & -\sin(2\pi k/n) \\ \sin(2\pi k/n) & \cos(2\pi k/n) \end{pmatrix} .$$

These matrices all commute (as they form an image of an abelian group) and are unitary, which means they can be simultaneously diagonalised over the complex numbers. (We will talk about this more later on.) In this particular case, we may take

$$T = \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix} ,$$

such that

$$T^{-1}(\delta \circ j)(\bar{k})T = \begin{pmatrix} e^{2\pi i k/n} & 0 \\ 0 & e^{-2\pi i k/n} \end{pmatrix} .$$

This idea, of simultaneous diagonalisation, will be useful later. We cannot quite do this in general, but we can do something like it. ◇

2.2.2 — New representations from old

In the previous chapter, we discussed a few ways of constructing new groups from ones we already had, such as taking subgroups or direct products of groups. we can do similar things for representations. For example, when is a subspace of a representation again a representation?

Definition 2.2.4. Let G be a group, and (V, ρ) a representation. A *subrepresentation* of (V, ρ) is a linear subspace $W \subseteq V$ such that for all $g \in G$ and $w \in W$, $\rho(g)(w) \in W$, i.e. $\rho(G)W = W$. We also say that W is an *invariant subspace*.

In other words, a subspace of a representation is a subspace which is *preserved* under the action of ρ.

Example 2.2.5. For any group G and representation (V, ρ) the subspaces $\{0\}, V \subseteq V$ are subrepresentations. They are the *trivial subrepresentations*. ◇

Definition 2.2.6. A representation is called *irreducible* or *simple* if it is non-trivial and has no non-trivial subrepresentations. Otherwise, it is called *reducible*.

Irreducible representations are in a way the building blocks of representations. They are the simplest examples (hence that name), and we can contruct others out of them, as we will see below. Irreducible represenations are usuful in physics as well: elementary particles behave in many respects as the irreducible representations of certain Lie groups.

The term 'irreducible representation' is central to representation theory, but it is also very long and cumbersome. Hence, it is often abbreviated, both in speech and in writing, to 'irrep'. I will not do this in these notes, but I will in the lectures.

EXAMPLE 2.2.7. The two-dimensional representation of D_{2n} given in example 2.1.9 is irreducible if $n \geq 3$: a non-trivial subrepresentation $W \subset V$ must be one-dimensional. That means it must be an eigenspace for both the actions of ρ and of σ, but these matrices do not have common eigenvectors.

However, if we consider the subgroup $\{e, \sigma\} \subseteq D_{2n}$, then we get a representation $\delta \circ i\colon \{e, \sigma\} \to \mathrm{GL}(2, \mathbb{C})$ which does have a subrepresentation: the vectors $\begin{pmatrix}1\\0\end{pmatrix}$ and $\begin{pmatrix}0\\1\end{pmatrix}$ are eigenvectors of $\delta(\sigma)$, so their eigenspaces are subrepresentations.

We have also seen in example 2.2.3 that the subgroup generated by ρ also has two non-trivial subrepresentations, the lines given by the basis vectors $\begin{pmatrix}1\\i\end{pmatrix}$ and $\begin{pmatrix}1\\-i\end{pmatrix}$.

So these latter two representations are reducible. ◇

In this previous example, we saw a few reducible representations which actually decomposed completely, in the sense that we found two different one-dimensional representations that together formed the two-dimensional one. This is a bit analogous to the direct product we saw for groups. Here, we call it the direct sum.

DEFINITION 2.2.8. Let G be a finite group and let (V, ρ) and (W, π) be two representations of G. Then their *direct sum* has the underlying vector space $V \oplus W = V \times W$, and the representation is given by

$$(\rho \oplus \pi)(g)(v, w) = (\rho(g)v, \pi(g)w)\,. \tag{2.1}$$

In terms of matrices, this means that the representation is block-diagonal:

$$(\rho \oplus \pi)(g) = \begin{pmatrix}\rho(g) & 0\\ 0 & \pi(g)\end{pmatrix}\,. \tag{2.2}$$

Direct sum representations are somehow the obvious reducible representations: by definition the break apart in two subrepresentations, given by the inclusions of V and W in $V \oplus W$. We could now ask the question: can we always do this for a reducible representation? In other words, if we have a subrepresentation W of a representation (V, ρ), can we always find a complement? This question is not quite as simple as it looks at first sight: yes, we can find a complement to W as a vector space by completing the basis of W to a basis of V, but this will in general not be an invariant subspace. How do we even find an invariant complement, if it exists?

First, let us give a definition for the nice situation. We will return to this question later.

DEFINITION 2.2.9. A representation is called *semisimple* or *completely reducible* if it is the direct sum of irreducible subrepresentations.

Returning to our question at the start of this section, of trying to list all different representations of a group G, we now see that there will be infinitely many: given any two representations, we can take their direct sum. In this sense, maybe we should restrict to only the simple representations, at least is we can show that all representations are semisimple.

This makes sense from a classification standpoint. But semisimple representations do appear quite often even when studying irreducible representations. We saw an example of this in example 2.2.7: when restricting an irreducible representation to a subgroup, it became reducible. We will see more applications both in physics and in mathematics.

There exists another way to combine representations. If the direct sum acts like a sum (indeed the dimensions add), there is also an operation which behave more as a product. This is called the tensor product. The definition of tensor products is quite complicated, but it turns out to be very useful in quantum physics: particles are representations of certain groups, and if we put two particles together, we do this mathematically by taking the tensor product of representations. I will not give the complete abstract definition of tensor products here, but give a good working definition.

DEFINITION 2.2.10. Let V and W be two vector spaces, with dimensions $\dim V = n$ and $\dim W = m$ and given bases $\{e_i \mid 1 \le i \le n\}$ and $\{f_k \mid 1 \le k \le m\}$, respectively. The *tensor product* of V and W, written as $V \otimes W$, is the vector space with basis $\{e_i \otimes f_k \mid 1 \le i \le n, 1 \le k \le m\}$. The number of basis elements, and hence the dimension of $V \otimes W$, is $\dim(V \otimes W) = nm$.

Now, let G be a finite group, and (V, ρ) and (W, π) to representations, with the same given bases. Let us write $\rho(g)_{i,j}$ for the matrix coefficients of $\rho(g)$ in the basis $\{e_i\}$ and $\pi(g)_{k,l}$ for the matrix coefficients of $\pi(g)$ in the basis $\{f_k\}$. Then the *tensor product representation* is $(V \otimes W, \rho \otimes \pi)$, where $\rho \otimes \pi$ has the block matrix form

$$(\rho \otimes \pi)(g) = \begin{pmatrix} \rho(g)_{1,1}\pi(g) & \cdots & \rho(g)_{1,n}\pi(g) \\ \vdots & \ddots & \vdots \\ \rho(g)_{n,1}\pi(g) & \cdots & \rho(g)_{n,n}\pi(g) \end{pmatrix} . \tag{2.3}$$

Here, every entry is an $m \times m$ block, given by expanding $\pi(g)$ in coefficients as well.

For two elements $v = \sum_{i=1}^n v_i e_i \in V$ and $w = \sum_{k=1}^m w_k f_k$, their tensor product is $v \otimes w = \sum_{i=1}^n \sum_{k=1}^m v_i w_k \, e_i \otimes f_k \in V \otimes W$. Such elements are sometimes called *pure tensors*. But not all elements of $V \otimes W$ are of this form, e.g. $e_1 \otimes f_1 + e_2 \otimes f_2$ is not. In quantum physics, if V and W are the state spaces of two particles, these non-pure tensors correspond to entangled particle states.

On a pure state, the tensor product representation acts as

$$(\rho \otimes \pi)(g)(v \otimes w) = (\rho(g)v) \otimes (\pi(g)w) . \tag{2.4}$$

The definition may be hard to unpack. Let us look at an example:

EXAMPLE 2.2.11. In the above definition, let $n = m = 2$. Then both $\rho(g)$ and $\pi(g)$ are 2×2 matrices, and $(\rho \otimes \pi)(g)$ will be a 4×4 matrix. If

$$\rho(g) = \begin{pmatrix} \rho(g)_{1,1} & \rho(g)_{1,2} \\ \rho(g)_{2,1} & \rho(g)_{2,2} \end{pmatrix} , \qquad \pi(g) = \begin{pmatrix} \pi(g)_{1,1} & \pi(g)_{1,2} \\ \pi(g)_{2,1} & \pi(g)_{2,2} \end{pmatrix} ,$$

then

$$(\rho \otimes \pi)(g) = \begin{pmatrix} \rho(g)_{1,1}\pi(g)_{1,1} & \rho(g)_{1,1}\pi(g)_{1,2} & \rho(g)_{1,2}\pi(g)_{1,1} & \rho(g)_{1,2}\pi(g)_{1,2} \\ \rho(g)_{1,1}\pi(g)_{2,1} & \rho(g)_{1,1}\pi(g)_{2,2} & \rho(g)_{1,2}\pi(g)_{2,1} & \rho(g)_{1,2}\pi(g)_{2,2} \\ \rho(g)_{2,1}\pi(g)_{1,1} & \rho(g)_{2,1}\pi(g)_{1,2} & \rho(g)_{2,2}\pi(g)_{1,1} & \rho(g)_{2,2}\pi(g)_{1,2} \\ \rho(g)_{2,1}\pi(g)_{2,1} & \rho(g)_{2,1}\pi(g)_{2,2} & \rho(g)_{2,2}\pi(g)_{2,1} & \rho(g)_{2,2}\pi(g)_{2,2} \end{pmatrix} . \qquad \diamond$$

Note that this tensor product is not a usual matrix product!

We will see that, even if (V, ρ) and (W, π) are irreducible, their tensor product may well be reducible. This is one of the reasons it may be useful to look at reducible representations.

In the case that $\dim V = 1$ or $\dim W = 1$, the tensor product $\rho(g) \otimes \pi(g)$ is just multiplication of a matrix by a number.

2.3 — Unitary and irreducible representations

2.3.1 — Review: unitary and Hermitian matrices

In order to proceed, we will need some complex linear algebra. This is not part of the course, but you may not have seen it before, and we do need it.

Definition 2.3.1. Let A be a complex matrix. Its *adjoint* $A^\dagger$ is its conjugate transpose: if $A = (a_{ij})$, then $A^\dagger = (\bar{a}_{ji})$. This is often also denoted by A^*.

The complex analogue of symmetric matrices is the following.

Definition 2.3.2. A square complex matrix A is *Hermitian* or *self-adjoint* if $A = A^\dagger$.

The fundamental result on Hermitian matrices is the following:

Theorem 2.3.3. *All eigenvalues of a Hermitian matrix are real. Any Hermitian matrix has an orthogonal basis of eigenvectors.*

Because of this, Hermitian matrices are of the utmost importance for quantum mechanics: observables are given by Hermitian matrices and measurements may return their eigenvalues.

Orthogonal matrices also have a complex analogue.

Definition 2.3.4. A matrix $A \in \mathrm{GL}(n, \mathbb{C})$ is *unitary* if $AA^\dagger = \mathrm{Id}$.

Just like orthogonal matrices turn one orthogonal basis into another in the real case, unitary matrices do the same in the complex case. The reason for the difference is that the standard inner product on $\mathbb{C}^n$ is $\langle v, w\rangle = v^\dagger w$, i.e. the entries of v have to be conjugated. A matrix U is then unitary if $\langle Uv, Uw\rangle = \langle v, w\rangle$, as can be seen by using that $U^\dagger = U^{-1}$.

One important property of unitary matrices is that they are diagonalisable and their eigenvalues lie on the unit circle, i.e. are of the shape $e^{i\vartheta}$ for some $\vartheta \in \mathbb{R}$.

Hermitian and unitary matrices are strongly related: any Hermitian matrix H can be diagonalised by a unitary matrix U, i.e. $U^\dagger H U$ is diagonal.

2.3.2 — The unitarity theorem

We will use the above notions for our study of representation theory.

Definition 2.3.5. A representation (V, ρ) of a group G is *unitary* if all $\rho(g)$ are, for $g \in G$.

Remark 2.3.6. Actually, this definition does not quite make sense. We need V to be an inner product space and all $\rho(g)$ to be unitary with respect to the inner product. We will gloss over this in future, and just assume all of this to be given.

Example 2.3.7. For a finite group G, any one-dimensional representation is unitary. Indeed, any $g \in G$ has finite order, say n. Then, writing $\rho(g) = re^{i\vartheta}$ in polar form, $1 = \rho(g)^n = r^n e^{in\vartheta}$, so $r = 1$. But then $\rho(g)^\dagger \rho(g) = e^{-i\vartheta} e^{i\vartheta} = 1$, so $\rho(g)$ is unitary. ◇

For higher dimensions, this is not the case:

Example 2.3.8. Consider the cyclic group of order 3, $\mathbb{Z}/3\mathbb{Z}$. It has a non-unitary two-dimensional representation, defined by setting $\rho(\bar{1}) = \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}$. ◇

However, we do know that, for a finite group G and a representation (V, ρ) of G, the eigenvalues of $\rho(g)$ must be roots of unity for any $g \in G$ – because the order of g must be finite. In particular, they have norm 1. As we also know that the eigenvalues of unitary matrices have norm 1, we can expect the matrices $\rho(g)$ to be similar to unitary matrices, and hopefully we can make this similarity transform uniformly for all $g \in G$.

This is the content of the following theorem.

THEOREM 2.3.9 (UNITARITY THEOREM). *Any finite-dimensional representation (V, ρ) of a finite group G is equivalent to a unitary representation.*

Proof. We will construct a unitary representation equivalent to ρ.

We first introduce a matrix to help us in the proof:

$$H = \sum_{g \in G} \rho(g)^\dagger \rho(g) .$$

Because G is finite, this is well-defined.

This H has the property that for any $h \in G$,

$$\begin{aligned}
\rho(h)^\dagger H \rho(h) &= \sum_{g \in G} \rho(h)^\dagger \rho(g)^\dagger \rho(g) \rho(h) \\
&= \sum_{g \in G} \rho(gh)^\dagger \rho(gh) \\
&= \sum_{k \in G} \rho(k)^\dagger \rho(k) \\
&= H ,
\end{aligned}$$

by using the rearrangement theorem 1.1.19.

By construction, $H^\dagger = H$, so H is Hermitian. Hence, there exist a unitary matrix U and a diagonal matrix D with real entries such that $D = U^\dagger H U$. Moreover, we find that if we write $A(g) = U^\dagger \rho(g) U$, then

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

Therefore, the j-th diagonal entry of D is given by $\sum_{g \in G} \sum_{i=1}^{\dim V} \overline{A(g)_{ij}} A(g)_{ij} > 0$.

Because of this, we may define $D^{1/2}$ and $D^{-1/2}$ by taking the positive square roots of all diagonal entries. We now define a new representation of G by $\pi(g) = D^{1/2} A(g) D^{-1/2}$. This is equivalent to ρ by construction, and we claim it is also unitary. This follows by using that $D^{1/2}$ is Hermitian:

$$\begin{aligned}
\pi(g)^\dagger \pi(g) &= D^{-1/2} U^\dagger \rho(g)^\dagger U D^{1/2} \cdot D^{1/2} U^\dagger \rho(g) U D^{-1/2} \\
&= D^{-1/2} U^\dagger \rho(g)^\dagger U D U^\dagger \rho(g) U D^{-1/2} \\
&= D^{-1/2} U^\dagger \rho(g)^\dagger H \rho(g) U D^{-1/2} \\
&= D^{-1/2} U^\dagger H U D^{-1/2} \\
&= D^{-1/2} D D^{-1/2} \\
&= \mathrm{Id} .
\end{aligned}$$

□

This result does not hold for non-finite groups! The problem in the proof is that H cannot be defined. For the integers $\mathbb{Z}$, the representation given by

$$\tau(k) = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix} . \tag{2.5}$$

is not unitary, and as these matrices are not diagonalisable, they are not similar to unitary matrices either.

However, it does hold for compact groups: there we can replace the sum by an integral.

2.3.3 — Irreducible representations

In this section, we will study the irreducible representations of a finite group G. We will see on the one hand that they are all finite-dimensional, and conversely, we will prove that all finite-dimensional representations are semisimple, i.e. decompose into irreducible representations.

We will start with the first result.

Theorem 2.3.10. *Let G be a finite group. Then any irreducible representation of G is finite-dimensional.*

Proof. We will consider any representation (V, ρ) of G and construct a finite-dimensional subrepresentation. If (V, ρ) is infinite-dimensional, this proves it is reducible.

Let $v \in V \setminus \{0\}$ be any non-zero vector, and consider the subspace W spanned by $\{\rho(h)v \mid h \in G\}$, the orbit of v under ρ. Because G is finite, W must be finite-dimensional. We claim this is a subrepresentation of (V, ρ).

To prove this, take any $w \in W$, which we can write as $w = \sum_{h \in G} \lambda_h \rho(h) v$ for some $\lambda_h \in \mathbb{C}$. Then, for any $g \in G$,

$$\rho(g)w = \rho(g) \sum_{h \in G} \lambda_h \rho(h) v = \sum_{h \in G} \lambda_h \rho(gh) v \,.$$

This is again in W, which proves that W is a subrepresentation. □

Why did we consider unitary representations in the previous section? Well, it gives us some more information on representations, and we can hope to use this extra information. Here, we will do just that: we will prove that all finite-dimensional unitary representations are semisimple, and hence all finite-dimensional representations are.

Theorem 2.3.11. *Any finite-dimensional unitary representation of any finite group is semisimple.*

Proof. Let G be a finite group and (V, ρ) a finite-dimensional unitary representation of G. We want to prove that V decomposes into irreducible subrepresentations. We will do this by induction: given any subrepresentation W, we construct a complement which is also a subrepresentation. Because V is finite-dimensional, we can only decompose it finitely many times, and we will end up with a decomposition into irreducible subrepresentations.

In general, finding a suitable complement is hard, there are too many options. This is where unitarity comes in: recall that unitary transformations are 'complex orthogonal matrices'. In other words, they preserve orthogonal bases. Therefore, we can look at the orthogonal complement $W^\perp$ of W. We will show this is an invariant subspace. Recall that

$$W^\perp = \{v \in V \mid \langle v, w \rangle = 0 \text{ for all } w \in W\} \,.$$

We want to show that if $v \in W^\perp$ and $g \in G$, then $\rho(g)v \in W^\perp$. So we will have to prove that $\langle \rho(g)v, w \rangle = 0$ for all $w \in W$. First, we calculate

$$\langle \rho(g)v, w \rangle = v^\dagger \rho(g)^\dagger w = v^\dagger \rho(g)^{-1} w = v^\dagger \rho(g^{-1}) w = \langle v, \rho(g^{-1}) w \rangle \,.$$

But we know that W is a subrepresentation, so $\rho(g^{-1})w \in W$, and therefore, $\langle \rho(g)v, w \rangle = \langle v, \rho(g^{-1})w \rangle = 0$. This proves that $\rho(g)v \in W^\perp$, and hence $W^\perp$ is a complementary subrepresentation of W in V. □

Corollary 2.3.12 (Maschke's theorem). *Any finite-dimensional representation of any finite group is semisimple.*

Proof. Let G be a finite group, and (V, ρ) a finite-dimensional representation of G. By the unitarity theorem 2.3.9, this representation is equivalent to a unitary representation (W, π) by some equivalence $T\colon V \to W$, i.e. $\rho(g) = T^{-1}\pi(g)T$. But by theorem 2.3.11, (W, π) is semisimple, so it decomposes as

$$W = \bigoplus_{j=1}^{k} W_j\,,$$

where all W_j are irreducible subrepresentations. Therefore, V decomposes as

$$V = \bigoplus_{j=1}^{k} T^{-1}W_j\,,$$

where $T^{-1}W_j = \{T^{-1}w \mid w \in W_j\}$ are irreducible subrepresentations, because they are equivalent (by T) to the W_j. Hence, this is also a decomposition into irreducible subrepresentations. □

Actually, this result does not require the unitarity theorem 2.3.9 – it holds over any field whose characteristic does not divide the order of the group. An alternative, more general, proof goes as follows.

Proof (of corollary 2.3.12). The setup is the same as for the proof of theorem 2.3.9: let G be a finite group, (V, ρ) a representation of G, and W a subrepresentation. We will find a complementary subrepresentation.

First, pick $U_0 \subseteq V$ to be *any* complementary subspace. This does not have to be a subrepresentation. Then any $v \in V$ can be written in a unique way as $v = w + u$, with $w \in W$ and $u \in U$. Define a projection $p_0\colon V \to W\colon w + u \mapsto w$. This is linear.

Now, we average over G: we define another projection

$$p\colon V \to W\colon v \mapsto \frac{1}{|G|}\sum_{g\in G} \rho(g) \circ p_0 \circ \rho(g)^{-1}(v)\,. \tag{2.6}$$

This is an intertwiner:

$$\begin{aligned} p \circ \rho(g)(v) &= \frac{1}{|G|}\sum_{h\in H} \rho(h) \circ p_0 \circ \rho(h)^{-1} \circ \rho(g)(v) \\ &= \frac{1}{|G|}\sum_{h\in H} \rho(h) \circ p_0 \circ \rho(g^{-1}h)^{-1}(v) \\ &= \frac{1}{|G|}\sum_{k\in H} \rho(gk) \circ p_0 \circ \rho(k)^{-1}(v) \\ &= \rho(g) \circ \frac{1}{|G|}\sum_{k\in H} \rho(k) \circ p_0 \circ \rho(k)^{-1}(v) \\ &= \rho(g) \circ p(v)\,, \end{aligned}$$

where we re-indexed the sum by $k = g^{-1}h$ using the rearrangement theorem 1.1.19.

Since it is an intertwiner, its kernel $\operatorname{Ker} p$ is a subrepresentation of V (this uses lemma 2.4.1, but there is no circular reasoning), and this is clearly a complement to W, so it is the complementary subrepresentation we were looking for. □

Again, this result does not hold for infinite groups. Again using the representation of the integers by equation (2.5), this representation is reducible, as $\begin{pmatrix}0\\1\end{pmatrix}$ is an eigenvector, but the matrix is not diagonalisable, and therefore this subrepresentation does not have a complement.

2.4 — Schur's lemma

In the previous section, we saw that every finite-dimensional representation of a finite group decomposed into irreducible subrepresentations. So these irreducible representations are the building blocks of all representation theory of finite groups. In order to understand representation theory, we should then focus on studying irreducibles.

In other parts of this course, we have found that, to understand something (groups, representations), it may be useful to also consider maps (homomorphisms, transformations) between these things. Now we know that all representations decompose into irreducibles, this also tells use there is a block decomposition of any transformation between representations, where the blocks represent the transformations between the irreducible components. So we may try to investigate what these blocks look like. Schur's lemma tells us essentially that these blocks are as simple as can be: any map between irreducible representations is either zero, or an equivalence. Moreover, the equivalences are not too complicated either: any transformation from an irreducible representation to itself is a multiple of the identity.

First, we give an introductory result.

Lemma 2.4.1. *Let G be a group, and let (V, ρ) and (W, π) be two representations of G. Let $T\colon V \to W$ be a transformation of representations. Then $\operatorname{Ker} T \subseteq V$ and $\operatorname{Im} T \subseteq W$ are subrepresentations.*

Proof. By definition of a transformation, $T\rho(g) = \pi(g)T$ for all $g \in G$.

Let us first consider $\operatorname{Ker} T$. Let $v \in \operatorname{Ker} T$ and $g \in G$. We have to prove that $\rho(g)v \in \operatorname{Ker} T$. But we know that $T\rho(g)v = \pi(g)Tv = \pi(g)0 = 0$, so this is indeed true. Hence, $\operatorname{Ker} T$ is a subrepresentation.

For $\operatorname{Im} T$, let us now take $w \in \operatorname{Im} T$, i.e. $w = Tv$ for some $v \in V$, and let again $g \in G$. Then $\pi(g)w = \pi(g)Tv = T\rho(g)v \in \operatorname{Im} T$, so $\operatorname{Im} T$ is also a subrepresentation. □

Lemma 2.4.2 (Schur's lemma). *Let G be a group, and let (V, ρ) and (W, π) be two irreducible representations of G. Let $T\colon V \to W$ be a transformation of representations. Then either $T = 0$ or T is invertible.*

If $(V, \rho) = (W, \pi)$ is finite-dimensional (e.g. if G is finite), then $T = \lambda \mathrm{Id}$ for some $\lambda \in \mathbb{C}$.

Proof. By lemma 2.4.1, $\operatorname{Ker} T \subseteq V$ and $\operatorname{Im} T \subseteq W$ are subrepresentations. But both (V, ρ) and (W, π) are irreducible, so they only have two subrepresentations: either $\{0\}$ or the entire space.

If $\operatorname{Ker} T = V$, then $\operatorname{Im} T = \{0\}$, and $T = 0$.

If $\operatorname{Ker} T = \{0\}$, then T is injective, and moreover $\operatorname{Im} T \neq \{0\}$, so $\operatorname{Im} T = W$, so T is also surjective. This means that T is invertible.

If $(V, \rho) = (W, \pi)$, then T must have an eigenvalue, $\lambda \in \mathbb{C}$. So $T - \lambda \mathrm{Id}$ is also a transformation, but this is not injective, as it is zero on the λ eigenspace of T. By the first part of the lemma, $T - \lambda \mathrm{Id} = 0$, so $T = \lambda \mathrm{Id}$. □

In this proof, we used that any linear transformation T from a non-trivial space to itself (i.e. any square matrix) has an eigenvalue. This is because we work over the complex numbers: by the fundamental theorem of algebra, the characteristic polynomial of T must have a root. This does not work over the real numbers – there are many real matrices without real eigenvalues. This is one of the main reasons we prefer working over the complex numbers in this part of the course.

One important direct result of Schur's lemma is that we have a very good handle on irreducible representations of abelian groups. For other groups, we will need some more tools, but this is already a good start.

Corollary 2.4.3. *All irreducible representations of finite abelian groups are one-dimensional.*

Proof. Let G be a finite group and (V, ρ) an irreducible representation of G. Because G is abelian, for any $g, h \in G$, $\rho(g)\rho(h) = \rho(h)\rho(g)$. In other words, $\rho(g)$ is a transformation from (V, ρ) to itself for any $g \in G$. By Schur's lemma 2.4.2, $\rho(g) = \lambda_g \mathrm{Id}$ for some λ_g. So all $\rho(g)$ are diagonal. This can only be irreducible if $\dim V = 1$. □

2.5 — Characters

Characters are certain functions attached to representations. While a representation of a large dimension n becomes quite unwieldy, as it requires an $n \times n$ matrix for each element of the group, a character is just a function $\chi\colon G \to \mathbb{C}$. Still, they somehow encode all of the information of the original representation. Moreover, they are quite useful in physical applications as well.

Characters are defined in terms of an important function in linear algebra called the trace. Because you may not have seen the trace before, this section starts with its definition and basic properties.

2.5.1 — Review: trace

The trace is an invariant of matrices very similar to the determinant, but it is often not addressed in a first course on linear algebra, so I will give the definition and some properties here.

Both the determinant and the trace are the same for similar matrices, and therefore can be expressed in terms of (generalised) eigenvalues. While the determinant is the product of all eigenvalues, the trace is the sum. Similarly, while the determinant has the multiplicative property $\det(AB) = \det(A)\det(B)$, the trace is a linear map, so in particular $\operatorname{Tr}(A+B) = \operatorname{Tr} A + \operatorname{Tr} B$.

Definition 2.5.1. Let $A = (A_{ij})$ be an $n \times n$ matrix. Its *trace* is the sum of its diagonal entries, i.e. $\operatorname{Tr} A = \sum_{i=1}^{n} A_{ii}$.

Example 2.5.2. A particular case, which is surprisingly useful to know, is $\operatorname{Tr}(\mathrm{Id}_{n\times n}) = \sum_{i=1}^{n} 1 = n$. ◇

Lemma 2.5.3. *The trace is a linear map* $\operatorname{Tr}\colon \operatorname{Mat}_{n\times n}(\mathbb{C}) \to \mathbb{C}$. *In other words, for* $A, B \in \operatorname{Mat}_{n\times n}(\mathbb{C})$ *and* $\lambda \in \mathbb{C}$, $\operatorname{Tr}(\lambda A + B) = \lambda \operatorname{Tr} A + \operatorname{Tr} B$.

Proof. This follows from $(\lambda A + B)_{ij} = \lambda A_{ij} + B_{ij}$. □

Lemma 2.5.4. *Let A be an $n \times m$ matrix and B an $m \times n$ matrix. Then* $\operatorname{Tr}(AB) = \operatorname{Tr}(BA)$.

This is often called the *cyclic property* of the trace.

Proof. This is an explicit calculation, using only the definition of the product of matrices and that of the trace:

$$\operatorname{Tr}(AB) = \sum_{i=1}^{n} (AB)_{ii} = \sum_{i=1}^{n}\sum_{j=1}^{m} A_{ij}B_{ji} = \sum_{j=1}^{m}\sum_{i=1}^{n} B_{ji}A_{ij} = \sum_{j=1}^{m} (BA)_{jj} = \operatorname{Tr}(BA)\,.$$ □

Corollary 2.5.5. *Similar matrices have equal traces. The trace of a matrix is the sum of its generalised eigenvalues, counted with multiplicities.*

Proof. Two matrices A and A' are similar if there exists an invertible B such that $A = BA'B^{-1}$. Then $\operatorname{Tr}(A) = \operatorname{Tr}(BA'B^{-1}) = \operatorname{Tr}(B^{-1}BA') = \operatorname{Tr} A'$.

In particular, any matrix is similar to its Jordan normal form, which has the generalised eigenvalues on the diagonal, with multiplicites. □

Definition 2.5.6. Let V be a finite-dimensional vector space and $T\colon V \to V$ a linear map. Then the *trace* of T is the trace of its matrix in any basis. Equivalently, it is the sum of its generalised eigenvalues with multiplicities.

By the corollary, the definition of the trace of T does not depend on the choice of basis.

2.5.2 — Characters

Definition 2.5.7. Let G be a group and (V,ρ) a representation of G. Its *character* is the function $\chi_\rho\colon G\to\mathbb{C}$ given by

$$\chi_\rho(g) = \mathrm{Tr}\left(\rho(g)\right) = \sum_{i=1}^{\dim V} (\rho(g))_{ii}\,. \tag{2.7}$$

In case (V,ρ) is irreducible, we call its character χ_ρ *simple*.

This is just a function, so it does not have to be a homomorphism.

Lemma 2.5.8. *For a one-dimensional representation (V,ρ), its character is equal to the representation itself, i.e. $\chi_\rho = \rho$.*

Example 2.5.9. By the above point, one-dimensional characters do not give us too much information we did not already have. So let us look at a higher-dimensional representation. We take the permutation representation of S_n of example 2.1.7. As permutation matrices only have zeroes and ones as entries, the trace counts the number of ones on the diagonal, i.e. the number of basis vectors that get mapped to themselves. This is nothing but the number of one-cycles of the permutation. If we look at the particular case of $n=3$, we find

$$\chi_\pi(e) = 3\,, \qquad \chi_\pi((1\,2)) = \chi_\pi((1\,3)) = \chi_\pi((2\,3)) = 1\,, \qquad \chi_\pi((1\,2\,3)) = \chi_\pi((1\,3\,2)) = 0\,. \tag{2.8}$$

◇

Example 2.5.10. Recall also the regular representation of any finite group G, given in example 2.1.8. This is a $|G|$-dimensional representation r_G by permutation matrices, and by the rearrangement theorem 1.1.19, we know that if $g\neq e$, then $r_G(g)$ does not fix any basis vector. Therefore, $\chi_{r_G}(g) = |G|\delta_{g,e}$. ◇

Lemma 2.5.11. *The character χ_ρ of any representation (V,ρ) of any group G is a* class function*, i.e. it is constant on conjugacy classes. In other words, for any $g,h\in G$, $\chi_\rho(hgh^{-1}) = \chi_\rho(g)$.*

Proof. This follows from corollary 2.5.5:

$$\chi_\rho(hgh^{-1}) = \mathrm{Tr}\left(\rho(hgh^{-1})\right) = \mathrm{Tr}\left(\rho(h)\rho(g)\rho(h)^{-1}\right) = \mathrm{Tr}\left(\rho(g)\right) = \chi_\rho(g)\,.$$

□

Example 2.5.12. Consider the defining representation of the dihedral group D_n given in example 2.1.9. We find that

$$\chi_\delta(\rho^k) = 2\cos(2\pi k/n)\,, \qquad \chi_\delta(\rho^k\sigma) = 0\,. \tag{2.9}$$

◇

Lemma 2.5.13. *Equivalent representations have the same character.*

Proof. Let G be a group and (V,ρ) and (W,π) two representations of G, with an equivalence $T\colon (V,\rho)\to(W,\pi)$. Then, using corollary 2.5.5, for any $g\in G$,

$$\chi_\rho(g) = \mathrm{Tr}\left(\rho(g)\right) = \mathrm{Tr}\left(T^{-1}\pi(g)T\right) = \mathrm{Tr}\left(\pi(g)\right) = \chi_\pi(g)\,.$$

□

Corollary 2.5.14. *For any character χ of a finite group G and any $g\in G$, $\chi(g^{-1}) = \overline{\chi(g)}$. In particular, if g and g^{-1} are conjugate in G, then $\chi(g)\in\mathbb{R}$.*

Proof. Pick a representation (V,ρ) for which χ is a character. By the unitarity theorem 2.3.9 and the previous lemma, we may choose (V,ρ) to be unitary. Therefore, $\chi(g^{-1}) = \mathrm{Tr}\left(\rho(g^{-1})\right) = \mathrm{Tr}\left(\rho(g)^\dagger\right) = \overline{\mathrm{Tr}\left(\rho(g)\right)} = \overline{\chi(g)}$. □

LEMMA 2.5.15. *The character is additive with respect to direct sums and multiplicative with respect to tensor products: if G is a group and (V,ρ) and (W,π) two representations, then*

$$\chi_{\rho\oplus\pi} = \chi_\rho + \chi_\pi\,, \qquad \chi_{\rho\otimes\pi} = \chi_\rho \cdot \chi_\pi\,. \tag{2.10}$$

Proof. Let us write $n = \dim V$ and $m = \dim W$. For any $g \in G$, the diagonal elements of $(\rho\oplus\pi)(g)$ are those of $\rho(g)$ joined with those of $\pi(g)$, so

$$\begin{aligned}\chi_{\rho\oplus\pi}(g) = \mathrm{Tr}\,\big((\rho\oplus\pi)(g)\big) &= \sum_{i=1}^{n+m}\big((\rho\oplus\pi)(g)\big)_{ii}\\ &= \sum_{i=1}^{n}\big(\rho(g)\big)_{ii} + \sum_{j=1}^{m}\big(\pi(g)\big)_{jj}\\ &= \mathrm{Tr}\,\big(\rho(g)\big) + \mathrm{Tr}\,\big(\pi(g)\big) = \chi_\rho(g) + \chi_\pi(g)\,.\end{aligned}$$

Furthemore, the diagonal elements of $(\rho\otimes\pi)(g)$ are all products of diagonal elements of $\rho(g)$ and $\pi(g)$, so

$$\begin{aligned}\chi_{\rho\otimes\pi}(g) = \mathrm{Tr}\,\big((\rho\otimes\pi)(g)\big) &= \sum_{i=1}^{n}\sum_{j=1}^{m}\big((\rho\otimes\pi)(g)\big)_{ij,ij}\\ &= \sum_{i=1}^{n}\big(\rho(g)\big)_{ii}\sum_{j=1}^{m}\big(\pi(g)\big)_{jj}\\ &= \mathrm{Tr}\,\big(\rho(g)\big)\,\mathrm{Tr}\,\big(\pi(g)\big) = \chi_\rho(g)\chi_\pi(g)\,.\end{aligned}$$

As the product of functions is defined pointwise, this finishes the proof. □

2.6 — ORTHOGONALITY

For this entire part of the course, we have been using linear algebra to talk about groups. In this section, we will do this as well, but in a different way.

In the previous section, we defined characters as certain functions $G \to \mathbb{C}$. So, in effect, they are just a collection of numbers, one for each $g \in G$. But a collection of numbers is nothing but a vector, in this case in a vector space of dimension $|G|$. This observation is more than just some wisecracking, and is one of the key points to really understanding representation theory.

DEFINITION 2.6.1. We define a vector space $\mathbb{C}^G = \{f\colon G \to \mathbb{C}\}$, with addition and scalar multiplication. It has a Hermitian inner product given by

$$\langle f_1, f_2\rangle = \frac{1}{|G|}\sum_{g\in G}\overline{f_1(g)}f_2(g)\,. \tag{2.11}$$

We also define the subspace $\mathrm{Cl}(G) \subseteq \mathbb{C}^G$ of *class functions*, i.e.

$$\mathrm{Cl}(G) = \{f \in \mathbb{C}^G \mid f(g) = f(hgh^{-1}) \text{ for all } g, h \in G\}\,. \tag{2.12}$$

For a class function f, we may write $f(C) = f(g)$ if $C \subseteq G$ is a conjugacy class and $g \in C$.

If we write $\mathcal{C}(G)$ for the set of conjugacy classes of G, then the inner product on $\mathrm{Cl}(G)$ can also be written as

$$\langle f_1, f_2\rangle = \frac{1}{|G|}\sum_{C\in\mathcal{C}(G)}|C|\overline{f_1(C)}f_2(C)\,. \tag{2.13}$$

EXAMPLE 2.6.2. For any representation (V, ρ), we have $\chi_\rho \in \mathrm{Cl}(G)$. But there are more easily-found elements of $\mathbb{C}^G$. In fact, all matrix coefficients of ρ are elements of $\mathbb{C}^G$: we define $\rho_{ij}\colon G \to \mathbb{C}\colon g \mapsto \big(\rho(g)\big)_{ij}$. ◇

The reason this vector space is so interesting is that many of these vectors are orthogonal.

THEOREM 2.6.3 (GREAT ORTHOGONALITY THEOREM). *Let G be a finite group and (V, ρ) and (W, π) two irreducible unitary representations of G. If (V, ρ) and (W, π) are inequivalent, then*

$$\langle \rho_{ij}, \pi_{kl} \rangle = 0 \qquad \textit{for all } 1 \le i, j \le \dim V \textit{ and } 1 \le k, l \le \dim W. \tag{2.14}$$

If $(V, \rho) = (W, \pi)$, then

$$\langle \rho_{ij}, \rho_{kl} \rangle = \frac{1}{\dim V} \delta_{ik} \delta_{jl} \,. \tag{2.15}$$

Proof. The main ingredient in the proof is Schur's lemma 2.4.2, along with a cleverly-picked matrix. We let X be any $\dim V \times \dim W$ matrix, and define

$$A = \sum_{g \in G} \rho(g)^\dagger X \pi(g) = \sum_{g \in G} \rho(g^{-1}) X \pi(g) \,.$$

We calculate

$$\begin{aligned} \rho(h) A &= \rho(h) \sum_{g \in G} \rho(g^{-1}) X \pi(g) \\ &= \sum_{g \in G} \rho(hg^{-1}) X \pi(gh^{-1}) \pi(h) \\ &= \sum_{k \in G} \rho(k^{-1}) X \pi(k) \pi(h) \\ &= A \pi(h) \,, \end{aligned}$$

where we used the rearrangement theorem 1.1.19. By Schur's lemma 2.4.2, if (V, ρ) and (W, π) are inequivalent, $A = 0$. In particular, taking $X = E_{ik}$ to be the matrix with mostly zeroes and a single one in the ith row and kth column, we get the first statement.

If $(V, \rho) = (W, \pi)$, then we find, again by Schur's lemma 2.4.2, that $A = \lambda_X \mathrm{Id}$ for some λ_X. Taking the trace of this, we find that

$$\lambda_X \dim V = \mathrm{Tr}(\lambda_X \mathrm{Id}) = \mathrm{Tr}(A) = \sum_{g \in G} \mathrm{Tr}\left(\rho(g^{-1}) X \rho(g)\right) = \sum_{g \in G} \mathrm{Tr}(X) = |G| \, \mathrm{Tr}\, X \,.$$

Taking X to be the same matrix as before, $\mathrm{Tr}\, E_{ik} = \delta_{ik}$, so $\lambda_{E_{ik}} = \frac{|G|}{\dim V} \delta_{ik}$ and

$$\begin{aligned} \langle \rho_{ij}, \rho_{kl} \rangle &= \frac{1}{|G|} \sum_{g \in G} \overline{\rho_{ij}(g)} \rho_{kl}(g) \\ &= \frac{1}{|G|} \sum_{g \in G} \big(\rho(g)^\dagger\big)_{ji} \big(\rho(g)\big)_{kl} \\ &= \frac{1}{|G|} \sum_{g \in G} \Big(\rho(g)^\dagger E_{ik} \rho(g)\Big)_{jl} \\ &= \frac{1}{|G|} A_{jl} \\ &= \frac{\lambda}{|G|} \delta_{jl} \\ &= \frac{1}{\dim V} \delta_{ik} \delta_{jl} \,. \end{aligned}$$

This is what we wanted to prove. □

COROLLARY 2.6.4. *Let G be a finite group. Its simple characters form an orthonormal set in $\mathrm{Cl}(G)$. In particular, the number of inequivalent irreducible representations of G is finite and bounded by the number of conjugacy classes.*

Proof. Let χ and χ' be two simple characters. They are characters of some irreducible representations (V,ρ) and (W,π), which we may choose to be unitary by the unitarity theorem 2.3.9 and by lemma 2.5.13. By the great orthogonality theorem 2.6.3, if $\chi \neq \chi'$,

$$\langle \chi, \chi' \rangle = \Big\langle \sum_{i=1}^{\dim V} \rho_{ii}, \sum_{j=1}^{\dim W} \pi_{jj} \Big\rangle = \sum_{i=1}^{\dim V} \sum_{j=1}^{\dim W} \langle \rho_{ii}, \pi_{jj} \rangle = 0\,,$$

while on the other hand

$$\langle \chi, \chi \rangle = \Big\langle \sum_{i=1}^{\dim V} \rho_{ii}, \sum_{j=1}^{\dim V} \rho_{jj} \Big\rangle = \sum_{i,j=1}^{\dim V} \langle \rho_{ii}, \rho_{jj} \rangle = \sum_{i,j=1}^{\dim V} \frac{1}{\dim V} \delta_{ij}\delta_{ij} = 1\,.$$

□

In fact, we can get more: the simple characters give a basis for the space of class functions.

THEOREM 2.6.5 (ROW ORTHOGONALITY FOR CHARACTERS). *Let G be a finite group. Its simple characters span $\mathrm{Cl}(G)$, and hence form an orthonormal basis. In particular, the number of inequivalent irreducible representations of G is equal to the number of its conjugacy classes.*

Proof. Given a class function $\alpha \in \mathrm{Cl}(G)$ and a representation (V,ρ), we will construct a transformation $T^\alpha_\rho \colon (V,\rho) \mapsto (V,\rho)$. We will show that, if α is orthogonal to all simple characters, T^α_ρ must be zero for any V. By considering the regular representation of examples 2.1.8 and 2.5.10, we find that α must be zero, and therefore that the characters span $\mathrm{Cl}(G)$.

Define

$$T^\alpha_\rho = \frac{1}{|G|} \sum_{g\in G} \overline{\alpha(g)} \rho(g)\,.$$

Then, for any $h \in G$,

$$\begin{aligned}
T^\alpha_\rho \rho(h) &= \frac{1}{|G|} \sum_{g\in G} \overline{\alpha(g)} \rho(gh) \\
&= \rho(h) \frac{1}{|G|} \sum_{g\in G} \overline{\alpha(g)} \rho(h^{-1}gh) \\
&= \rho(h) \frac{1}{|G|} \sum_{k\in G} \overline{\alpha(hkh^{-1})} \rho(k) \\
&= \rho(h) \frac{1}{|G|} \sum_{k\in G} \overline{\alpha(k)} \rho(k) \\
&= \rho(h) T^\alpha_\rho\,,
\end{aligned}$$

where we used the rearrangement theorem 1.1.19 for the third equality and the fact that α is a class function for the fourth.

If (V,ρ) is simple, by Schur's lemma 2.4.2, we now know that $T^\alpha_\rho = \lambda^\alpha_\rho \mathrm{Id}$ for some $\lambda^\alpha_\rho \in \mathbb{C}$. Taking the trace of this equality, we find that

$$\dim V \cdot \lambda^\alpha_\rho = \frac{1}{|G|} \sum_{g\in G} \overline{\alpha(g)} \chi_\rho(g) = \langle \alpha, \chi_\rho \rangle\,.$$

Now assume that α is in the orthogonal complement of the span of simple characters. Then $\langle \alpha, \chi_\rho \rangle = 0$ for all irreducible ρ, so $T^\alpha_\rho = 0$. But as all representations are semisimple, this must then hold for all representations ρ.

In particular, if we take the regular representation, and look at the basis vector associated to the unit of G (confusingly written as e_e), then we find that

$$0 = T^\alpha_{r_G}(e_e) = \frac{1}{|G|} \sum_{g \in G} \overline{\alpha(g)} r_G(g) e_e = \frac{1}{|G|} \sum_{g \in G} \overline{\alpha(g)} e_g \,,$$

from which we can conclude that $\alpha(g) = 0$ for all $g \in G$. □

There is a second orthogonality result, which in a sense goes into the other direction (I guess it is orthogonal to this one, pun fully intended).

Corollary 2.6.6 (Column orthogonality for characters). *Let G be a finite group, and $\{\chi_1, \ldots, \chi_k\}$ its irreducible characters (of which we now know there are finitely many). Let $C, C' \subset G$ be two conjugacy classes. Then*

$$\sum_{i=1}^{k} \overline{\chi_i(C)} \chi_i(C') = \frac{|G|}{|C|} \delta_{C,C'} \,. \tag{2.16}$$

Proof. Consider the matrix A whose rows are indexed by irreducible representations ρ of G and whose columns are indexed by conjugacy classes C of G, with coefficients $A_{\rho,C} = \sqrt{|C|} \chi_\rho(C)$. By theorem 2.6.5, this is a square matrix whose rows are orthonormal for the inner product equation (2.13). Therefore, its columns must also be orthonormal, i.e.

$$\frac{1}{|G|} \sum_{i=1}^{k} \sqrt{|C|}\, \overline{\chi_i(C)} \sqrt{|C'|} \chi_i(C') = \delta_{C,C'} \,.$$

Bringing factors to the other side gives the result. □

2.6.1 — Character tables

A slightly different variant of the matrix A in the proof of corollary 2.6.6 is a very important tool for representation theory, and one of the most concrete ways to study the representations of a group:

Definition 2.6.7. Let G be a finite group. Its *character table* is the table (or matrix) whose rows are indexed by irreducible representations, and whose columns are indexed by conjugacy classes, such that the (ρ, C) coefficient is $\chi_\rho(C)$.

Example 2.6.8. For the cyclic groups $\mathbb{Z}/n\mathbb{Z}$, we saw a lot of one-dimensional representations in example 2.1.10. With all our tools at hand, we now see these are all irreducible representations, as we had n of them, exactly as many as the number of conjugacy classes/elements of $\mathbb{Z}/n\mathbb{Z}$. The character table is $X_{\rho_j, \bar{k}} = e^{2\pi i jk/n}$. ◇

Example 2.6.9. For symmetric groups S_n with $n > 1$, we have already found two irreducible representations: the trivial representation τ and the sign representation ε, cf. examples 2.1.6 and 2.1.11. Both are one-dimensional. We also know that conjugacy classes of the symmetric group are given by cycle types, by theorem 1.4.19.

In case $n = 3$, we know that there are three cycle types, given by the partitions $(1,1,1)$, $(2,1)$, and (3) (take a look at the exercises for this). Therefore, there must also be a third simple character χ. So the character table of S_3 looks like

S_3	$(1,1,1)$	$(2,1)$	(3)
τ	1	1	1
ε	1	-1	1
χ			

This table can be completed by column orthogonality, and because we know $\chi(e) = \dim \chi > 0$, to give

S_3	$(1,1,1)$	$(2,1)$	(3)
τ	1	1	1
ε	1	-1	1
χ	2	0	-1

So we know that $\chi = \chi_\rho$ for some 2-dimensional representation (U, ρ). Also, comparing to the permutation character χ_π from example 2.5.9, we see that $\chi_\pi = \tau + \chi_\rho$. This tells us that the permutation representation contains two irreducible representations: τ and ρ. You can check that the subspace spanned by $e_1 + e_2 + e_3$ is indeed equivalent to the trivial representation, and its orthogonal complement $U = \{v_1 e_1 + v_2 e_2 + v_3 e_3 \mid v_1 + v_2 + v_3 = 0\}$ is also a subrepresentation, as π is unitary, by the proof of corollary 2.3.12. So $(U, \pi|_U) \cong \rho$. ◇

Example 2.6.10. Consider the dihedral group D_8. We have the defining character from example 2.1.9 of dimension 2. Because of column orthogonality, corollary 2.6.6 for the class of the identity, all other irreducible representations must be one-dimensional (as one must be trivial), and there must be four in total. This means that for such a representation π, $\pi(\sigma\rho^3) = \pi(\rho\sigma) = \pi(\rho)\pi(\sigma) = \pi(\sigma)\pi(\rho)$, so $\pi(\rho)^2 = \pi(\sigma)^2 = 1$, and we can get all four options. Let $\pi_{\pm,\pm}$ denote these possibilities. The character table looks like

D_8	e	ρ^2	$\{\rho, \rho^3\}$	$\{\sigma, \rho^2\sigma\}$	$\{\rho\sigma, \rho^3\sigma\}$
$\pi_{+,+}$	1	1	1	1	1
$\pi_{+,-}$	1	1	1	-1	-1
$\pi_{-,+}$	1	1	-1	1	-1
$\pi_{-,-}$	1	1	-1	-1	1
δ	2	-2	0	0	0

Note that this is the same character table as we would get for the quaternion group Q, of which we found all irreducible representations in an exercise (we did not show we found them all, but by now that is easy to check). So, while characters are very useful, they do not retain all information of groups. (However, in section 2.7, we will find a way of distinguishing between these two tables, using a bit more of the group structure.) ◇

2.6.2 — Decomposition

Now we have a vector space of class functions with an orthonormal basis given by the simple characters, we can use this to decompose characters. By corollary 2.3.12, any character is a sum of simple characters, i.e. a linear combination with non-negative integer coefficients. Because we are working in an inner product space, we can find these coefficients explicitly.

Proposition 2.6.11. *Let G be a finite group with irreducible representations $\{(V_i, \rho_i) \mid i = 1, \ldots, k\}$ with characters χ_i. Then any representation (W, π) can be decomposed as*

$$(W, \pi) \cong \bigoplus_{i=1}^{k} m_i (V_i, \rho_i)\,, \tag{2.17}$$

where

$$m_i = \langle \chi_\pi, \chi_i \rangle = \frac{1}{|G|} \sum_{C \in \mathcal{C}(G)} |C| \overline{\chi_\pi(C)} \chi_i(C)\,. \tag{2.18}$$

Proof. This follows directly from row orthogonality, corollary 2.6.6, and additivity of characters, lemma 2.5.15. □

COROLLARY 2.6.12. *Let G be a finite group. Its regular representation contains all of its irreducible representations, with multiplicity equal to their dimension. That is,*

$$(\mathbb{C}^G, r_G) \cong \bigoplus_{i=1}^{k} \dim(V_i)(V_i, \rho_i)\,. \tag{2.19}$$

In particular, taking dimensions,

$$|G| = \sum_{i=1}^{k} (\dim V_i)^2\,.$$

Proof. By proposition 2.6.11, we have to calculate inner products $\langle \chi_{r_G}, \chi_i \rangle$. But $\chi_{r_G}(g) = |G|\delta_{g,e}$ by example 2.5.10, so

$$m_i = \langle \chi_{r_G}, \chi_i \rangle = \frac{1}{|G|} \sum_{g\in G} |G|\delta_{e,g}\chi_i(g) = \chi_i(e) = \dim V_i\,.$$ □

Moreover, because these coefficients are non-negative integers, we can get a criterion for simplicity out of this.

PROPOSITION 2.6.13. *Let G be a finite group and (V,ρ) representation of G. Then (V,ρ) is irreducible if and only if*

$$\langle \chi_\rho, \chi_\rho \rangle = 1\,. \tag{2.20}$$

If (V,ρ) is reducible, then

$$\langle \chi_\rho, \chi_\rho \rangle > 1\,. \tag{2.21}$$

Proof. By proposition 2.6.11, $(V,\rho) \cong \bigoplus_{i=1}^{k} m_i(V_i,\rho_i)$. Therefore,

$$\langle \chi_\rho, \chi_\rho \rangle = \sum_{i,j=1}^{k} m_i m_j \langle \chi_i, \chi_j \rangle = \sum_{i=1}^{k} m_i^2\,.$$

Clearly, this is equal to 1 if and only if exactly one of the m_i equals 1 and all other are 0, but this means exactly that $(V,\rho) \cong (V_i,\rho_i)$, so it is irreducible.

In the other case, the sum is strictly larger than 1. □

From the proof, we see we can get some more information. For example, if $\langle \chi_\rho, \chi_\rho \rangle = 2$, then ρ must be the sum of two inequivalent irreducible representations.

EXAMPLE 2.6.14. Let us consider S_4. It has five conjugacy classes, given by the cycle types/partitions (1^4), $(2,1^2)$, (2^2), $(3,1)$, and (4). These have sizes

$$|C_{(1^4)}| = 1\,, \qquad |C_{(2,1^2)}| = 6\,, \qquad |C_{(2^2)}| = 3\,, \qquad |C_{(3,1)}| = 8\,, \qquad |C_{(4)}| = 6\,.$$

We know of two one-dimensional representations, which are the trivial representation τ and the sign representation ε.

We also have the four-dimensional permutation representation π, but this is reducible: it has a subrepresentation spanned by $e_1 + e_2 + e_3 + e_4$, which is equivalent to the trivial representation. This subrepresentation has a complement by corollary 2.3.12, which is the subspace $S = \{c_1e_1 + c_2e_2 + c_3e_3 + c_4e_4 \mid c_1 + c_2 + c_3 + c_4 = 0\}$. This is called the *standard representation* (S, s), and it is three-dimensional. If we denote its character by χ_s, then $\chi_s + \tau = \chi_\pi$, so we can calculate

$$\langle \chi_s, \chi_s \rangle = \langle \chi_\pi - \tau, \chi_\pi - \tau \rangle = \frac{1}{|S_4|} \sum_{C\in\mathcal{C}(S_4)} |C|\big(\chi_\pi(C) - \tau(C)\big)^2\,.$$

We know that $\chi_\pi(C)$ equals the number of one-cycles of an element of C by example 2.5.9, and $\tau(C) = 1$ for any C, so

$$\begin{aligned}\langle \chi_s, \chi_s \rangle &= \frac{1}{24}\Big(1 \cdot (4-1)^2 + 6 \cdot (2-1)^2 + 3 \cdot (0-1)^2 + 8 \cdot (1-1)^2 + 6 \cdot (0-1)^2\Big) \\ &= \frac{1}{24}\Big(9 + 6 + 3 + 0 + 6\Big) = 1\,,\end{aligned}$$

and so by proposition 2.6.13, χ_s is a simple character.

Then $\varepsilon \cdot \chi_s$ must also be simple, as it is the character of $\varepsilon \otimes s$, and any subrepresentation of $\varepsilon \otimes s$ yields a subrepresentation of s by tensoring with ε again. Moreover, it is different from any of the other simple characters we have found. So we get an incomplete character table

S_4	(1^4)	$(2,1^2)$	(2^2)	$(3,1)$	(4)
τ	1	1	1	1	1
ε	1	-1	1	1	-1
χ_ρ					
χ_s	3	1	-1	0	-1
$\varepsilon\chi_s$	3	-1	-1	0	1

The one leftover row can be completed by orthogonality, and this gives

S_4	(1^4)	$(2,1^2)$	(2^2)	$(3,1)$	(4)
τ	1	1	1	1	1
ε	1	-1	1	1	-1
χ_ρ	2	0	2	-1	0
χ_s	3	1	-1	0	-1
$\varepsilon\chi_s$	3	-1	-1	0	1

Note that χ_ρ is the character given by composing the quotient homomorphism $q\colon S_4 \to S_4/V_4 \cong S_3$ from example 1.5.19 with the two-dimensional simple character χ of S_3 found in example 2.6.9, so we can realise ρ as $\pi|_U \circ q$. More precisely, ρ is only defined up to isomorphism, as we only know its character, and $\pi|_U \circ q$ is a representative of this isomorphism class. ◇

2.7 — Real, complex, quaternionic representations

For this entire part of the course, we have looked at representations over the complex numbers. But you may wonder (as some students asked this during lectures) what happens if we look at other fields, e.g. the real numbers. The general answer is quite complicated, but for the reals, we can actually answer it completely.

Moreover, this is a useful question to ask: any representation has a complex conjugate, and in terms of elementary particles, an antiparticle transforms as the conjugate representation. So real representations correspond to particles which are their own antiparticle.

First of all, we should consider carefully what a real representation should be. For a single number or matrix, we say it is real if its imaginary part vanishes, or better if it equals its complex conjugate. For representations we want this as well. However, now we come to the question: if a representation is given, is it equivalent to a real representation? Or even: given a character, is it the character of a real representation? These are the questions we will answer here.

DEFINITION 2.7.1. Let G be a group and (V, ρ) a representation (on a complex vector space). Its *complex conjugate representation* is the representation $(\bar{V}, \bar{\rho})$ given by $\bar{\rho}(g) = \overline{\rho(g)}$. As a consequence, $\chi_{\bar{\rho}}(g) = \overline{\chi_\rho(g)}$.

A representation (V, ρ) of a group G is called *complex* if it is not equivalent to its complex conjugate representation.

By the above discussion, you may hope that any representation which is not complex is real (i.e. equivalent to a representation with real entries). This is not quite true, although the converse clearly holds. Let us study the non-complex representations further.

LEMMA 2.7.2. *Let G be a finite group and (V, ρ) a unitary irreducible finite-dimensional representation which is equivalent to its complex conjugate by some matrix S:*

$$\bar{\rho}(g) = S\rho(g)S^{-1} . \tag{2.22}$$

Then S is either symmetric or antisymmetric.

Proof. If we take the transpose of the relation above, and using unitarity, we get

$$\rho(g^{-1}) = \rho(g)^\dagger = (S^{-1})^T \rho(g)^T S^T .$$

This holds for any $g \in G$, and therefore also for g^{-1}:

$$\rho(g) = (S^{-1})^T \rho(g^{-1})^T S^T .$$

Substituting $\rho(g^{-1})^T = S\rho(g)S^{-1}$ into this yields

$$\rho(g) = (S^{-1})^T S\rho(g)S^{-1}S^T = (S^{-1}S^T)^{-1}\rho(g)(S^{-1}S^T) .$$

Therefore, $S^{-1}S^T$ is a transformation from (V, ρ) to itself, and hence by Schur's lemma 2.4.2, it is equal to $\lambda\mathrm{Id}$ for some $\lambda \in \mathbb{C}$. In other words, $S = \lambda S^T$. But then

$$S = \lambda S^T = \lambda(\lambda S^T)^T = \lambda^2 S ,$$

and as $S \neq 0$, we find $\lambda = \pm 1$. This means exactly that S is symmetric ($\lambda = 1$) or anti-symmetric ($\lambda = -1$). □

COROLLARY 2.7.3. *Let G be a finite group and (V, ρ) an irreducible representation which is equivalent to its complex conjugate by some matrix S:*

$$\bar{\rho}(g) = S\rho(g)S^{-1} . \tag{2.23}$$

Then S is either symmetric or antisymmetric.

Proof. The representation (V, ρ) is equivalent to a unitary one by theorem 2.3.9. Let the equivalence be E. Then by the theorem, $E^{-1}SE$ is symmetric or anti-symmetric, but then so is S itself. □

We want to distinguish between the two cases in the lemma. But how do we know we can? Maybe there could be both a symmetric and an anti-symmetric matrix giving the required equivalence. And in any case, this is quite hard to check. It may also be hard to check if a representation is complex. As usual, to resolve the questions, we turn to characters.

DEFINITION 2.7.4. Let G be a finite group and χ a character of G. The *Frobenius-Schur indicator* of χ is the number

$$\mathrm{FS}(\chi) = \frac{1}{|G|}\sum_{g\in G}\chi(g^2) . \tag{2.24}$$

THEOREM 2.7.5. *Let G be a finite group and (V, ρ) an irreducible representation of G. Then*

$$\mathrm{FS}(\chi_\rho) = \begin{cases} 0 & (V, \rho) \text{ is complex} \\ 1 & (V, \rho) \cong (\bar{V}, \bar{\rho}) \text{ by a symmetric matrix} \\ -1 & (V, \rho) \cong (\bar{V}, \bar{\rho}) \text{ by an anti-symmetric matrix} \end{cases} \tag{2.25}$$

Proof. This is just a sketch to give you an idea; this is not part of the course material.

The vector space $V^\vee \otimes V^\vee$ (where $V^\vee$ is the dual space to V; a unitary inner product induces an isomorphism $V^\vee \cong \bar{V}$) is the space of bilinear forms on V. It has a representation $\rho^\vee \otimes \rho^\vee$, where $\rho^\vee = (\rho^{-1})^T$, and this decomposes into the subspace of symmetric forms $(\mathrm{Sym}^2 V^\vee, \mathrm{Sym}^2 \rho^\vee)$ and the subspace of anti-symmetric forms $(\bigwedge^2 V^\vee, \bigwedge^2 \rho^\vee)$. Then $\chi_{\rho^\vee}(g^2) = \chi_{\mathrm{Sym}^2 \rho^\vee}(g) - \chi_{\bigwedge^2 \rho^\vee}(g)$. By lemma 2.7.2 and Schur's lemma 2.4.2, the trivial representation τ occurs at most once in $(V^\vee \otimes V^\vee, \rho^\vee \otimes \rho^\vee)$ (i.e. there is up to scaling at most one equivalence $S\colon (V, \rho) \to (V^\vee, \rho^\vee)$), and this will always be a subrepresentation of either $(\mathrm{Sym}^2 V^\vee, \mathrm{Sym}^2 \rho^\vee)$ or $(\bigwedge^2 V^\vee, \bigwedge^2 \rho^\vee)$. Then

$$\begin{aligned} \mathrm{FS}(\chi) &= \frac{1}{|G|} \sum_{g \in G} \overline{\chi_{\rho^\vee}(g^2)} \\ &= \frac{1}{|G|} \sum_{g \in G} \overline{\chi_{\mathrm{Sym}^2 \rho^\vee}(g) - \chi_{\bigwedge^2 \rho^\vee}(g)} \\ &= \left\langle \mathrm{Sym}^2 \rho^\vee - \textstyle\bigwedge^2 \rho^\vee, \tau \right\rangle , \end{aligned}$$

and by the previous discussion and proposition 2.6.11, this evaluates to the value in the theorem. □

DEFINITION 2.7.6. Let G be a finite group and (V, ρ) an irreducible representation. We call (V, ρ) *real* if $\mathrm{FS}(\chi_\rho) = 1$ and *quaternionic* if $\mathrm{FS}(\chi_\rho) = -1$.

LEMMA 2.7.7. *An irreducible representation (V, ρ) of a finite group is real if and only if it is equivalent to a representation by real matrices.*

Proof. First, we take (V, ρ) unitary. We want to do the same for S. We see

$$\begin{aligned} S^\dagger S \rho(g) &= S^\dagger \bar{\rho}(g) S \\ &= (\bar{\rho}(g)^\dagger S)^\dagger S \\ &= (\bar{\rho}(g^{-1}) S)^\dagger S \\ &= (S \rho(g^{-1}))^\dagger S \\ &= \rho(g^{-1})^\dagger S^\dagger S \\ &= \rho(g) S^\dagger S \end{aligned}$$

By Schur's lemma 2.4.2, we then know that $S^\dagger S = \lambda \mathrm{Id}$ for some $\lambda \in \mathbb{C}$.

But S is defined only up to rescaling, so we can choose an appropriate rescaling so that $S^\dagger S = \mathrm{Id}$.

Now that we have S unitary, we can prove the lemma. We assume that the representation is real, so that S is a symmetric unitary matrix. So it is diagonalisable with eigenvalues in $U(1)$. Using this, we find $S = W^2$ for some unitary symmetric matrix W. Thus $\bar{\rho}(g) = S\rho(g)S^{-1}$ becomes $\bar{\rho}(g) = W^2 \rho(g) W^{-2}$. We conjugate by W to get

$$W^{-1} \bar{\rho}(g) W = W \rho(g) W^{-1} .$$

Using the fact that W is unitary symmetric, we have $W^{-1} = W^\dagger = \bar{W}$, and $W = (\bar{W}^{-1})$. So we can write:

$$W \rho(g) W^{-1} = W^{-1} \bar{\rho}(g) W = \bar{W} \bar{\rho}(g) \bar{W}^{-1} = \overline{W \rho(g) W^{-1}}.$$

It thus follows that $W\rho(g)W^{-1}$ has only real entries, so (V, ρ) is equivalent to a representation with real entries. □

Example 2.7.8. The trivial representation is always real, as it is given by real matrices ($1 \in \mathbb{R}$). ◇

Example 2.7.9. All irreducible representations of symmetric groups are real. This follows from more advanced techniques than we have seen, but you can check this for a few cases by calculating Frobenius-Schur indicators. ◇

Example 2.7.10. Many of the irreducible representations ρ_j of cyclic groups $\mathbb{Z}/n\mathbb{Z}$ we have seen in example 2.1.10 are complex: they are real if $j \in \{0, n/2\}$, because then $\operatorname{Im}\rho_j \subseteq \{\pm 1\}$, but otherwise they have at least one strictly complex value. By example 2.2.2, two representations on $\mathbb{C}$ are equivalent if and only if they are equal, so if ρ_j has a strictly complex value, it will not be equivalent to its complex conjugate.

Alternatively, we can use that for any $k > 1$ and kth root of unity ϑ,

$$\sum_{i=0}^{k-1} \vartheta^i = 0\,, \tag{2.26}$$

to show that $\mathrm{FS}(\rho_j) = 0$. ◇

Example 2.7.11. The two-dimensional irreducible representation of Q found in one of the exercises, which had matrices

$$\rho(i) = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \qquad \rho(j) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \qquad \rho(k) = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \tag{2.27}$$

is quaternionic:

$$\mathrm{FS}(\rho) = \frac{1}{|Q|}\sum_{g\in Q}\chi_\rho(g^2) = \frac{1}{8}(2\chi_\rho(1) + 6\chi_\rho(-1)) = \frac{1}{8}(2\cdot 2 + 6\cdot -2) = -1\,. \tag{2.28}$$

◇

Remark 2.7.12. The names real, complex, and quaternionic come from the analogon of Schur's lemma 2.4.2 over the real numbers. Over the complex numbers it states that the set (in fact the ring, if you know that term) of transformations from an irreducible representation to itself is isomorphic to $\mathbb{C}$. Over a field k which is not algebraically closed (such as $\mathbb{R}$), this set is isomorphic to some finite rank division algebra over k. In case of $k = \mathbb{R}$, by a result called the Frobenius theorem, there are three such division algebras: $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{H}$. These correspond exactly to the names we have given.

2.8 — Application: crystals

Let us look at an application of representation theory to actual physics, in particular to crystals. Crystals are solids that have a very geometric shape with large, flat faces. This is because of their microscopic structure: the atoms – or molecules, or ions – are arranged in a lattice.

Definition 2.8.1. A *(Bravais) lattice* Λ is an additive discrete subgroup of a real vector space V which spans that space.

In other words, a lattice in V is the $\mathbb{Z}$-linear span of a basis $\{e_1, \dots, e_{\dim V}\}$ of V:

$$\Lambda = \Big\{ \sum_{i=1}^{\dim V} \lambda_i e_i \,\Big|\, \lambda_i \in \mathbb{Z} \Big\}\,. \tag{2.29}$$

Of course, for an actual crystal, $\dim V = 3$, although it is possible to create two-dimensional or one-dimensional crystals as well. Higher dimensions are hard to make in real life, but the theory still exists.

Lattices have two kinds of symmetries: *translations* $t_v \colon w \mapsto v + w$ by any vector v in the lattice, and *point symmetries*, i.e. rotations and reflections which fix a point (we choose this point to be the origin $0 \in V$). In our group theoretic language, this means that there is an additive subgroup of translations $\Lambda \subset V$ and a (multiplicative) subgroup of rotations and reflections $R \subseteq \mathrm{O}(V)$, which together generate some group of symmetries S which acts on V. This is not a representation, as translations are not linear, but the action of R is a representation.

As we understand the translational symmetry part pretty well – it is just the lattice itself – we can focus a bit on the subgroup R. There are a lot of possible subgroups of $\mathrm{O}(V)$, both finite (e.g. dihedral groups) and infinite (e.g. a circle group), and it would be interesting to know which of these groups can actually occur as symmetries of a lattice. That would tell us something about the possible lattices, and hence about the possible crystals we may find in nature. The crystallographic restriction theorem tells us that this group is actually very restricted (hence the name).

Theorem 2.8.2 (Crystallographic restriction theorem). *Let V be a real vector space of dimension 2 or 3, and $\Lambda \subset V$ a lattice. If $r \in R$ is a rotational symmetry of Λ with angle α, then $\alpha = 2\pi/n$ for $n \in \{1, 2, 3, 4, 6\}$.*

Proof. In case $\dim V = 3$, we are looking at a three-dimensional representation (V, ρ) of R. Any rotation in three dimensions is rotation around an axis. In an appropriate orthonormal basis,

$$\rho(r) = \begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

so in particular $\chi_\rho(r) = 2\cos\alpha + 1$.

However, we also have a basis of lattice vectors, and as the lattice by definition is the $\mathbb{Z}$-linear span, and $\rho(r)$ must send lattice vectors to lattice vectors, the matrix $\rho(r)$ with respect to this basis has integer entries. In particular, its trace is integral.

But the trace of a linear transformation is independent of basis, so comparing, $2\cos\alpha \in \mathbb{Z}$. There are five options, $2\cos\alpha \in \{2, 1, 0, -1, -2\}$, and these correspond to $\alpha \in \{2\pi, 2\pi/6, 2\pi/4, 2\pi/3, 2\pi/2\}$.

In case $\dim V = 2$, the argument is similar, but in an orthonormal basis,

$$\rho(r) = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix},$$

so $\chi_\rho(r) = 2\cos\alpha$. □

The complete classification of these groups is known, and quite some information can be found on e.g. Wikipedia: in 2d, the groups are called wallpaper groups. In 3d, they are space groups.

This is not the entire story, though. Quasicrystals may have other symmetries, such as the five-fold symmetry of Penrose tilings. These may occur as projections of higher-dimensional lattices to a lower-dimensional subspace by the cut-and-project method, using that rotations in higher order have more options (although still finitely many for a given dimension).

Chapter 3 — Lie groups and Lie algebras

Up to now, we have mostly been concerned with finite (or discrete) groups, although we have also seen examples of what you may call continuous groups. The latter are the focus of this last part of the course. We will partially repeat the story before, considering representations, in this case.

This part of the course will be less thorough than the previous parts, and is mostly intended to give a basic overview of Lie theory for physicists.

3.1 — Lie groups

There are several different ways of making sense of the notion of 'continuous' groups. The main idea is that a continuous group is modeled on the continuum, i.e. the real line, or higher-dimensional variants. It turns out that we get a very nice theory if we import differentiability from $\mathbb{R}$. In fact, we will want to be able to differentiate infinitely many times, i.e. have smooth, or C^∞, functions. We will mostly encounter polynomial functions, which are always smooth – you know quite well how to differentiate them.

The formal definition is as follows.

Definition 3.1.1. A *Lie group* is a real smooth manifold G with a smooth multiplication map $m\colon G\times G\to G$, a unit $e\in G$, and a smooth inverse map $i\colon G\to G$ satisfying the three axioms of a group, G1–G3 of definition 1.1.1.

The *dimension* of a Lie group is its dimension as a manifold.

A Lie group is *compact* if its underlying topological space is.

However, in this course we do not assume you know what a manifold is, so this definition may not make sense to you. The idea is that a smooth manifold is a 'space' (a topological one, if you know what that is), which has smooth coordinates near any point.

Let us look at a few examples.

Example 3.1.2. Any (countable) group is a zero-dimensional Lie group with the discrete topology. It is compact if and only if it is finite. ◇

Example 3.1.3. The real line $\mathbb{R}$ is a Lie group of dimension one: addition $\mathbb{R}^2\to\mathbb{R}\colon (x,y)\mapsto x+y$ is a smooth map, and inversion $\mathbb{R}\to\mathbb{R}\colon x\mapsto -x$ is as well.

Similarly, any finite-dimensional real vector space is a Lie group under addition, with the same dimension.

These examples are not compact. ◇

Example 3.1.4. The circle group $\mathrm{U}(1)$ is a compact one-dimensional Lie group. It has a (local) coordinate given by an angle ϑ, and in such a coordinate, the operation becomes addition modulo 1, as in example 1.5.18. So this is again smooth. ◇

Example 3.1.5. The multiplicative group $\mathbb{R}^\times$ is a non-compact one-dimensional Lie group as well. The multiplication $\mathbb{R}^\times\times\mathbb{R}^\times\to\mathbb{R}^\times\colon (x,y)\mapsto xy$ is smooth and the inverse $\mathbb{R}^\times\to\mathbb{R}^\times\colon x\to\frac{1}{x}$ is as well (because we exclude $x=0$). ◇

3.1.1 — Matrix Lie groups

For this course, and in fact for most applications in physics, we can take a far more concrete approach to Lie groups: not as a smooth manifold with smooth maps, but as groups of matrices. Let us start with the most important examples.

Example 3.1.6. The general linear groups $\mathrm{GL}(n,\mathbb{R})$ are smooth Lie groups. To see that they are smooth manifolds, we first see that $\mathrm{Mat}_{n\times n}(\mathbb{R})$, the space of n by n real matrices, is a real vector space of dimension n^2, so a smooth manifold. Then $\mathrm{GL}(n,\mathbb{R})$ is an open subset given by $\det \neq 0$, and therefore also a smooth manifold of dimension n^2. The multiplication is smooth, as we can check in the standard coordinates (which are just the matrix coefficients). To check that the inverse is also smooth, we need the determinant $\det\colon \mathrm{GL}(n,\mathbb{R}) \to \mathbb{R}^\times$ to be smooth. This is true, because the determinant is polynomial in all matrix coefficients, and polynomial functions are smooth.

These Lie groups are not compact. ◇

Example 3.1.7. Similarly, the complex general linear groups $\mathrm{GL}(n,\mathbb{C})$ are smooth Lie groups, but now of dimension $2n^2$: we need two real coordinates for each complex matrix entry. ◇

If we want to consider other matrix groups, the following theorem, which we will not prove, is essential.

Theorem 3.1.8 (Closed subgroup theorem). *Let G be a Lie group, and $H \subseteq G$ a closed subgroup. Then H is also a Lie group.*

The condition that H be closed is essential here. For example, $\mathbb{Q} \subset \mathbb{R}$ is a subgroup which is not closed, and it is certainly not a Lie group.

If you do not quite know what a closed subset is, that is fine. The zero set of any smooth function f, i.e. the set $\{x \mid f(x) = 0\}$, is closed, and this is what we will use.

Definition 3.1.9. A closed subgroup of some $\mathrm{GL}(n,\mathbb{R})$ or $\mathrm{GL}(n,\mathbb{C})$ is called a *matrix Lie group*. They are Lie groups by theorem 3.1.8.

All groups given in examples 1.2.5 and 1.2.6 are matrix Lie groups. Let us give some more details on them.

Example 3.1.10. The special linear groups $\mathrm{SL}(n,\mathbb{R})$ and $\mathrm{SL}(n,\mathbb{C})$ are closed subgroups given by the condition $\det = 1$. This cuts down the dimension by one, so $\dim \mathrm{SL}(n,\mathbb{R}) = n^2 - 1$ and $\dim \mathrm{SL}(n,\mathbb{C}) = 2(n^2 - 1)$ – in the latter case the *complex* dimension came down by one.

These Lie groups are not compact. ◇

Example 3.1.11. The orthogonal group $\mathrm{O}(n)$ is a compact Lie group of dimension $\frac{n(n-1)}{2}$: specifying an orthogonal matrix is the same as specifying an orthonormal basis. For the first vector you require the norm to be 1, which gives $n-1$ independent coordinates. For the second, you also require it to be orthogonal to the first, so you get $n-2$ more coordinates. Iterating gives the dimension.

The special orthogonal group $\mathrm{SO}(n)$ is a compact sub-Lie group of the same dimension: the determinant of an orthogonal matrix is already ± 1, so enforcing it to be 1 does not reduce the number of coordinates. ◇

Example 3.1.12. The unitary group $\mathrm{U}(n)$ is a compact Lie group of dimension n^2, which can be calculated in a similar way to example 3.1.11.

The special unitary group $\mathrm{SU}(n)$ is a compact sub-Lie group of dimension $n^2 - 1$: any unitary matrix U has $\det U \in \mathrm{U}(1)$, so fixing this determinant to be 1 lowers the dimension by 1. ◇

Example 3.1.13. The symplectic group $\mathrm{Sp}(2n,\mathbb{R})$ is a non-compact Lie group of dimension $n(2n+1)$, and similarly, $\mathrm{Sp}(2n,\mathbb{C})$ is a non-compact Lie group of dimension $2n(2n+1)$.

The compact symplectic group $\mathrm{Sp}(2n)$ is a compact Lie group of dimension $n(2n+1)$. It can be seen as the quaternionic unitary group $\mathrm{U}(n,\mathbb{H})$. ◇

But there are more examples. Just to name a few:

Example 3.1.14. The group of upper-triangular matrices $\mathrm{T}(n,\mathbb{R}) \subseteq \mathrm{GL}(n,\mathbb{R})$ is a non-compact sub-Lie group of dimension $\frac{n(n+1)}{2}$. The subgroup of upper-triangular matrices with only ones on the diagonal $\mathrm{N}(n,\mathbb{R}) \subset \mathrm{T}(n,\mathbb{R})$ is a non-compact Lie group of dimension $\frac{n(n-1)}{2}$.

Similarly, we get $\mathrm{T}(n,\mathbb{C})$ and $\mathrm{N}(n,\mathbb{C})$. ◇

Example 3.1.15. The groups of diagonal matrices $\mathrm{D}(n,\mathbb{R}) \subseteq \mathrm{GL}(n,\mathbb{R})$ and $\mathrm{D}(n,\mathbb{C}) \subseteq \mathrm{GL}(n,\mathbb{C})$. They are isomorphic to $(\mathbb{R}^\times)^n$ and $(\mathbb{C}^\times)^n$, and have dimensions n and $2n$, respectively. ◇

Example 3.1.16. The group of affine transformations on the line $\mathrm{A}(1,\mathbb{R})$ is the group of maps $f_{a,b}\colon \mathbb{R} \to \mathbb{R}\colon x \to ax+b$, with $(a,b) \in \mathbb{R}^\times \times \mathbb{R}$, under composition. It is isomorphic to the matrix group of matrices of the form

$$T_{a,b} = \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} . \tag{3.1}$$

It has dimension 2 and is not compact. ◇

As usual, we also have a notion of maps between Lie groups.

Definition 3.1.17. Let G and H be Lie groups. A *Lie group homomorphism* is a smooth map $f\colon G \to H$ which is also a group homomorphism. A *Lie group isomorphism* is a bijective Lie group homomorphism.

Remark 3.1.18. There is some subtlety regarding the definition of a Lie group isomorphism. In general, an isomorphism of a certain type should be a map with an inverse of the same type. E.g. a group isomorphism is a group homomorphism with an inverse which is also a group homomorphism.

For abstract groups, the inverse of a bijective homomorphism is always a homomorphism, so our definition 1.3.1 is the right definition.

However, the inverse of a bijective smooth map does not have to be smooth (or even continuous). An example is the map $f\colon [0,1) \to \mathrm{U}(1)\colon x \mapsto e^{2\pi i x}$, whose inverse is discontinuous at $1 \in \mathrm{U}(1)$.

It turns out that for Lie groups, this issue cannot happen, and the definition above is correct, but this is fairly complicated to prove.

3.2 — Lie algebras

Now we have seen some examples of Lie groups, it is time to study their properties, and see if we can say something useful about them. The key to this is the interaction between the algebraic aspects – the group structure – and the analytic aspects – the manifold structure.

In calculus, as well as other analytic or geometric theories, we often look at functions locally, i.e. near a point. Taylor's theorem tells us that certain functions can be reconstructed from their derivatives, so from the local aspects. However, you still need to have the value of all derivatives at a chosen point.

For Lie groups, the situation is much better: we only have to look at first derivatives to understand nearly everything. Moreover, we only need to know these first derivatives at a special point: the unit of the Lie group. In more geometric terms, we want to consider the tangent space, the space of tangent vectors, at the identity. Let us try to understand this with an example.

Example 3.2.1. Consider the special orthogonal group $\mathrm{SO}(3)$, i.e. the group of rotations in $\mathbb{R}^3$. Pick an $A \in \mathrm{SO}(3)$ very close to the identity. This means we may write $A = \mathrm{Id} + M$, with M very small (a tangent). The condition that $A^TA = \mathrm{Id}$ translates to

$$\mathrm{Id} = (\mathrm{Id} + M^T)(\mathrm{Id} + M) \approx \mathrm{Id} + M^T + M \,, \tag{3.2}$$

where we neglected higher-order terms in M. This means that we need $M^T = -M$. We have now determined that the tangent space of $\mathrm{Id} \in \mathrm{SO}(3)$ is the space of antisymmetric matrices.

This only works for M very small. How do we deal with matrices farther from the identity, so when M^2 cannot be neglected? This is where the group theory starts interacting with the analysis. Any rotation is a rotation about an axis by a certain angle ϑ. Let us write it as $R(\vartheta)$. But then $R(\vartheta) = R(\vartheta/n)^n$ for any n, and we may take n arbitrarily large to get close to the identity.

Hence, if $R(\vartheta) = \mathrm{Id} + \vartheta M + \mathcal{O}(\vartheta^2)$, we find that

$$R(\vartheta) = \lim_{n\to\infty} \Big(1 + \frac{\vartheta}{n} M + \mathcal{O}\big(\frac{\vartheta}{n}\big)^2\Big)^n = e^{\vartheta M} . \tag{3.3}$$

So we can recover any element of the group by exponentiation!

What we mean by exponentiation is the Taylor series, i.e.

$$e^M = \sum_{k=0}^{\infty} \frac{1}{k!} M^k . \tag{3.4}$$

This makes sense in the vector space of matrices, and if we start with an antisymmetric M, we do in fact recover an element $e^M \in \mathrm{SO}(3)$, as $(e^M)^T e^M = e^{M^T} e^M = e^{-M} e^M = \mathrm{Id}$ and $\det e^M = e^{\mathrm{Tr}\, M} = e^0 = 1$, using that antisymmetric matrices must have only zeroes on the diagonal, and hence zero trace.

Now we found we can recover the set G (in fact the manifold) from its tangent space, what about the group operations? Of course, we have the unit, as that is the exponent of 0. If $A = e^{\vartheta M}$, then $A^{-1} = e^{-\vartheta M}$, so we also get inverses. Multiplication is a bit more complicated, though. Let us see what happens.

For real (or complex) numbers, we know that $e^x e^y = e^{x+y}$. This also works for matrices A and B, *as long as they commute*. But what if they do not? Let us first consider again the case that $A = \mathrm{Id} + M$ and $B = \mathrm{Id} + N$ with M, N small. Then we see that

$$AB = (\mathrm{Id} + M)(\mathrm{Id} + N) = \mathrm{Id} + M + N + MN , \tag{3.5}$$

while

$$BA = (\mathrm{Id} + N)(\mathrm{Id} + M) = \mathrm{Id} + M + N + NM . \tag{3.6}$$

So it seems that to understand the multiplication, we need to record the *commutator* $[M, N] = MN - NM$. and use that for corrections. This is not the same commutator as in definition 1.2.11, but it is related: it is in some sense an infinitesimal version of it – a derivative.

This is indeed the right thing to do, for two reasons: the first is that the commutator of two tangent vectors is again a tangent vector, so we remain in this tangent space we had. The second is that $e^C = e^A e^B$ can in fact be solved for C in a universal way involving only iterated commutators, by something called the *Baker–Campbell–Hausdorff* formula.

All in all, this tells us that all structure of the Lie group SO(3) can be encoded in its linearisation, i.e. tangent space, if we remember the commutator. This tangent space, $\mathfrak{so}(3)$, is the space of antisymmetric 3 by 3 matrices. ◇

The structure we have found in this example is a particular case of a Lie algebra. We will be able to do this for any Lie group, and then study Lie algebras in their own right. But let us define them first.

Definition 3.2.2. Let k be a field (we usually take $k = \mathbb{R}$). A *Lie algebra* $\mathfrak{g}$ over k is a k-vector space together with a binary operation $[\cdot,\cdot] \colon \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$, called the *Lie bracket*, which satisfies

L1 *Bilinearity:* for any $\lambda, \mu \in k$, $x, x', y, y' \in \mathfrak{g}$,

$$[\lambda x + x', y] = \lambda[x, y] + [x', y] , \qquad [x, \mu y + y'] = \mu[x, y] + [x, y'] ; \tag{3.7}$$

L2 *Antisymmetry:* for any $x, y \in \mathfrak{g}$,

$$[x, y] = -[y, x]\,; \tag{3.8}$$

L3 *Jacobi identity:* for any $x, y, z \in \mathfrak{g}$,

$$[x, [y, z]] + [y, [z, x]] + [z, [x, y]] = 0\,. \tag{3.9}$$

A Lie algebra $\mathfrak{g}$ is *abelian* if $[x, y] = 0$ for all $x, y \in \mathfrak{g}$.

A *Lie algebra morphism* from a Lie algebra $\mathfrak{g}$ to a Lie algebra $\mathfrak{h}$ is a linear map $f\colon \mathfrak{g} \to \mathfrak{h}$ such that $f([x, y]) = [f(x), f(y)]$ for all $x, y \in \mathfrak{g}$.

Now, the point of this definition is the following result.

THEOREM 3.2.3. *For any Lie group G, its tangent space at the origin,* $\mathrm{Lie}(G)$*, is a Lie algebra over $\mathbb{R}$, and there is an exponential map* $\exp\colon \mathrm{Lie}(G) \to G$ *which is a diffeomorphism (smooth bijection) near the origin.*

Given a Lie group homomorphism $f\colon G \to H$, there is a unique Lie algebra homomorphism $df\colon \mathrm{Lie}(G) \to \mathrm{Lie}(H)$ such that $e^{df(x)} = f(e^x)$. In particular, closed subgroups correspond to sub-Lie algebras.

REMARK 3.2.4. It is customary to denote Lie algebras by lower case *Fraktur* letters, e.g. $\mathfrak{a}, \mathfrak{b}, \mathfrak{c}$, &c. This is done both for generic Lie algebras and for the Lie algebras associated to concrete Lie groups, in which case the Lie algebra is denoted by the lower case Fraktur version of the upper case non-Fraktur name of the group.

These letters are very hard to write on a whiteboard (a blackboard would be better), so during lectures I will substitute them with *Sütterlin*.

In the case of matrix Lie groups, the construction works as in example 3.2.1: we consider matrices close to the identity, see what conditions we get to understand the vector space of the Lie algebra, and the Lie bracket is the commutator. The map df is really the derivative, or linearisation, of f at the identity.

REMARK 3.2.5. The exponential map is a diffeomorphism near the identity, meaning that it sends small neighbourhoods of the origin of $\mathrm{Lie}(G)$ to small neighbourhoods of the identity in G. But it is in general neither surjective nor injective globally. We will see examples of both.

However, it is surjective in case G is connected and compact. This is an important case, including, e.g., $\mathrm{SO}(n)$, $\mathrm{U}(n)$, $\mathrm{SU}(n)$, and $\mathrm{Sp}(n)$.

EXAMPLE 3.2.6. For the Lie group $\mathbb{R}$ from example 3.1.3, the Lie algebra is $\mathbb{R}$ itself, with trivial Lie bracket (so it is commutative). The exponential map is the identity.

This may sound a bit strange, but recall from example 1.3.7 that the additive group $\mathbb{R}$ is isomorphic to the multiplicative group $\mathbb{R}_{>0}$ via the exponential. So we could see $\mathbb{R}_{>0}$ as a Lie group with Lie algebra $\mathbb{R}$. ◇

EXAMPLE 3.2.7. For the circle Lie group $\mathrm{U}(1)$ from example 3.1.4, its Lie algebra is also $\mathbb{R}$, and the exponential map is $\exp\colon \mathbb{R} \to \mathrm{U}(1)\colon x \mapsto e^{ix}$. Here, we see the exponential map is not injective, as $\exp(2\pi) = \exp(0)$. ◇

So we see that different Lie groups may have the same Lie algebra.

EXAMPLE 3.2.8. The Lie algebra associated to $\mathrm{GL}(n, \mathbb{R})$ is the vector space of all n by n matrices, with the commutator $[A, B] = AB - BA$. It is denoted $\mathfrak{gl}(n, \mathbb{R})$. The exponential map is given by the power series expansion

$$e^X = \sum_{k=0}^{\infty} \frac{1}{k!} X^k\,, \tag{3.10}$$

which converges for small enough X.

Even in the case $n = 1$, this shows that the exponential map need not be surjective: $e^x > 0$ for all $x \in \mathfrak{gl}(1, \mathbb{R}) \cong \mathbb{R}$, so $-1 \notin \mathrm{Im}\exp$.

The Lie algebra $\mathfrak{gl}(n, \mathbb{C})$ is constructed similarly. ◇

Example 3.2.9. For the Lie algebra associated to $\mathrm{SL}(n, \mathbb{R})$, we need to check what the relation $\det A = 1$ corresponds to. For this, we use the identity $\det e^M = e^{\mathrm{Tr}\, M}$, which can be checked directly on upper triangular matrices and extended to all matrices by conjugation. From this, we see that $\mathfrak{sl}(n, \mathbb{R})$ is the Lie algebra of n by n real matrices with trace 0.

Again, the Lie algebra $\mathfrak{sl}(n, \mathbb{C})$ has a similar construction. ⋄

Example 3.2.10. We have already seen in example 3.2.1 that $\mathfrak{so}(3) = \mathrm{Lie}(\mathrm{SO}(3))$ is the space of 3 by 3 antisymmetric matrices. There was nothing particular about the number 3 in that part of the analysis; in general $\mathfrak{so}(n) = \mathrm{Lie}(\mathrm{SO}(n))$ is the space of n by n antisymmetric matrices.

In a completely analogous way, $\mathfrak{u}(n) = \mathrm{Lie}(\mathrm{U}(n))$ is the space of n by n anti-Hermitian matrices. And $\mathfrak{su}(n) = \mathrm{Lie}(\mathrm{SU}(n))$ is the space of n by n traceless anti-Hermitian matrices. ⋄

3.2.1 — Bases for Lie algebras

Lie algebras are very commonly given by giving an explicit basis, together with an expression of the Lie bracket in terms of the basis vectors. The basis elements of a Lie algebra are often called the *(infinitesimal) generators* of the corresponding Lie group. We will go over a few simple cases here, after we give the motivation for the name generator.

Proposition 3.2.11. *Let G be a connected Lie group. Then G is generated (in the sense of definition 1.5.22) by elements of the form e^X with $X \in \mathrm{Lie}(G)$.*

Remark 3.2.12. The word 'connected' means that it is possible to make a continuous path between any two points in the Lie group. E.g., the circle group $\mathrm{SO}(2) \cong \mathrm{U}(1)$ is connected, but the group $\mathrm{O}(2)$, which looks like two disjoint circles, is not.

Examples of connected Lie groups are $\mathrm{GL}(n, \mathbb{C})$, $\mathrm{SL}(n, \mathbb{R})$, $\mathrm{SL}(n, \mathbb{C})$, $\mathrm{SO}(n)$, $\mathrm{U}(n)$, $\mathrm{SU}(n)$, $\mathrm{Sp}(2n, \mathbb{R})$, $\mathrm{Sp}(2n, \mathbb{C})$, and $\mathrm{Sp}(n)$.

Examples of disconnected Lie groups are $\mathrm{GL}(n, \mathbb{R})$ and $\mathrm{O}(n)$ – both with two connected components identified by the sign of the determinant – but also all non-trivial discrete groups, and more generally the product of any Lie group with a discrete group.

In case of a not necessarily connected Lie group, the elements of the form e^X generate a subgroup of G, which is the connected component containing the identity.

This proposition does not mean that every element of a connected Lie group G can be written as e^X for some $X \in \mathrm{Lie}(G)$, though! For example, $\begin{pmatrix} -1 & -1 \\ 0 & -1 \end{pmatrix} \in \mathrm{SL}(2, \mathbb{C})$ cannot be written this way.

There is an important difference in convention between physics and mathematics on bases for Lie algebras. The most important Lie groups for physics are compact Lie groups, whose Lie algebras naturally consist of anti-Hermitian matrices (see e.g. example 3.2.10). However, physical observables should be Hermitian, as their eigenvalues – which correspond to the outcomes of measurements – should be real. Therefore, physicists introduce extra factors of i to force the generators to be Hermitian. This has the side effect that in the physics convention, factors of i pop up in other places as well, such as in the Lie bracket. Specifically, physicists multiply their Lie algebras by $-i$, such that their exponential map becomes $\exp\colon \mathrm{Lie}(G) \to G\colon B \mapsto e^{iB}$.

Mathematicians do not usually introduce these factors of i, as they clutter equations, and make the generators not lie in the Lie algebra in general. Also, over general fields, i is not even defined.

We will stick to the physics convention from now on, but please keep in mind this difference in convention when looking for other resources, or in future courses.

Example 3.2.13. In the case of $\mathrm{SO}(2)$ (which is isomorphic to the circle group $\mathrm{U}(1)$, cf. example 1.5.18), any matrix is of the form

$$A = \begin{pmatrix} \cos\vartheta & -\sin\vartheta \\ \sin\vartheta & \cos\vartheta \end{pmatrix} . \tag{3.11}$$

Writing this up to first order in ϑ and remembering that we use the physics convention, we get $A = \mathrm{Id} + iX\vartheta + \mathcal{O}(\vartheta^2)$, where

$$X = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix} \tag{3.12}$$

is the generator of $\mathfrak{so}(2)$. The Lie bracket is trivial: $[X, X] = 0$.

In the picture of $\mathrm{U}(1)$, we get $A = e^{i\vartheta} = 1 + i\vartheta + \mathcal{O}(\vartheta^2)$, so in this case the generator is 1. (What we are doing here is viewing different representations of isomorphic Lie groups.) ◇

EXAMPLE 3.2.14. In the case of SO(3), studied in example 3.2.1, a standard choice of basis for $\mathfrak{so}(3)$ is

$$L_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad L_2 = \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix}, \quad L_3 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{3.13}$$

and in this basis, the Lie bracket is $[L_j, L_k] = i \sum_{l=1}^{3} \varepsilon_{jkl} L_l$, where ε is the Levi-Civita symbol, given by

$$\varepsilon_{jkl} = \begin{cases} 1 & \text{if } (j\,k\,l) = (1\,2\,3) \in S_3; \\ -1 & \text{if } (j\,k\,l) = (1\,3\,2) \in S_3; \\ 0 & \text{if an index is repeated.} \end{cases} \tag{3.14}$$

Another way of thinking about these generators is in the terms of tangent vectors. A tangent vector is really a first order differential operator, and looking from this point of view, the basis vectors above can be represented as follows:

$$L_1 = iz\partial_y - iy\partial_z, \quad L_2 = ix\partial_z - iz\partial_y, \quad L_3 = iy\partial_x - ix\partial_y. \tag{3.15}$$

If you have taken quantum mechanics courses before, you may recognise these as the angular momentum operators. This is of course no coincidence: angular momentum is nothing more than the speeds of rotation in different directions. ◇

EXAMPLE 3.2.15. Let us now consider SU(2). We have seen in example 3.2.10 that $\mathfrak{su}(2)$ is the space of 2 by 2 traceless anti-Hermitian matrices, and we can take a basis

$$T_1 = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad T_2 = \frac{1}{2}\begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}, \quad T_3 = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{3.16}$$

These matrices are called *Pauli matrices*, cf. also the exercise on irreducible representations of the quaternion group. The Lie bracket for these matrices is

$$[T_j, T_k] = i \sum_{l=1}^{3} \varepsilon_{jkl} T_l. \tag{3.17}$$

This is the exact same as we found for $\mathfrak{so}(3)$! So we find that $\mathfrak{su}(2) \cong \mathfrak{so}(3)$. We will see what this means for the Lie groups in section 3.2.2. ◇

EXAMPLE 3.2.16. Another fairly small Lie group we could consider is $\mathrm{GL}(2, \mathbb{R})$. By example 3.2.8, $\mathfrak{gl}(2, \mathbb{R})$ is the vector space of all real 2 by 2 matrices, with the commutator. This is not a compact Lie group, and does not have any condition of being (anti-)Hermitian, so we will *not* use the physicists' convention here.

We could choose the standard basis for the space of matrices, but that is not quite convenient. In fact, this algebra has a non-trivial *centre*: the subspace of elements x for which $[x, y] = 0$ for all elements

y. In this case, this centre is the one-dimensional space of multiples of the identity matrix, and it has a complement $\mathfrak{sl}(2,\mathbb{R})$. So a better basis would be

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad E = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad F = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad H = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{3.18}$$

where the last three elements span $\mathfrak{sl}(2,\mathbb{R})$. Of course, this is also a basis for $\mathfrak{gl}(2,\mathbb{C})$ or $\mathfrak{sl}(2,\mathbb{C})$ as complex Lie algebras. ◇

3.2.2 — From Lie algebras to Lie groups

We have seen that any Lie group gives rise to a Lie algebra, cf. theorem 3.2.3. How do we go back? The answer cannot be quite straightforward, as by examples 3.2.14 and 3.2.15 different Lie groups can have isomorphic algebras. So what is happening?

In fact, we can go in the other direction, but we will not reach all Lie groups. the statement is the following.

Theorem 3.2.17. *For any real Lie algebra $\mathfrak{g}$, there is a simply connected Lie group G such that* $\mathrm{Lie}(G) = \mathfrak{g}$, *and this G is unique up to unique isomorphism.*

Moreover, for any morphism $f\colon \mathfrak{g} \to \mathfrak{h}$, there is a unique morphism $F\colon G \to H$, where G and H are the simply connected Lie groups as above, such that $dF = f$.

Ignoring whatever 'simply connected' means, this does tell us that we can go back to Lie groups. But what does 'simply connected' mean? There are two parts to this.

Firstly, the space must be connected. This is the same condition as in proposition 3.2.11.

Secondly, any loop drawn on the space must be contractable to a point. For some two-dimensional examples, the sphere S^2 is simply connected, but the torus $S^1 \times S^1$ is not.

From the viewpoint of algebraic topology, simply connected spaces are special because any connected space has a 'universal cover' which is simply connected. In the case of a Lie group, this is what we get if we take the Lie algebra and go back to a Lie group.

All of this may sound like completely nonsense to you, so let us look at some examples.

Example 3.2.18. There is exactly one one-dimensional Lie algebra $\mathfrak{g}$ up to isomorphism, as for any two element $x, y \in \mathfrak{g}$, we must have $[x,y] = 0$, by bilinearity and antisymmetry (which implies that $[x,x] = 0$). So this $\mathfrak{g}$ is abelian.

But we have seen three one-dimensional Lie groups in examples 3.1.3 to 3.1.5. These were $\mathbb{R}$, $\mathbb{R}^\times$, and $\mathrm{U}(1)$. The first of these is the simply connected Lie group reconstructed in theorem 3.2.17 (and the exponential map is the identity here, as found in example 3.2.6). The second is not connected, as the components of positive and negative numbers are disjoint. The last is connected, but not simply connected, as the loop given by the circle itself is not contractible.

And there are in fact maps $\exp\colon \mathbb{R} \to \mathbb{R}^\times$ and $f \circ q\colon \mathbb{R} \to \mathbb{R}/\mathbb{Z} \to \mathrm{U}(1)$, which we found in example 1.3.7 and example 1.5.18. ◇

Example 3.2.19. Let us consider $\mathfrak{so}(3) \cong \mathfrak{su}(2)$, as in examples 3.2.14 and 3.2.15. We will use the bases $\{L_j\}$ and $\{T_j\}$ from those examples. On Lie algebras, the isomorphism is given by $f\colon \mathfrak{su}(2) \to \mathfrak{so}(3)\colon T_j \mapsto L_j$. By remark 3.2.5, any element in $SU(2)$ or $SO(3)$ is in the image of the relevant exponential map, as both groups are compact and connected. So the only possible map should have the shape

$$F\colon SU(2) \to SO(3)\colon \exp\left(i \sum_{j=1}^{3} \vartheta_j T_j\right) \mapsto \exp\left(i \sum_{j=1}^{3} \vartheta_j L_j\right). \tag{3.19}$$

This is in fact a well-defined map, but it is not an isomorphism: it is not injective. For example, $\exp(2\pi i T_1) = -\mathrm{Id}$, while $\exp(2\pi i L_1) = \mathrm{Id}$. This is the only issue that may occur, though: the kernel of the map F is

$\operatorname{Ker} F = \{\mathrm{Id}, -\mathrm{Id}\}$. This tells us that F is two-to-one, a double cover. We will see this more thoroughly in example 3.3.12.

Returning to theorem 3.2.17, $SU(2)$ is the simply connected Lie group related to $\mathfrak{su}(2) \cong \mathfrak{so}(3)$, and is the universal cover of $SO(3)$. The fact that this non-trivial cover of $SO(3)$ exists, is what gives rise to fermions in quantum mechanics: they transform not as a representation of $SO(3)$ under rotations, but as a representation of $SU(2)$. ⋄

3.3 — Lie representations

We have determined the objects of study for the remainder of the course: these are Lie groups and Lie algebras. We have seen that we can go from Lie groups to Lie algebras, and we can go back in a restricted way.

In the first part of the course, we studied abstract groups, and then we studied their representations in case the group was finite. We can do something similar here. We will introduce the notions of representations of Lie groups and Lie algebras, and connect the two.

3.3.1 — Lie group representations

As before, we start with the story for Lie groups.

Definition 3.3.1. Let G be a Lie group. A *representation* of G is a pair (V, ρ), where V is a complex vector space and ρ is a Lie group homomorphism $\rho\colon G \to \mathrm{GL}(V)$.

Remark 3.3.2. Even though Lie groups are intrinsically real objects, we define their representations on complex vector spaces. This is for the same reason as before: we want to be able to use Schur's lemma 2.4.2. Restricting to real representations is possible in the same way as before, cf. section 2.7.

We will give a few basic examples here.

Example 3.3.3. For any Lie group G, the *trivial representation* is the same as before, i.e. the map $\tau\colon G \to \mathbb{C}^\times : g \mapsto 1$. ⋄

Example 3.3.4. For any matrix Lie group G , by definition 3.1.9 a closed Lie subgroup of $\mathrm{GL}(n, \mathbb{R})$ or $\mathrm{GL}(n, \mathbb{C})$, its *defining representation* is the inclusion $j\colon G \to \mathrm{GL}(n, \mathbb{C})$. Here, if G was defined as a closed subgroup of $\mathrm{GL}(n, \mathbb{R})$ (e.g. $G = \mathrm{O}(n)$), we use the inclusion $\mathrm{GL}(n, \mathbb{R}) \subset \mathrm{GL}(n, \mathbb{C})$ to view it as a closed subgroup of $\mathrm{GL}(n, \mathbb{C})$ anyways. ⋄

Example 3.3.5. Any Lie group G has a representation on its Lie algebra $\mathfrak{g} = \mathrm{Lie}(G)$ by conjugation. This is called the *adjoint representation* and is given as

$$\mathrm{Ad}\colon G \to \mathrm{GL}(\mathfrak{g})\colon g \mapsto \mathrm{Ad}_g\,, \qquad \mathrm{Ad}_g(x) = gxg^{-1}\,. \tag{3.20}$$

There are a few things to check here, try and do this yourself.

This representation may remind you of the conjugation action of proposition 1.4.13. ⋄

Many constructions we had for group representations carry over to representations of Lie groups: there are still transformations between representations, equivalences, direct sums, tensor products, irreducible representations, and characters.

For *compact* Lie groups, many results from the representation theory of *finite* groups go through. The main observation is that one should replace the sum $\frac{1}{|G|}\sum_{g\in G} f(g)$ by an integral $\int_G f(g)dg$, where dg is the *Haar measure* on G, the (left) translation-invariant measure with weight $\int_G dg = 1$. For example, for the circle group $\mathrm{U}(1) = \{e^{i\vartheta} \mid 0 \le \vartheta < 2\pi\}$, the Haar measure is $\frac{d\vartheta}{2\pi}$.

Let us give a list of results that hold in this context:

- The unitarity theorem 2.3.9: any finite-dimensional representation of a compact Lie group is equivalent to a unitary one;
- Theorem 2.3.10: any irreducible representation of a compact Lie group is finite-dimensional;
- Maschke's theorem, corollary 2.3.12: any finite-dimensional representation of a compact Lie group is semisimple;
- Schur's lemma 2.4.2: transformations between irreducible representations are either 0 or invertible. Transformations from an irreducible representation to itself are multiples of the identity;
- Corollary 2.4.3: all irreducible representations of abelian Lie groups are one-dimensional;
- All of the results on characters in section 2.5.2;
- Part of corollary 2.6.4 and all of theorem 2.6.5: simple characters of Lie groups are an orthonormal set, and in fact a basis of the space of square-integrable functions on the Lie group. However, there are infinitely many conjugacy classes, and infinitely many simple characters as well;
- Propositions 2.6.11 and 2.6.13: representations can be decomposed and the multiplicities calculated via the inner product.
- Part of corollary 2.6.12: the regular representation is the Hilbert space of square-integrable functions on the Lie group. This is infinite-dimensional, so it has no character, but the multiplicity of any irreducible representation in it is still equal to its dimension.

In most cases, the proofs go through when replacing sums by integrals.

One of the important changes is the number of irreducible representations: there are infinitely many for Lie groups. This can already be seen in the very simplest example.

Example 3.3.6. Consider the circle group $\mathrm{U}(1)$. This is abelian, so all its irreducible representations must be one-dimensional. As we may restrict to unitary representations, an irreducible representation of $\mathrm{U}(1)$ is a Lie group homomorphism $\rho\colon \mathrm{U}(1) \to \mathrm{U}(1)$. This means we need to have $\rho(e^{i\vartheta}) = e^{in\vartheta}$ for some $n \in \mathbb{R}$. Moreover, we need $\rho(e^{2\pi i}) = 1$, so we need $n \in \mathbb{Z}$. But any value of $n \in \mathbb{Z}$ works, so we get infinitely many irreducible representations $\rho_n\colon \mathrm{U}(1) \to \mathrm{U}(1)\colon e^{i\vartheta} \mapsto e^{in\vartheta}$ for all $n \in \mathbb{Z}$.

As characters form a basis for the space of square-integrable functions on a compact Lie group by the above, in this case we get a decomposition of square-integrable functions on $\mathrm{U}(1)$, i.e. periodic functions with period 2π, in *Fourier modes*. ◇

So writing down a character table is not really achievable for a Lie group.

How do we then find these irreducible representations? This is where Lie algebras come to the rescue.

3.3.2 — Lie algebra representations

The definition of a representation of Lie algebras is very similar to that of a Lie group, definition 3.3.1.

Definition 3.3.7. Let $\mathfrak{g}$ be a Lie algebra (over a field k). A *representation* of $\mathfrak{g}$ is a pair (V, r) of a k-vector space V and a homomorphism of Lie algebras $r\colon \mathfrak{g} \to \mathfrak{gl}(V)$.

If $\mathfrak{g}$ is a real Lie algebra, we often still consider its representations on a complex vector space. These are then the representations of its *complexification*, $\mathfrak{g}_{\mathbb{C}} = \mathfrak{g} \oplus i\mathfrak{g}$.

Of course, this definition is made in such a way that it is related to Lie group representations.

Proposition 3.3.8. *Let G be a Lie group, and (V, ρ) a Lie group representation of G. Then $d\rho\colon$ $\mathrm{Lie}(G) \mapsto \mathrm{Lie}(\mathrm{GL}(V)) = \mathfrak{gl}(V)$ is a Lie algebra representation of* $\mathrm{Lie}(g)$.

Conversely, if G is simply connected, then for any Lie algebra representation (V, r) of $\mathrm{Lie}(G)$*, there is a Lie group representation (V, ρ) of G such that $r = d\rho$, and this ρ is unique up to equivalence.*

Example 3.3.9. For any Lie algebra $\mathfrak{g}$, the *trivial representation* is $t\colon \mathfrak{g} \to \mathfrak{gl}(1, \mathbb{C}) \cong \mathbb{C}\colon g \mapsto 0$. This is related to the trivial representation τ for Lie groups given in example 3.3.3: on the one hand, τ is constant, so its derivative $d\tau$ is indeed equal to 0, i.e. to t. On the other hand, $e^{t(g)} = e^0 = 1 = \tau(e^g)$. ◇

Example 3.3.10. Applying the construction of proposition 3.3.8 to example 3.3.4, the Lie algebra of any matrix Lie group of definition 3.1.9 has a *defining representation*: if $G \subseteq \mathrm{GL}(n, \mathbb{R})$ is a matrix Lie group, and we write j for the inclusion homomorphism, then $dj\colon \mathrm{Lie}(G) \to \mathfrak{gl}(n, \mathbb{R})$ is a homomorphism of Lie algebras by theorem 3.2.3, so this is a representation. This is just the natural inclusion map.

For example, if $G = \mathrm{SU}(n)$, the defining Lie group representation is $j\colon \mathrm{SU}(n) \to \mathrm{GL}(n, \mathbb{C})\colon A \mapsto A$, and $dj\colon \mathfrak{su}(n) \to \mathfrak{gl}(n, \mathbb{C})\colon x \mapsto x$. ◇

Example 3.3.11. Applying the construction of proposition 3.3.8 to example 3.3.5, and realising that any Lie algebra is the Lie algebra of some Lie group by theorem 3.2.17, we find that any Lie algebra $\mathfrak{g}$ has a representation on itself which is called the *adjoint representation* $\mathrm{ad}\colon \mathfrak{g} \to \mathfrak{gl}(\mathfrak{g})\colon x \mapsto \mathrm{ad}_x$. Explicitly, it is given by $\mathrm{ad}_x(y) = [x, y]$. There are several things to check here:

- This is indeed a map $\mathfrak{g} \to \mathfrak{gl}(\mathfrak{g})$, i.e. $\mathrm{ad}_x \in \mathfrak{gl}(\mathfrak{g})$. This holds because the Lie bracket is linear in its second argument.
- This map is linear. This holds because the Lie bracket is linear in its first argument.
- The map is a homomorphism of Lie algebras, i.e. $[\mathrm{ad}_x, \mathrm{ad}_x] = \mathrm{ad}_{[x,y]}$. Let us check how the left-hand side acts on a third element z:
$$\begin{aligned} \left[\mathrm{ad}_x, \mathrm{ad}_y\right](z) &= \mathrm{ad}_x\left(\mathrm{ad}_y(z)\right) - \mathrm{ad}_y\left(\mathrm{ad}_x(z)\right) \\ &= \mathrm{ad}_x\left([y,z]\right) - \mathrm{ad}_y\left([x,z]\right) \\ &= \left[x, [y,z]\right] - \left[y, [x,z]\right]. \end{aligned} \tag{3.21}$$
On the other hand, the right-hand side gives
$$\mathrm{ad}_{[x,y]}(z) = \left[[x,y], z\right], \tag{3.22}$$
and these two are equal by the antisymmetry and the Jacobi identity. Check this yourself, it is a good exercise!

The adjoint represention in some sense (but not all senses!) is the analogue of the regular representation for finite groups of example 2.1.8. ◇

Example 3.3.12. Let us consider the adjoint representation of $\mathfrak{su}(2)$. We can do this in the basis found in example 3.2.15. As $\dim \mathfrak{su}(2) = 3$, we have to associate to each T_j a 3 by 3 matrix ad_{T_j} such that in terms of coefficients
$$\mathrm{ad}_{T_j} T_k = \sum_{l=1}^{3} (\mathrm{ad}_{T_j})_{kl} T_l\,. \tag{3.23}$$
But we know that $\mathrm{ad}_{T_j} T_k = [T_j, T_k]$ by definition, and we have an explicit expression for these Lie brackets from example 3.2.15, so $(\mathrm{ad}_{T_j})_{kl} = i\varepsilon_{jkl}$. We find that
$$\mathrm{ad}_{T_1} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad \mathrm{ad}_{T_2} = \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix}, \quad \mathrm{ad}_{T_3} = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{3.24}$$
but these are exactly the matrices L_i from example 3.2.14. So in this case, the adjoint representation of $\mathfrak{su}(2) \cong \mathfrak{so}(3)$ is equal to the defining representation of the latter. But this is a very special occurence.

Retranslating this to Lie groups, we get a map $\mathrm{Ad}\colon \mathrm{SU}(2) \mapsto \mathrm{SO}(3)$, where we see $\mathrm{SO}(3)$ as the space of special orthogonal transformations of the vector space $\mathfrak{su}(2)$, and as we know that $\mathrm{Ad}_U(X) = UXU^{-1} = UXU^\dagger$, we find that $\mathrm{Ker}\,\mathrm{Ad} = \{\mathrm{Id}, -\mathrm{Id}\}$, so we have constructed a two-to-one map from $\mathrm{SU}(2)$ to $\mathrm{SO}(3)$, as anticipated in example 3.2.19. ◇

3.3.3 — Representations of $\mathfrak{sl}(2, \mathbb{C})$

Let us consider complex representations of a very specific Lie algebra: namely $\mathfrak{sl}(2, \mathbb{C})$. We can actually construct all irreducible representations (up to equivalence) explicitly. We do this in several steps.

What is a representation of $\mathfrak{sl}(2, \mathbb{C})$? Concretely, we need to find a complex vector space V and three matrices $e = r(E)$, $f = r(F)$, and $h = r(H)$ from V to V such that they satisfy the Lie bracket of example 3.2.16

Let us start by assuming that h is diagonal. This is not much of an assumption, as we only look for representations up to equivalence. As we are looking for finite-dimensional representations, it has finitely many eigenvalues, which we call *weights*. Let us pick one, a, such that $a + 2$ is not a weight. We write v_a for its eigenvector, i.e. $hv_a = av_a$. For now, we only know $a \in \mathbb{C}$.

For convenience, we normalise to $|v_a| = 1$.

Because we know that $[h, e] = 2e$, we see that $ev_a = 0$ is also an eigenvector of h, as $hev_a = [h, e]v_a + ehv_a = 2ev_a + aev_a = (a + 2)ev_a$, but h did not have $a + 2$ as an eigenvalue. Because of this, we call a a *highest weight*.

On the other hand, because $[h, f] = -2f$, we find that fv_a is an eigenvector of weight $a - 2$, if it is non-zero. We may write it as v_{a-2}. By iteration, we find that $f^k v_a$ has weight $a - 2k$. We call it $v_{a-2k} = f^k v_a$. Because we assumed that V was finite-dimensional, this must stop at some point, i.e. there must be some k such that $v_{a-2k} \neq 0$, but $fv_{a-2k} = 0$. But what is this k? For this, we need to analyse the action of e. We can already see that $ev_b = N_b v_{b+2}$ for some N_b.

Using that $[e, f] = h$, we find that

$$N_b v_{b+2} = ev_b = efv_{b+2} = (fe + h)v_{b+2} = (N_{b+2} + b + 2)v_{b+2} \tag{3.25}$$

Therefore, $N_b = N_{b+2} + b + 2$.

Also, $N_a = 0$, so $N_{a-2} = a$, $N_{a-4} = 2a - 2$, and in general,

$$N_{a-2s} = \sum_{k=0}^{s-1} a - 2k = sa - s(s-1)\,. \tag{3.26}$$

Now, let us consider v_{a-2k} again. We chose it such that $fv_{a-2k} = 0$. But then

$$\begin{aligned} 0 &= efv_{a-2k} \\ &= (fe + h)v_{a-2k} \\ &= (N_{a-2k} + a - 2k)v_{a-2k}\,. \end{aligned} \tag{3.27}$$

This tells us that $N_{a-2k} = 2k - a$, and solving $ka - k(k-1) = 2k - a$ for k gives two solutions: either $k = -1$, which is impossible as it must be an non-negative integer, or $k = a$. In particular, this must be an integer. So the lowest weight state is v_{-a}, and there are $a + 1$ states.

Theorem 3.3.13. *For any non-negative integer a, these exists a unique irreducible $\mathbb{C}$-representation r_a of $\mathfrak{sl}(2, \mathbb{C})$ of dimension $a + 1$, as constructed above. These are all irreducible $\mathbb{C}$-representations of $\mathfrak{sl}(2, \mathbb{C})$.*

We have not actually proved that the representations we constructed above are irreducible, and we will not. It is beyond the scope of this course.

Corollary 3.3.14. *The complex representations of $\mathfrak{su}(2)$ have the same structure as those of $\mathfrak{sl}(2, \mathbb{C})$.*

Proof. For complex representations, we may complexify $\mathfrak{su}(2)$, cf. definition 3.3.7. But $\mathfrak{su}(2)_{\mathbb{C}} = \mathfrak{su}(2) \oplus i\mathfrak{su}(2) = \mathfrak{sl}(2, \mathbb{C})$, as any (traceless) matrix is uniquely the sum of a Hermitian and an anti-Hermitian matrix. □

REMARK 3.3.15. For $\mathfrak{su}(2)$- or $\mathfrak{so}(3)$-representations, the usual choice is to rescale the matrices $L_+ = \frac{1}{\sqrt{2}}(L_3 - iL_2) = \frac{1}{\sqrt{2}}E$, $L_- = \frac{1}{\sqrt{2}}(L_3 + iL_2) = \frac{1}{\sqrt{2}}F$, and $L_1 = \frac{1}{2}H$, where the L_k are the basis as in example 3.2.14.

In this basis, $[L_1, L_+] = L_+$, $[L_1, L_-] = -L_-$, and $[L_+, L_-] = L_1$. In this context, L_+ and L_- are called *ladder operators*, or *raising* resp. *lowering operator*. Note that L_+ and L_- are actually not in $\mathfrak{su}(2)$, but in its complex version $\mathfrak{su}(2) + i\mathfrak{su}(2) = \mathfrak{sl}(2, \mathbb{C})$.

With this normalisation, the weights of L_1 are half-integers, and the highest weight is as well. These representations are called *spin representations* s_j and $j = \frac{a}{2}$ is called the *spin*. Notation is $|j, m\rangle$ for the normalised eigenvalue $m = \frac{b}{2}$ state in the representation with highest weight j. In particular, $-j \leq m \leq j$, and they are both half-integers such that $j - m \in \mathbb{Z}$.

In this normalisation, we get

$$\begin{aligned} s_j(L_1)|j, m\rangle &= m|j, m\rangle\,, \\ s_j(L_+)|j, m\rangle &= \sqrt{\frac{1}{2}(j+m+1)(j-m)}|j, m+1\rangle\,, \\ s_j(L_-)|j, m\rangle &= \sqrt{\frac{1}{2}(j+m)(j-m+1)}|j, m-1\rangle\,. \end{aligned} \tag{3.28}$$

We are using Dirac's bra-ket notation here. That is, we write vectors as $|v\rangle$ (this is called a *ket*), and their duals as $\langle v| = |v\rangle^\dagger$ (this is a *bra*). In this notation, the inner product is $\langle u|v\rangle = \sum_{i=1}^n \bar{u}_i v_i$.

By proposition 3.3.8, we now know that these are also the irreducible representations of $\mathrm{SU}(2)$, as $\mathrm{SU}(2)$ is simply connected. But not all of these are representations of $\mathrm{SO}(3)$. Let us investigate what is going on here.

First of all, how do we get an explicit representation for $\mathrm{SU}(2)$ out of this? Well, any element $A \in \mathrm{SU}(2)$ is the exponent $A = e^X$ of some $X \in \mathfrak{su}(2)$, so $\sigma_j(A) = e^{s_j(X)}$ as matrices. What we often do, is to write this out for the basis we have. For example,

$$e^{i\vartheta s_j(L_1)}|j, m\rangle = e^{i\vartheta m}|j, m\rangle\,. \tag{3.29}$$

The others are harder to write, of course, as they are not diagonal. On the other hand, they are *nilpotent*: $s_j(L_+)^{j+1} = s_j(L_-)^{j+1} = 0$. So in the series expansion of the exponential, $e^x = \sum_{k=0}^\infty \frac{1}{k!}x^k$, only finitely many terms contribute.

This is already sufficient to see what issues we may have in the case of $\mathrm{SO}(3)$. We have seen in example 3.3.12 that the adjoint representation gave a surjective map $\mathrm{Ad}\colon \mathrm{SU}(2) \to \mathrm{SO}(3)$ with kernel $\operatorname{Ker}\mathrm{Ad} = \{\pm\mathrm{Id}\}$. So this is a quotient map, and $\mathrm{SU}(2)/\{\pm\mathrm{Id}\} \cong \mathrm{SO}(3)$. For a representation $\sigma_j\colon \mathrm{SU}(2) \mapsto \mathrm{GL}(2j+1, \mathbb{C})$ to 'descend' to one of $\mathrm{SO}(3)$, we need it to be trivial on this kernel. In other words, we need that $\sigma_j(-\mathrm{Id}) = \mathrm{Id} \in \mathrm{GL}(2j+1, \mathbb{C})$.

Now, $-\mathrm{Id} = e^{2\pi i L_1}$, so by equation (3.29), we find

$$\sigma_j(-\mathrm{Id})|j, m\rangle = e^{2\pi i m}|j, m\rangle\,. \tag{3.30}$$

If j is an integer, then m is too, and hence all eigenvalues of $\sigma_j(-\mathrm{Id})$ are equal to 1, i.e. $\sigma_j(-\mathrm{Id}) = \mathrm{Id}$. In this case, we get a representation of $\mathrm{SO}(3)$.

However, if j is a strict half-integer, then m is too, and the eigenvalues of $\sigma_j(-\mathrm{Id})$ are equal to -1. So these cases do not give representations of $\mathrm{SO}(3)$. We have proved the following result.

PROPOSITION 3.3.16. *The irreducible representations of* $\mathrm{SU}(2)$ *are the same as those of* $\mathfrak{su}(2)$. *They are indexed by the maximal weight or spin j, which can be any non-negative half-integer, and* $\dim\sigma_j = 2j+1$.

This gives a representation of $\mathrm{SO}(3)$ *if and only if j is an integer.*

This proposition is the reason for the particular choice of $L_1 \in \mathfrak{so}(3)$: it is chosen such that its highest weights are all non-negative integers.

REMARK 3.3.17. In physics, irreducible representations are mostly denoted by their dimension (often, but not always, in bold-face), so e.g. **3** is the spin-1 representation. In case there are several irreducible representations of the same dimension, these numbers may be decorated. E.g. if some 6-dimensional representation is complex, we could write **6** and $\bar{\mathbf{6}}$ for that representation and its complex conjugate.

There are several more useful things to remark in this example. The first of these is the existence of a quadratic *Casimir invariant*. We can construct an element $L^2 = L_1^2 + L_2^3 + L_3^2 = L_+L_- + L_-L_+ + L_1^2$. This is not an element of $\mathfrak{so}(3)$ (in fact it lies in something called its universal enveloping algebra), but it is useful to look at anyways: it does naturally act on any representation of $\mathfrak{so}(3)$. Its most striking feature is that it commutes with all of $\mathfrak{so}(3)$, so by Schur's lemma, it must act by a scalar in any representation s_j. By calculating, we find that

$$s_j(L^2)|j,j\rangle = j(j+1)|j,j\rangle\,, \tag{3.31}$$

so we get that $s_j(L^2) = j(j+1)\mathrm{Id}$. This scalar, $j(j+1)$, corresponds to the total angular momentum squared of a quantum-mechanical system.

Secondly, there is some more structure relating these irreducible representations. In fact, the representations of $\mathrm{SU}(2)$ are the *symmetric powers* of the defining representation, $\sigma_j = \mathrm{Sym}^{2j}\,\sigma_{1/2}$. These symmetric powers are a generalisation of the symmetric square from assignment 3.

Thirdly, we can come back to the differential operator representation of $\mathfrak{so}(3)$, given by equation (3.15). Written in this way, they act naturally on the space of harmonic functions $f\colon \mathbb{R}^3 \to \mathbb{C}$ (with coordinates x, y, z), i.e. such that $\Delta f = \nabla^2 f = 0$. By explicit computation, the linear functions form a representation equivalent to the defining (spin-1) representation of $\mathrm{SO}(3)$, and homogeneous degree j harmonic polynomials form the spin-j representation.

If we change to spherical coordinates (r, ϑ, φ), we can ignore the r-dependence, as rotations leave it fixed (this is conventional in physics and is the reason we can restrict to harmonic polynomials). Using the story above, we can find a basis of functions $Y_j^m(\vartheta,\varphi)$ such that $L_1 Y_j^m = mY_j^m$ and $L^2 Y_j^m = j(j+1)Y_j^m$. These are called the *spherical harmonics*, and they are essential to understanding spherically symmetrical problems. In particular, any study of the hydrogen atom is sure to include them.

Finally, there is an explicit formula for tensor products of $\mathfrak{su}(2)$-representations.

PROPOSITION 3.3.18 (CLEBSCH-GORDAN FORMULA). *Denote as above by s_j the spin-j representation of $\mathfrak{su}(2)$. Then*

$$s_j \otimes s_k \cong \bigoplus_{l=0}^{2\min\{j,k\}} s_{j+k-l}\,. \tag{3.32}$$

This formula is used to calculate the combined spin of pairs (and other combinations) of particles. For example, combining three spin-$\frac{1}{2}$ particles does not always result in a spin-$\frac{3}{2}$ particle, but may also give a spin-$\frac{1}{2}$ particle – this is actually the case with protons and neutrons, both of which are spin-$\frac{1}{2}$ particles made up of three spin-$\frac{1}{2}$ quarks.

3.3.4 — PROJECTIVE AND SPIN REPRESENTATIONS

In the previous section, we saw that not all representations of $\mathrm{SU}(2)$ are also representations of $\mathrm{SO}(3)$. However, the ones that are not are still of physical relevance. In this section, we investigate this a bit deeper.

In quantum physics, a 'system' consists of two main components. A space of states, which is a complex *Hilbert space*, so a complete inner product space $\mathcal{H}$ (if $\dim\mathcal{H} < \infty$, it is automatically complete), together with a *Hamiltonian* H, which is a (bounded) linear Hermitian operator on $\mathcal{H}$.

The *wave function* ψ is then a function $\psi\colon \mathbb{R} \to \mathcal{H}\colon t \to |\psi(t)\rangle$ which is a solution to the Schrödinger equation

$$i\hbar\frac{\partial}{\partial t}|\psi(t)\rangle = H|\psi(t)\rangle\,. \tag{3.33}$$

The eigenvalues of H are the energy levels. They are real, because H is Hermitian, and there is an eigenbasis. The wave function describes the time-evolution of the state of the system in the sense that the probability of measuring the system in a state $|v\rangle \in \mathcal{H}$ at time t is equal to $|\langle v|\psi(t)\rangle|^2$. (For us to be able to measure it, $|v\rangle$ must be an eigenstate of another operator which we use for the measurement, but that is irrelevant for now.)

The most important part of this for us is that the physical information, i.e. the possible measurements and their probabilities, only depend on ψ up to a complex phase. If the system has some symmetry – which we would then like to understand via representation theory – the wave function may still transform under this symmetry by such a phase. This is the physics motivation for projective representations.

Definition 3.3.19. Let G be a group. A *projective representation* of G is pair $(V, \tilde{\rho})$ of a complex vector space V and a homomorphism $\tilde{\rho}\colon G \mapsto \mathrm{PGL}(V) = \mathrm{GL}(V)/\mathbb{C}^\times$, where $\mathbb{C}^\times$ should be interpreted as the subgroup of non-zero multiples of the identity matrix.

In less fancy language, a projective representation is a collection of matrices $\{\rho(g) \in \mathrm{GL}(V)\,|g \in G\}$ such that there exist some $c(g,h) \in \mathbb{C}^\times$ with

$$\rho(g)\rho(h) = c(g,h)\rho(gh)\,. \tag{3.34}$$

If we want to make the contrast with usual representations clear, we call the latter *linear representations*.

If we also require that this projective representation be unitary – as is required in quantum physics –, we are dealing with a map $\tilde{\rho}\colon G \to \mathrm{U}(V)/\mathrm{U}(1)$, or equivalently, we want all $\rho(g) \in \mathrm{U}(V)$ and $c(g,h) \in \mathrm{U}(1)$. As $\mathrm{U}(1)$ is the group of complex phases, this does fit with our motivation.

This tells us that, from a quantum point of view, if we want to understand rotationally symmetric systems, we should study projective representations of $\mathrm{SO}(3)$. But it turns out that these are exactly the representations of $\mathrm{SU}(2)$!

In the case that the representation is finite-dimensional, this is because we can look at $d\tilde{\rho}$, and then every coset of $\mathfrak{u}(1)$ contains exactly one traceless representative. If we choose $d\rho$ to always be this traceless representative, this gives an actual representation of $\mathfrak{so}(3)$, so a representation of $\mathrm{SU}(2)$. However, this need not always descend to $\mathrm{SO}(3)$. This is a particular case of the following result.

Proposition 3.3.20. *Let G be a Lie group and $(V, \tilde{\rho})$ a finite-dimensional projective representation of G. Then there exists a linear representation ρ of the universal cover $\hat{G}$ of G which lifts $\tilde{\rho}$.*

The point of this proposition is that we may go to the Lie algebra $\mathrm{Lie}(G)$, there only look at traceless matrices, and then back to the simply connected Lie group related to that. This simply connected Lie group is exactly the 'universal cover' of G. In our case of interest, $G = \mathrm{SO}(3)$, and $\hat{G} = \mathrm{SU}(2)$.

In the case the representation is infinite-dimensional, which happens a lot in physics, we need something more.

Theorem 3.3.21 (Bargmann's theorem). *Let G be a Lie group such that $H^2(\mathrm{Lie}(G);\mathbb{R}) = 0$ and $(V, \tilde{\rho})$ an infinite-dimensional projective unitary representation of G. Then there exists a linear representation ρ of the universal cover $\hat{G}$ of G which lifts $\tilde{\rho}$.*

I will not explain what $H^2(\mathfrak{g};\mathbb{R})$ is for a Lie algebra $\mathfrak{g}$, but just mention that it is zero for all semisimple Lie algebras (we will introduce these in the next section, $\mathfrak{su}(n)$ is included), as well as the algebra of the Poincaré group, the group of symmetries of Minkowski space-time.

If we restrict our attention to $\mathrm{SO}(p,q)$, the special orthogonal group with respect to a pseudo-Hermitian form of signature (p,q), these are actually very nearly simply connected. In fact, they all have a universal double cover $\mathrm{Spin}(p,q) \to \mathrm{SO}(p,q)$ which is called the *spin group*. In the particular case of $\mathrm{SO}(3)$, we have seen in example 3.3.12 that $\mathrm{Spin}(3) \cong \mathrm{SU}(2)$. This has the consequence that the $c(p,q)$ of definition 3.3.19 must have values in $\{\pm 1\}$. We call this kind of projective representation *spin representations*.

Then objects which transform as linear representations for $\mathrm{SO}(p,q)$ are called *tensors*, while those that transform as a spin representations are called *spinors*.

Again going back to $\mathrm{SO}(3)$, we see that the representations of $\mathfrak{su}(2)$ of half-integer spin are spinor representations, while those of integer spin are tensor representations. We also have special names for the cases $j = 0$ and $j = 1$: objects transforming like those are called *scalars* and *vectors*, respectively.

As a last note, the extension to $\mathrm{SO}(p,q)$ is physically meaningful. In particular, when moving to relativistic quantum dynamics, the right group is $\mathrm{SO}(1,3)$. This is the (proper) Lorentz group, the group of Lorentz transformations. The Dirac equation, which is a relativistic version of the Schrödinger equation, is generally written in terms of a specific 4-dimensional representation of $\mathrm{Spin}(1,3)$. This dimension was required to represent the group in the right way (also preserving parity), and this caused Dirac to predict the existence of anti-matter, one of the greatest predictions of theoretical physics in the previous century.

3.4 — Semisimple Lie algebras

In this section, we will consider a special class of Lie algebras: the ones that are reductive. One of the reasons this class is interesting, is representation-theoretic: we have seen that any representation of a finite group decomposes into irreducibles, i.e. is semisimple. The reductive Lie algebras are exactly those for which a similar result holds.

Reductive Lie algebras are naturally decomposed in two parts: an abelian part and a semisimple part. The abelian part is easily described, so we will mostly focus on the semisimple part.

Translating this back to Lie groups, we will see that this theory is intimately related to compact Lie groups. We have already seen that many results for finite groups have direct analogues for compact Lie groups, except that there may be infinitely many irreducible representations. In this section, we will give a classification of the connected compact Lie groups.

We will first focus on Lie algebras. For now, we assume all Lie algebras and representations are over $\mathbb{C}$.

Definition 3.4.1. An *ideal* I of a Lie algebra $\mathfrak{g}$ is a subrepresentation of its adjoint representation. Explicitly, it is a subspace $I \subseteq \mathfrak{g}$ such that for all $x \in \mathfrak{g}$ and $y \in I$, $[x,y] \in I$.

A Lie algebra $\mathfrak{g}$ is *simple* if it is not abelian and its only ideals are 0 and $\mathfrak{g}$ itself.

A Lie algebra is *semisimple* if it is a direct sum of simple Lie algebras.

A Lie algebra is *reductive* if it is the direct sum of a semisimple Lie algebra and an abelian Lie algebra.

Ideals are very analogous to normal subgroups. Both are a stronger version of subobject (sub-Lie algebra resp. subgroup) that are closed under the action of the object on itself (adjoint action resp. conjugation action).

We exclude the abelian case because it behaves differently with representations: simple Lie algebras can be studied via their adjoint representations, which is very interesting as $[\mathfrak{g},\mathfrak{g}] = \mathfrak{g}$ for $\mathfrak{g}$ simple. On the other hand, for an abelian Lie algebra $\mathfrak{a}$, $[\mathfrak{a},\mathfrak{a}] = 0$. Of course, abelian Lie algebras are not very hard to understand either, so we can add them without harm. This is why reductive Lie algebras are well-behaved.

Example 3.4.2. Many Lie algebras we have encountered are simple: these include $\mathfrak{sl}(n,\mathbb{C})$, $\mathfrak{so}(n,\mathbb{C})$, and $\mathfrak{sp}(n,\mathbb{C})$.

By definition, all (semi)simple or abelian Lie algebras are reductive. A more interesting example is $\mathfrak{gl}(V)$, which is the sum of $\mathfrak{sl}(V)$ and the subspace of scalar multiples of the identity. ◇

EXAMPLE 3.4.3. In the exercises, we have seen that $\mathfrak{t}(3, \mathbb{R})$ is not (semi)simple: $\mathfrak{n}(3, \mathbb{R})$ is a non-trivial ideal which has no complementary ideal. This also works for $\mathfrak{t}(n, \mathbb{R})$ or $\mathfrak{t}(n, \mathbb{C})$.

In fact, these examples are called soluble or solvable, which means that for some k, any k-fold Lie bracket will be zero identically. We will not go into this further, but the idea is that solvable Lie algebras can be studied by their non-trivial ideals.

Think of this as analogous to the Jordan decomposition, if you have seen that: any matrix can be decomposed into a semisimple part (i.e. diagonalisable) and a nilpotent part, and this decomposition is unique. Solvable Lie algebras correspond to this nilpotent part. ◇

The study of semisimple Lie algebras mimics section 3.3.3, where we constructed irreducible representations of $\mathfrak{sl}(2, \mathbb{C})$ by diagonalising one of the basis elements and constructing a representation by hand using ladder operators. This is called a highest weight construction.

In $\mathfrak{sl}(2, \mathbb{C})$, we could only diagonalise one basis element at a time. But in general, we want to diagonalise as many as we can.

DEFINITION 3.4.4. Let $\mathfrak{g}$ be a semisimple Lie algebra. A *Cartan subalgebra* $\mathfrak{h} \subset \mathfrak{g}$ is a maximal abelian subalgebra of diagonalisable elements.

The *rank* of $\mathfrak{g}$, written as $\operatorname{rk} \mathfrak{g}$, is the dimension of such a Cartan subalgebra (this is independent of the choice).

EXAMPLE 3.4.5. For $\mathfrak{sl}(m, \mathbb{C})$, we can take $\mathfrak{h}$ to be the space of diagonal matrices. This has dimension $m - 1$ (as we impose that the trace be zero), so this is the rank of $\mathfrak{sl}(m, \mathbb{C})$. Note that the rank is different from both the dimension of $\mathfrak{sl}(m, \mathbb{C})$, which is $m^2 - 1$, and the dimension of its defining representation, which is m.

In this context, we often write $A_n = \mathfrak{sl}(n + 1, \mathbb{C})$. ◇

EXAMPLE 3.4.6. For the orthogonal Lie algebras, $\mathfrak{so}(m, \mathbb{C})$, the Lie algebra of m by m antisymmetric matrices, we have to distinguish between m being even or odd.

If $m = 2n$ is even, we can take $\mathfrak{h}$ to be spanned by $\{E_{2i,2i+1} - E_{2i+1,2i} \,|\, 1 \leq n\}$, showing that $\operatorname{rk} \mathfrak{so}(2n, \mathbb{C}) = n$. This is often denoted $D_n = \mathfrak{so}(2n, \mathbb{C})$. (Here $E_{j,k}$ is the matrix with one non-zero element a 1 at position (j, k).)

For $m = 2n + 1$ odd, the same set is still maximal, so $\operatorname{rk} \mathfrak{so}(2n + 1, \mathbb{C}) = n$ as well. This is denoted $B_n = \mathfrak{so}(2n + 1, \mathbb{C})$. ◇

EXAMPLE 3.4.7. The symplectic Lie algebras, $\mathfrak{sp}(2n, \mathbb{C})$, are the Lie algebras of $\mathrm{Sp}(2n, \mathbb{C})$, introduced in example 1.2.6. Concretely, they consist of those $2n$ by $2n$ matrices A such that $A^T \Omega = -\Omega A$, where

$$\Omega = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} . \tag{3.35}$$

A Cartan subalgebra is given by all diagonal matrices, and spanned by $\{E_{j,j} - E_{n+j,n+j} \,|\, 1 \leq j \leq n\}$. So $\operatorname{rk} \mathfrak{sp}(2n, \mathbb{C}) = n$. This is often denoted $C_n = \mathfrak{sp}(2n, \mathbb{C})$. ◇

Now, if we take any representation (V, r) of a simple matrix Lie algebra $\mathfrak{g}$, we can find a basis of common eigenvectors $\{v_i\}$ for all of $\mathfrak{h}$, i.e. for any $h \in \mathfrak{h}$, there are complex numbers $a_i(h)$ such that $hv_i = a_i(h)v_i$. These a_i are then elements of the dual space $\mathfrak{h}^\vee$, and they are called the *weights* of the representation ($\mathfrak{h}^\vee$ is called the *weight space*).

In particular, we may do this for the adjoint representation. Its non-zero weights are called *roots*, and we will denote them by α. The collection of roots is denoted Φ, and called a *root system*. This gives a decomposition

$$\mathfrak{g} = \mathfrak{g}_0 \oplus \bigoplus_{\alpha \in \Phi} \mathfrak{g}_\alpha \,, \tag{3.36}$$

where $\mathfrak{g}_\alpha$ is the eigenspace for the root α (and similarly for $\mathfrak{g}_0$).

Example 3.4.8. Let us consider $A_2 = \mathfrak{sl}(3, \mathbb{C})$. By example 3.4.5, we may take as a Cartan subalgebra the subspace of diagonal matrices. Let us pick a nice basis for the algebra.

$$H_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad H_2 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}, \tag{3.37}$$

$$E_1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad E_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad E_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \tag{3.38}$$

$$F_1 = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad F_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \qquad F_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}. \tag{3.39}$$

We see that all of these basis elements are eigenvectors for $\mathfrak{h} = \operatorname{span}\{H_1, H_2\}$, and the roots are

$$\alpha_{E_1} = (2, 0), \qquad \alpha_{E_2} = (1, \sqrt{3}), \qquad \alpha_{E_3} = (-1, \sqrt{3}), \tag{3.40}$$

$$\alpha_{F_1} = (-2, 0), \qquad \alpha_{F_2} = (-1, -\sqrt{3}), \qquad \alpha_{F_3} = (1, -\sqrt{3}), \tag{3.41}$$

◇

where we write $\alpha = (\alpha(H_1), \alpha(H_2))$.

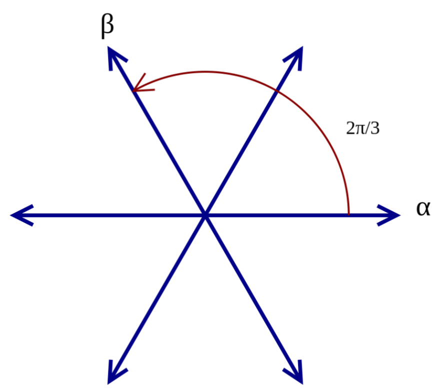

Figure 3.1: The root system of A_2. Image by Wikipedia user Smithers888, taken from https://commons.wikimedia.org/wiki/File:Root_system_A2.svg on March 22nd, 2022.

Now, it turns out that in the case of (semi)simple Lie algebras, there is a lot we can say about this decomposition, and this is the basis of one of the most famous classifications in abstract mathematics.

First of all, it turns out that $\mathfrak{g}_0 = \mathfrak{h}$, and all $\mathfrak{g}_\alpha$ are one-dimensional. Secondly, we may define root systems abstractly and classify those.

Definition 3.4.9. Let E be a finite-dimensional real vector space with inner product $(\cdot, \cdot)$. A subset $\Phi \subset E$ is called a *root system* if

R1 Φ is finite, spans E, and does not contain 0;

R2 If $\alpha \in \Phi$, the only multiples of α in Φ are $\pm\alpha$;

R3 If $\alpha \in \Phi$, the reflection in the hyperplane perpendicular to α preserves Φ. in other words, if $\alpha, \beta \in \Phi$, then

$$\sigma_\alpha(\beta) = \beta - 2\frac{(\beta, \alpha)}{(\alpha, \alpha)}\alpha \in \Phi; \tag{3.42}$$

R4 If $\alpha, \beta \in \Phi$, then $2\frac{(\beta,\alpha)}{(\alpha,\alpha)} \in \mathbb{Z}$.

REMARK 3.4.10. There is a subtlety here: what is this inner product? For a simple Lie algebra $\mathfrak{g}$, there is a canonical bilinear form, called the *Killing form*. It is defined on all of $\mathfrak{g}$ by $\kappa(x, y) = \mathrm{Tr}(\mathrm{ad}(x)\,\mathrm{ad}(y))$.

In case of a matrix Lie algebra, we may also take $\tilde{\kappa}(x, y) = \mathrm{Tr}(xy)$, and this is the same as the Killing form up to scaling.

In either case, when we restrict it to $\mathfrak{h}$, it is non-degenerate, and we may take a real slice on which it is positive-definite, to get an inner product. Then we take the dual inner product on $\mathfrak{h}_{\mathbb{R}}^{\vee}$. From the definition, we see that rescaling the inner product does not change anything.

In example 3.4.8, we chose H_1 and H_2 such that they are real and orthogonal, and with the same norm $\tilde{\kappa}(H_1, H_1) = \tilde{\kappa}(H_2, H_2) = 2$, so we can interpret this root system as lying in $\mathbb{R}^2$ with the standard inner product.

We can construct new root systems by adding them: if we have $\Phi_1 \subset E_1$ and $\Phi_2 \subset E_2$, then taking the direct sum $E_1 \oplus E_2$, where the terms are mutually orthogonal, $\Phi_1 \cup \Phi_2 \subset E_1 \oplus E_2$ is also a root system.

DEFINITION 3.4.11. A root system Φ is *reducible* if it can be decomposed as $\Phi = \Phi_1 \cup \Phi_2$, both non-empty, such that $(\alpha, \beta) = 0$ for all $\alpha \in \Phi_1$ and $\beta \in \Phi_2$.

A root system that is not reducible is called *irreducible*.

The point of this definition is two-fold: firstly, we can classify them, and secondly, there is a correspondence between simple Lie algebras and irreducible root systems.

From R4, it follows that, if the angle between roots α and β is ϑ, then $4\cos^2\vartheta = 4\frac{(\beta,\alpha)^2}{(\alpha,\alpha)(\beta,\beta)} \in \mathbb{Z}$. This severely limits the possible angles between and ratios of lengths of roots. Especially considering that by R2, $4\cos^2\vartheta = 4$ only occurs for $\beta = \pm\alpha$.

DEFINITION 3.4.12. Let $\Phi \subset E$ be a root system. A *base* Δ for Φ is subset of Φ which is a basis of E, and such that all roots are an integer linear combination of base elements, with either all coefficients positive, or all negative.

Elements of a base are called *simple roots*.

Bases always exist. For an example, see figure 3.1, where $\alpha = \alpha_{E_1}$ and $\beta = \alpha_{E_3}$ are the simple roots.

For any base Δ and simple roots $\alpha, \beta \in \Delta$, the angle between α and β must be at least $\frac{\pi}{2}$, i.e. we need $(\alpha, \beta) \leq 0$. For otherwise, $\sigma_\alpha(\beta) = \beta - 2\frac{(\beta,\alpha)}{(\alpha,\alpha)}\alpha$ would be a root with both positive and negative coefficients in the base.

DEFINITION 3.4.13. Let $\Phi \subset E$ be a root system, and Δ any base for Φ. Its *Dynkin diagram* is comprised of the following data:

- For each simple root, a node;
- Between the nodes for distinct roots α, β, a total of $4\frac{(\beta,\alpha)^2}{(\alpha,\alpha)(\beta,\beta)}$ edges;
- If two roots α and β are connected by multiple edges, an arrow pointing towards the shorter root.

These diagrams do not depend on the choice of base, and the root system can be recovered from the Dynkin diagram.

THEOREM 3.4.14. *If Φ is an irreducible root system, its Dynkin diagram is one of the diagrams listed in figure 3.2, i.e. it is one of*

- *A_n for $n \geq 1$;*
- *B_n for $n \geq 2$;*

- *C_n for $n \geq 3$;*
- *D_n for $n \geq 4$;*
- *E_6, E_7, E_8, F_4, or G_2.*

Each Dynkin diagram corresponds to one root system and Lie algebra, up to equivalence. In particular, the Lie algebras A_n, B_n, C_n, and D_n of examples 3.4.5 to 3.4.7 correspond to the Dynkin diagrams with the same labels. These are called the classical Lie algebras*.*

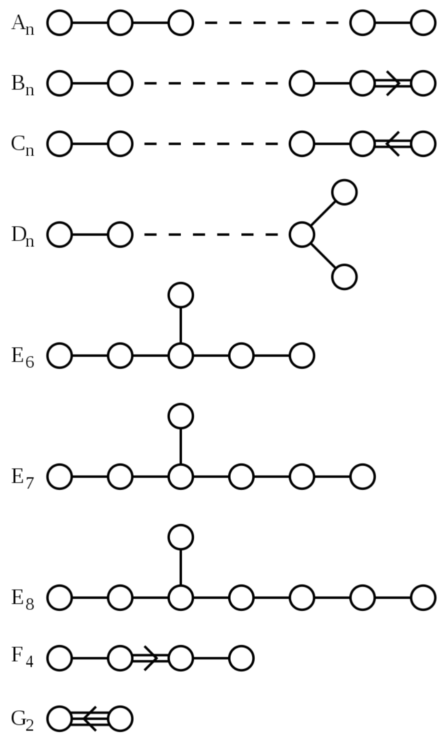

Figure 3.2: The list of Dynkin diagrams. Image by Wikipedia user R. A. Nonenmacher, taken from https://commons.wikimedia.org/wiki/File:Connected_Dynkin_Diagrams.svg on March 22nd, 2022.

The exceptional Lie algebras in this list, E_6, E_7, E_8, F_4, and G_2, can also be realised as explicit matrix Lie algebras (i.e. they have a faithful representation), but we will not give them here. There are actual physical theories using them: e.g. E_8 is used in a model of string theory called 'heterotic string theory' (although to what degree string theory is physics is up for debate).

The restrictions of the indices in the list are to avoid redundancies. Staring at the diagrams for a while, you may convince yourself that $A_1 = B_1 = C_1 = D_1$, $C_2 = B_2$, $D_2 = A_1 \times A_1$, and $D_3 = A_3$. Similarly, the exceptional cases can also be defined for lower indices, but then $E_1 = F_1 = G_1 = A_1$, $E_2 = A_1 \times A_1$, $F_2 = A_2$, $E_3 = A_2 \times A_1$, $F_3 = B_3$, $E_4 = A_4$, and $E_5 = D_5$. All of these identifications lead to isomorphisms on the level of Lie algebras, e.g. $\mathfrak{so}(4) \cong \mathfrak{su}(2) \times \mathfrak{su}(2)$.

This classification occurs in many different areas of abstract mathematics as well. Sometimes, a classification only contains the diagrams of type A_n, D_n, and E_6, E_7, E_8, because they have only single edges. We call these diagrams *simply laced* and such a classification an *ADE classification*.

EXAMPLE 3.4.15. Let us consider the Dynkin diagram G_2 and try to reconstruct the root system and the Lie algebra.

The Dynkin diagram has two nodes, so the root system (E, Φ) must be two-dimensional, with a base $\Delta = \{\alpha, \beta\}$. Let us take α to be the shorter root. We may choose $E = \mathbb{R}^2$, and $\alpha = (1, 0)$ (so $(\alpha, \alpha) = 1$). From the Dynkin diagram, we know that $4\frac{(\beta,\alpha)^2}{(\beta,\beta)} = 3$, so the angle between α and β must be $\frac{5}{6}\pi$ (as we need it to be obtuse for a base). We also know that $2\frac{(\beta,\alpha)}{(\beta,\beta)} \in \mathbb{Z}$, and that $2(\beta, \alpha) \in \mathbb{Z}$. This tells us that $2\frac{(\beta,\alpha)}{(\beta,\beta)} = -1$ and $2(\beta, \alpha) = -3$. It follows that $(\beta, \beta) = 3$. Without loss of generality, we may take $\beta = (-\frac{3}{2}, \frac{\sqrt{3}}{2})$.

Then, we may fill in the complete root system by reflections. First of all, $-\alpha = \sigma_\alpha(\alpha)$ and $-\beta = \sigma_\beta(\beta)$ must exist. We have found the actual numbers used in reflection of axiom R3, so we find that

- $\sigma_\beta(\alpha) = \alpha + \beta \in \Phi$, so also $-\alpha - \beta \in \Phi$;
- $\sigma_\alpha(\beta) = 3\alpha + \beta \in \Phi$, so also $-3\alpha - \beta \in \Phi$;
- $\sigma_\alpha(\alpha + \beta) = 2\alpha + \beta \in \Phi$, so also $-2\alpha - \beta \in \Phi$;
- $\sigma_\beta(3\alpha + \beta) = 3\alpha + 2\beta \in \Phi$, so also $-3\alpha - 2\beta \in \Phi$.

If we make a drawing of this, we get figure 3.3, from which it is clear that there are no more possible roots without violating the possible angles, or R2.

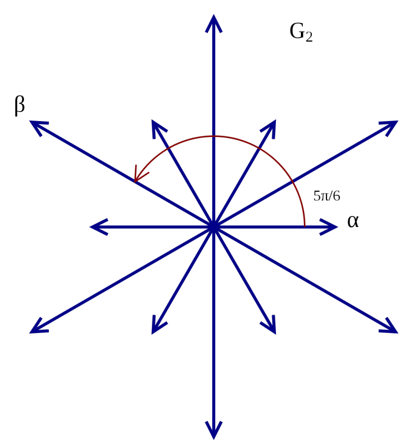

Figure 3.3: The root system of G_2. Image by Wikipedia user Smithers888, taken from https://commons.wikimedia.org/wiki/File:Root_system_G2.svg on March 28th, 2022.

The next step is to construct the Lie algebra $\mathfrak{g}$. From equation (3.36) and the paragraph after example 3.4.8, we know that $\mathfrak{g} = \mathfrak{h} \oplus \bigoplus_{a\in\Phi} \mathfrak{g}_a$, where $\dim \mathfrak{h} = \dim E = 2$, and $\dim \mathfrak{g}_\alpha = 1$ for any $a \in \Phi$. Let us pick a basis for $\mathfrak{h}$, $\{h_1, h_2\}$, dual to the basis of E, i.e. the eigenvalues of h_1 are given by the first components of the roots, and the eigenvalues of h_2 by the second components.

We may also pick a basis element $e_a \in \mathfrak{g}_a$ for all $a \in \Phi$. Then we already know that $[h_i, e_a] = a_i e_a$, by definition of what roots must be, and to construct all of $\mathfrak{g}$, we only need to find the values of $[e_a, e_b]$ for $a, b \in \Phi$. Note that there is still freedom here, as we may scale the e_a.

By the Jacobi identity, we find that if $x \in \mathfrak{g}_a$ and $y \in \mathfrak{g}_b$, then $[x, y] \in \mathfrak{g}_{a+b}$, i.e. the Lie bracket adds roots. In our case, this means that $[e_a, e_b]$ is proportional to e_{a+b} if $a + b \in \Phi$, or lies in $\mathfrak{h}$ if $a + b = 0$, or is zero if else. So we can *define* (because we still had scaling freedom)

$$e_{\alpha+\beta} = [e_\alpha, e_\beta] \qquad e_{-\alpha-\beta} = [e_{-\alpha}, e_{-\beta}] \tag{3.43}$$

$$e_{2\alpha+\beta} = [e_\alpha, e_{\alpha+\beta}] \qquad e_{-2\alpha-\beta} = [e_{-\alpha}, e_{-\alpha-\beta}] \tag{3.44}$$
$$e_{3\alpha+\beta} = [e_\alpha, e_{2\alpha+\beta}] \qquad e_{-3\alpha-\beta} = [e_{-\alpha}, e_{-2\alpha-\beta}] \tag{3.45}$$
$$e_{3\alpha+2\beta} = [e_\beta, e_{3\alpha+\beta}] \qquad e_{-3\alpha-2\beta} = [e_{-\beta}, e_{-3\alpha-\beta}] \tag{3.46}$$

Now, how do we relate positive to negative roots? It turns out that $[e_a, e_{-a}]$ is proportional to $a^T \in \mathfrak{h}$, and we have enough scaling freedom left to choose $[e_\alpha, e_{-\alpha}] = \alpha^T = h_1$ and $[e_\beta, e_{-\beta}] = \beta^T = -\frac{3}{2}h_1 + \frac{\sqrt{3}}{2}h_2$. This is sufficient to calculate any Lie bracket. As an example,

$$\begin{aligned}
[e_{\alpha+\beta}, e_{-\alpha-\beta}] &= \big[[e_\alpha, e_\beta], [e_{-\alpha}, e_{-\beta}]\big] \\
&= -\big[e_{-\alpha}, [e_{-\beta}, [e_\alpha, e_\beta]]\big] - \big[e_{-\beta}, [[e_\alpha, e_\beta], e_{-\alpha}]\big] \\
&= \big[e_{-\alpha}, [e_\alpha, [e_\beta, e_{-\beta}]]\big] + \big[e_{-\alpha}, [e_\beta, [e_{-\beta}, e_\alpha]]\big] \\
&\quad + \big[e_{-\beta}, [[e_\beta, e_{-\alpha}], e_\alpha]\big] + \big[e_{-\beta}, [[e_{-\alpha}, e_\alpha], e_\beta]\big] \\
&= \big[e_{-\alpha}, [e_\alpha, \beta^T]\big] + 0 + 0 - \big[e_{-\beta}, [\alpha^T, e_\beta]\big] \\
&= -(\beta, \alpha)[e_{-\alpha}, e_\alpha] - (\alpha, \beta)[e_{-\beta}, e_\beta] \\
&= (\alpha, \beta)(\alpha^T + \beta^T) = -\frac{3}{2}\big(h_1 + -\frac{3}{2}h_1 + \frac{\sqrt{3}}{2}h_2\big) \\
&= \frac{3}{4}h_1 - \frac{3\sqrt{3}}{4}h_2\,.
\end{aligned} \tag{3.47}$$

◇

3.4.1 — Compact Lie groups

Now it is time so translate the above story back to Lie groups. In particular, we want to relate back to compact Lie groups.

We have seen above that abelian Lie algebras are essential in understanding the classification: we had the abelian component in the reductive algebra, and in a semisimple algebra, we also looked for a maximal abelian subalgebra, the Cartan subalgebra. So let us first consider abelian Lie algebras.

A real abelian Lie algebra $\mathfrak{a}$ is isomorphic to $\mathbb{R}^n$ with trivial Lie bracket. Its simply connected Lie group is again $\mathbb{R}^n$. But this is not compact. The compact Lie groups whose Lie algebra is abelian must be quotients of $\mathbb{R}^n$, and it turns out that this means the group is a torus group: a product of circle groups.

Considering any compact connected Lie group, we may find a maximal torus subgroup, in the same way we found a maximal abelian subalgebra in a semisimple Lie algebra. It turns out that the same constructions go through, and we get a very close correspondence.

Theorem 3.4.16. *Let G be a compact Lie group. Then $\mathrm{Lie}(G)_\mathbb{C}$ is a reductive Lie algebra.*

Conversely, every semisimple complex Lie algebra $\mathfrak{g}$ has a unique compact real form, i.e. a unique real sub-Lie algebra $\mathfrak{k}$ such that $\mathfrak{g} = \mathfrak{k}_\mathbb{C}$, and the simply connected Lie group of $\mathfrak{k}$ is compact.

The compact real forms of the classical Lie algebras are

- $A_n = \mathrm{SU}(n+1)$*;*
- $B_n = \mathrm{Spin}(2n+1)$*;*
- $C_n = \mathrm{Sp}(2n)$*;*
- $D_n = \mathrm{Spin}(2n)$.

Therefore, any connected compact Lie group is a quotient of the product of a torus and compact real forms as above by a finite central subgroup.

There do exist other real forms of semisimple algebras. For example, considering $\mathfrak{sl}(n,\mathbb{C})$, the compact real form is $\mathfrak{su}(n)$, but a non-compact real form could be $\mathfrak{sl}(n,\mathbb{R})$. For another example, the real forms $\mathfrak{so}(p,n-p)$ of $\mathfrak{so}(n,\mathbb{C})$ are non-compact unless $p \in \{0,n\}$. The classification of all real forms, the semisimple real Lie algebras, is a bit more involved. Look up "Satake diagram" if you are interested.

EXAMPLE 3.4.17. We have already seen that $\mathrm{SO}(n) = \mathrm{Spin}(n)/\{\pm 1\}$. In fact, this was the definition of $\mathrm{Spin}(n)$. ◇

3.5 — Representations of the Lorentz and Poincaré groups

At the end of section 3.3.4, we noted that $\mathrm{SO}(1,3)$ is the (proper) Lorentz group, the group of Lorentz transformations, and that this is the point symmetry group of Minkowski spacetime. This group is therefore worth studying, and we will look at it here.

Minkowski spacetime has more symmetries though: it is not only Lorentz invariant, but also translation invariant. The larger symmetry group containing Lorentz transformations and spacetime translations, is called the Poincaré group. We will also give some idea of its representations here.

The Lorentz group $\mathrm{SO}(1,3)$ is not compact, as one-parameter subgroups of Lorentz boosts from hyperbolas and not circles. However, it is close to a compact Lie group: as noted below theorem 3.4.16, $\mathfrak{so}(1,3)$ is a non-compact real form of $\mathfrak{so}(4,\mathbb{C})$: $\mathfrak{so}(1,3)_{\mathbb{C}} \cong \mathfrak{so}(4,\mathbb{C})$. So its representation theory is still semisimple: any representation is the sum of irreducible representations. But beware: these irreducible representations are not unitary! (Or rather, they are unitary with respect to the indefinite metric.)

With the classification of semisimple Lie groups, theorem 3.4.14, we actually know that $\mathfrak{so}(4,\mathbb{C}) = D_2 = A_1 \times A_1 = \mathfrak{sl}(2,\mathbb{C}) \oplus \mathfrak{sl}(2,\mathbb{C})$. We can do even better though, by taking an actual explicit description of $\mathrm{SO}(1,3)$.

$$\mathrm{SO}(3,1) = \{A \in \mathrm{GL}(4,\mathbb{R}) \mid A^T \eta A = \eta\}\,, \qquad \text{where} \quad \eta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}\,. \tag{3.48}$$

We then get that

$$\mathfrak{so}(1,3) = \{X \in \mathfrak{gl}(4,\mathbb{R}) \,|\, A^T\eta = -A\eta\}\,, \tag{3.49}$$

for which we may write an explicit basis, in the physics convention:

$$\begin{aligned} J_1 &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \end{pmatrix}, \quad J_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & i \\ 0 & 0 & 0 & 0 \\ 0 & -i & 0 & 0 \end{pmatrix}, \quad J_3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -i & 0 \\ 0 & i & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \\ K_1 &= \begin{pmatrix} 0 & i & 0 & 0 \\ i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad K_2 = \begin{pmatrix} 0 & 0 & i & 0 \\ 0 & 0 & 0 & 0 \\ i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad K_3 = \begin{pmatrix} 0 & 0 & 0 & i \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ i & 0 & 0 & 0 \end{pmatrix}, \end{aligned} \tag{3.50}$$

such that

$$[J_l, J_m] = i \sum_{n=1}^{3} \varepsilon_{lmn} J_n\,, \tag{3.51}$$

$$[J_l, K_m] = i\sum_{n=1}^{3} \varepsilon_{lmn} K_n\,, \tag{3.52}$$

$$[K_l, K_m] = -i\sum_{n=1}^{3} \varepsilon_{lmn} J_n\,. \tag{3.53}$$

The J_l are just the L_l of example 3.2.14 with extra zeroes in the zeroth row and column: these are the rotations. The K_l are the boosts. The sign in equation (3.53) is the main difference between $\mathrm{SO}(1,3)$ and $\mathrm{SO}(4)$.

There is a trick to understand this better: we combine these generators as $Y_l = \frac{1}{2}(J_l + iK_l)$, then

$$[Y_l, Y_m] = i\sum_{n=1}^{3} \varepsilon_{lmn} Y_n\,. \tag{3.54}$$

As this is the same Lie bracket as for $\mathfrak{so}(3)_{\mathbb{C}} \cong \mathfrak{su}(2)_{\mathbb{C}} \cong \mathfrak{sl}(2,\mathbb{C})$, cf. examples 3.2.14 and 3.2.15, this tells us that $\{J_l, K_l\}_{l=1,2,3}$ is actually a real basis for $\mathfrak{sl}(2,\mathbb{C})$. So $\mathfrak{so}(1,3) \cong \mathfrak{sl}(2,\mathbb{C})$ as *real* Lie algebras. As $\mathrm{SL}(2,\mathbb{C})$ is simply connected, we have found that $\mathrm{Spin}(1,3) \cong \mathrm{SL}(2,\mathbb{C})$ is the double cover of $\mathrm{SO}(1,3)$.

What does all this mean for representation theory? On the level of complex Lie algebras, we are looking at $\mathfrak{sl}(2,\mathbb{C}) \oplus \mathfrak{sl}(2,\mathbb{C})$, and its irreducible representations are given by tensor products of two irreducible representations of $\mathfrak{sl}(2,\mathbb{C})$, cf. the last exercise sheet. Therefore, they are indexed by pairs (j_1, j_2), where $j_1, j_2 \in \{0, \frac{1}{2}, 1, \frac{3}{2}, \dots\}$. These all give representations of $\mathrm{SL}(2,\mathbb{C})$ and for $\mathrm{SO}(1,3)$ we get in a way similar to proposition 3.3.16, we get

Proposition 3.5.1. *The irreducible projective representations of* $\mathrm{SO}(1,3)$ *are indexed by pairs* (j_1, j_2), *where* $j_1, j_2 \in \{0, \frac{1}{2}, 1, \frac{3}{2}, \dots\}$, *with dimension* $(2j_1+1)(2j_2+1)$. *They are linear representations if* $j_1 + j_2 \in \mathbb{Z}$ *and spin otherwise.*

Now, let us change our focus to the Poincaré group. This contains not only the rotations J_l and boosts K_l, but also translations P_μ. You may think of this as somewhat analogous to the affine group introduced in example 3.1.16, only this time we have four dimensions and we require orthogonality with respect to η.

If we write $M^{\mu\nu}$ for the matrix with an i on position (μ,ν), a $-i$ on position (ν,μ), and zeroes elsewhere, then

$$J_l = \frac{1}{2}\sum_{m,n=1}^{3} \varepsilon_{lmn} M^{\mu\nu}\,, \quad K_l = M_{l4}\,, \tag{3.55}$$

where we use η to raise and lower indices. In this notation, the Lie bracket is given by

$$[P_\mu, P_\nu] = 0\,, \tag{3.56}$$

$$[M_{\mu\nu}, P_\rho] = i(\eta_{\mu\rho}P_\nu - \eta_{\nu\rho}P_\mu)\,, \tag{3.57}$$

$$[M_{\mu\nu}, M_{\rho\sigma}] = i(\eta_{\mu\rho}M_{\nu\sigma} - \eta_{\mu\sigma}M_{\nu\rho} - \eta_{\nu\rho}M_{\mu\sigma} + \eta_{\nu\sigma}M_{\mu\rho})\,. \tag{3.58}$$

This group is non-compact, so its representation theory is more complicated than what we are used to. Still, the unitary irreducible representations (which are infinite-dimensional or trivial) have been classified, and this is called *Wigner's classification*. More precisely, it is a classification of the physical representations; we will see what this means in a moment.

The idea of the classification is to first look at a subgroup, and then extend representations to the entire group. This is known as *inducing representations* in mathematics or the theory of *little groups* in physics. We start with the subgroup of translations, or on the Lie algebra side the abelian subalgebra generated by the P_μ. As this is an abelian Lie algebra, we may simulateneously diagonalise *all* matrices (so the entire thing is Cartan) and pick an eigenvector:

$$P_\mu|p\rangle = p_\mu|p\rangle\,. \tag{3.59}$$

We note here that the Poincaré group has two Casimirs (which were combinations of the generators that are diagonal in any representation). One is $P^2 = \sum_{\mu=0}^{3} P_\mu P^\mu$, whose eigenvalue is m^2, so it gives the mass of the representation/particle. The other is $W^2 = \sum_{\mu=0}^{3} W_\mu W^\mu$, where

$$W_\mu = -\frac{1}{2} \sum_{\nu,\rho,\sigma=0}^{3} \epsilon_{\mu\nu\rho\sigma} M^{\nu\rho} P^\sigma \tag{3.60}$$

is the *Pauli-Lubanski pseudo-vector*. We may use their eigenvalues to label representation.

The actual "little group" we start with is the subgroup of the Poincaré group that preserves this set of eigenvectors p_μ, i.e. comprising those elements g such that $P_\mu g|p\rangle = p_\mu g|p\rangle$. This little group depends on what p_μ actually is:

1. If $p_\mu = 0$, we call this the *vacuum*;
2. If $p_\mu \neq 0$, but has zero mass, then we may change to the frame of reference such that $p_\mu = (p, p, 0, 0)$, with $p > 0$;
3. If $m > 0$, then we may go to the rest frame and have $p_\mu = (m, 0, 0, 0)$.

In the first case, the little group is the entire group. But this means that the representation is unitary and finite-dimensional for the entire group, so it must be trivial.

In the massive third case, the little group is the subgroup SO(3) of rotations, whose irreducible unitary projective representations we know: they are the spin representations s_j. So a massive particle is classified by its mass and its spin.

In the massless second case, the little group is the double cover of SE(2), the group of rotations and translations in two dimensions. Its Lie algebra has a basis $\{J, P_1, P_2\}$, with Lie bracket

$$[J, P_1] = iP_2\,, \quad [J, P_2] = -iP_1\,, \quad [P_1, P_2] = 0\,. \tag{3.61}$$

Then we may use induced representations again on the subgroup generated by translations, and we get an eigenstate

$$P_j|k\rangle = k_j|k\rangle\,, j = 1, 2\,. \tag{3.62}$$

In case $k = 0$, the little group is $\mathrm{SO}(2) \cong \mathrm{U}(1)$, and the representations are indexed by a half integer (remember we took a double cover of SE(2)) called the *helicity* h. (This is equal to the chirality for massless particles, although the two differ for massive ones.)

In case $k \neq 0$, the little group is $\mathrm{SO}(1) \cong \{e\}$, and this gives 'continuous spin' representations, which seem to be unphysical.

Theorem 3.5.2 (Wigner's classification). *Physical unitary irreducible representations of the Poincaré group are one of the following:*

1. *The trivial, vacuum, representation;*
2. *The zero mass representations, indexed by helicity* $h \in \{0, \frac{1}{2}, 1, \frac{3}{2}, \dots\}$;
3. *The positive mass representations, indexed by mass* $m \in \mathbb{R}_{>0}$ *and spin* $s \in \{0, \frac{1}{2}, 1, \frac{3}{2}, \dots, \}$.

Note that these representations (apart from the vacuum) are infinite-dimensional: a free particle may be decribed by e.g. its position and momentum at a certain time, and these give continuous labels of the basis vector in the irreducible representation associated to that particle.

3.6 — Particle physics

This section is non-examinable. Much of the material is adapted from the Wikipedia page on the Standard Model and the relevant page of Vincent Bouchard's note for the 2020 edition of this course.

In the previous section, we found that certain well-known properties of elementary particles, namely mass and spin/helicity, arise from the representation they transform under when rotating or translating in Minkowski spacetime. Such invariant numbers are called *quantum numbers*. But there are other quantum numbers, such as the charges of particles for certain forces. These quantum numbers of elementary particles are captured by internal symmetries of the the system that do not affect the physics. This is called *gauge theory*, and the internal symmetries form the *gauge group* of the theory.

3.6.1 — The Standard Model

The most famous, and succesful, theory of particle physics is called the Standard Model. This is a particular example of a gauge theory, whose gauge group is $\mathrm{SU}(3) \times \mathrm{SU}(2) \times \mathrm{U}(1)$. Here, the $\mathrm{SU}(3)$ factor represents the strong force and the $\mathrm{SU}(2) \times \mathrm{U}(1)$ factor represents the electroweak force. (Although electromagnetism on its own is a $\mathrm{U}(1)$ gauge theory, it turns out we cannot make sense of the weak force as $\mathrm{SU}(2)$ gauge theory, the two need to be combined.)

This group is not semisimple, due to the $\mathrm{U}(1)$ factor, but it is reductive. For the $\mathrm{U}(1)$, the irreducible representations are indexed by an integer, cf example 3.3.6, but the correct notion of projective representation here is a representation of a finite cover, which is indexed by a rational number. This rational number is called the *weak hypercharge* Y in our current situation.

With this in mind, we can classify elementary particles according to their irreducible representation of $\mathrm{SU}(3) \times \mathrm{SU}(2) \times \mathrm{U}(1)$. These are conventionally given as $(\mathbf{p}, \mathbf{q})_n$, where $\mathbf{p}$ is the p-dimensional irreducible representation of $\mathrm{SU}(3)$, $\mathbf{q}$ is the q-dimensional irreducible representation of $\mathrm{SU}(2)$, and n is the weak hypercharge. The table is then

Particle	Name/gauge	Representation
Spin 1		
B	weak hypercharge	$(\mathbf{1}, \mathbf{1})_0$
W	weak isospin	$(\mathbf{1}, \mathbf{3})_0$
g	gluon/colour	$(\mathbf{8}, \mathbf{1})_0$
Spin $\frac{1}{2}$	(three generations)	
q_L	left-handed quark	$(\mathbf{3}, \mathbf{2})_{\frac{1}{3}}$
u_R	right-handed up-quark	$(\mathbf{3}, \mathbf{1})_{\frac{4}{3}}$
d_R	right-handed down-quark	$(\mathbf{3}, \mathbf{1})_{-\frac{2}{3}}$
ℓ_L	left-handed lepton	$(\mathbf{1}, \mathbf{2})_{-1}$
ℓ_R	right-handed electron	$(\mathbf{1}, \mathbf{1})_{-2}$
Spin 0		
H	Higgs	$(\mathbf{1}, \mathbf{2})_1$

Some notes about this:

- The gauge bosons (spin 1 particles) are the force carriers. They transform according to the adjoint representation of their corresponding group. Informally, they *are* the (Lie algebra of) the Lie group.
- The matter sector fermions (spin $\frac{1}{2}$ particles) all transform according to trivial or defining representations of $\mathrm{SU}(3)$ and $\mathrm{SU}(2)$. This is why it makes sense to say they do (or do not) have a charge.

- This picture is incomplete, as it does not allow for massive neutrinos. To accommodate for them, we should probably add a right-handed neutrino, which should transform as $(\mathbf{1},\mathbf{1})_0$, the trivial representation. But these have not been observed yet, and this would be very hard as they do not interact with any force except gravity.
- The matter sector is tripled for three generations, and then doubled again for antiparticles. An antiparticle transforms as the dual, or conjugate, representation of the usual particle.
- The gauge group does not commute with the Poincaré group, so mass and the charges cannot be measured simultaneously. The mass eigentates are what we are more familiar with as particles. For example, the left-handed lepton doublet ℓ_L has two mass eigenstates, which are the left-handed electron e_L and the left-handed electron neutrino $\nu_{e,L}$.
- Furthermore, the Higgs has a non-zero vacuum expectation value, and as such causes spontaneous symmetry breaking at low energies (i.e. close to the vacuum) from $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ to $\mathrm{U}(1)_{\mathrm{em}}$. I.e., the observed preserved quantity is the electro-magnetic charge, which is a combination of weak hypercharge Y and isospin T_3. The resulting gauge boson is the photon γ, while the remaining three mass eigenstates are the $W^\pm$ bosons and the Z boson.

From a representation-theoretic point of view, you may ask the question: why do these representations appear in nature, and no others? You would not be the first to ask! It could be that this is just the way it is, and there is no deeper reason. But that would not be very satisfying, so could we find a better reason? And would that perhaps involve new ideas that we can test? One approach to this is the direction of Grand Unified theories, or GUTs.

3.6.2 — Grand Unified Theories

The central idea of GUTs is that, similarly to the symmetry breaking of the Higgs mechanism, there is an even larger symmetry group which is broken at some larger energy scale. What would be some good candidates for this larger group?

On the level of root systems, we are looking at $A_2 \times A_1$ (the factor U(1) is not semisimple, but you could think of a one-dimensional root system with no roots). The 'obvious' first try is to package them together in $A_4 = \mathrm{SU}(5)$. Then the question is how irreducible representations of SU(5) decompose when restricted to the subgroup $\mathrm{SU}(3) \times \mathrm{SU}(2) \times \mathrm{U}(1)$. The lowest-dimensional ones, and the relevant ones for the Standard Model, are

$$\mathbf{1} \to (\mathbf{1},\mathbf{1})_0\,, \tag{3.63}$$

$$\mathbf{5} \to (\mathbf{3},\mathbf{1})_{-\frac{2}{3}} \oplus (\mathbf{1},\mathbf{2})_1\,, \tag{3.64}$$

$$\bar{\mathbf{5}} \to (\bar{\mathbf{3}},\mathbf{1})_{\frac{2}{3}} \oplus (\mathbf{1},\mathbf{2})_{-1}\,, \tag{3.65}$$

$$\mathbf{10} \to (\mathbf{3},\mathbf{2})_{\frac{1}{3}} \oplus (\bar{\mathbf{3}},\mathbf{1})_{-\frac{4}{3}} \oplus (\mathbf{1},\mathbf{1})_1\,, \tag{3.66}$$

$$\mathbf{24} \to (\mathbf{8},\mathbf{1})_0 \oplus (\mathbf{1},\mathbf{3})_0 \oplus (\mathbf{1},\mathbf{1})_0 \oplus (\mathbf{3},\mathbf{2})_{-\frac{5}{3}} \oplus (\bar{\mathbf{3}},\mathbf{2})_{\frac{5}{3}}\,. \tag{3.67}$$

So we find that $\bar{\mathbf{5}} \oplus \mathbf{10}$ generates exactly all left-handed matter of the Standard Model, while all gauge bosons come from $\mathbf{24}$, the adjoint representation of SU(5). The right-handed matter would be in the dual, $\mathbf{5} \oplus \overline{\mathbf{10}}$. The right-handed neutrinos can be added by hand as the trivial representation.

Unfortunately, this model predicts proton decay at a far higher rate than has been observed (it has never been observed yet). So this is not quite the right model. But it looks appealing.

We can go further, to e.g. $D_5 = \mathrm{Spin}(10)$. By abuse of notation, this is often called the SO(10) GUT. In this case the lowest-dimensional branching rules from SO(10) to $\mathrm{SU}(5) \times \mathrm{U}(1)_\chi$ are (the $\mathrm{U}(1)_\chi$ is relevant for the Higgs symmetry breaking)

$$\mathbf{1} \to \mathbf{1}_0\,, \tag{3.68}$$

$$\mathbf{10} \to \mathbf{5}_{-2} \oplus \bar{\mathbf{5}}_2\,, \tag{3.69}$$

$$\mathbf{16} \to \mathbf{10}_1 \oplus \bar{\mathbf{5}}_{-3} \oplus \mathbf{1}_5\,, \tag{3.70}$$

$$\mathbf{45} \to \mathbf{24}_0 \oplus \mathbf{10}_{-4} \oplus \overline{\mathbf{10}}_4 \oplus \mathbf{1}_0\,. \tag{3.71}$$

So in this case, all (left-handed) matter, including the antineutrino, comes from $\mathbf{16}$, while all gauge bosons come from $\mathbf{45}$. The 'ordinary' Higgs comes from $\mathbf{10}$, and this is actually an issue, as this would also give a triplet for the Standard Model, whose masses cannot be incorporated consistently in the Standard Model without proton decay.

It is possible to go even further. Considering Dynkin diagrams, we could now go to e.g. E_6, E_7, or E_8. The first or these actually unifies the Higgs with all the other matter. Going farther, however, we turn into trouble: E_7 and E_8 have no complex representations, and these are needed to get chirality by a Higgs-like symmetry breaking. But again, there are ways aroung this in e.g. string theory, where $E_8 \times E_8$ heterotic string theory is an actual possibility.