

Lectures on Group Theory for Physicists

A Unified Course: Parts A–E

Karl Landsteiner

October 14, 2026

Contents

I	Group Theory: Discrete and Finite Groups	5
1	Introduction: Groups in Physics	5
2	The Definition of a Group	6
2.1	Axioms	6
2.2	First examples	7
3	Finite and Continuous, Abelian and Non-Abelian Groups	8
3.1	A continuous group: $SO(2)$	8
3.2	Discrete groups	9
3.3	A non-abelian discrete group: symmetries of the square	9
3.4	Continuous non-abelian groups: $SO(N)$ and $SU(N)$	11
4	Homomorphisms and Isomorphisms	12
5	Subgroups, the Rearrangement Theorem, and Cayley's Theorem	14
5.1	Subgroups	14
5.2	The Rearrangement Theorem	15
5.3	Cayley's theorem: every group is a permutation group	15
6	Conjugacy Classes	16
7	Normal (Invariant) Subgroups and Simple Groups	18
7.1	Cosets: a first look	19
8	Cosets and Lagrange's Theorem	19
8.1	When left and right cosets coincide	21
9	Quotient (Factor) Groups	22
10	Representations, Kernel, Image, and the First Isomorphism Theorem	24
10.1	Representations	24
10.2	Kernel and image	26
10.3	The First Homomorphism (Isomorphism) Theorem	26
11	Automorphisms and the Centre	28
11.1	The centre of a group	28

12 Direct and Semidirect Products	29
12.1 The direct product	29
12.2 The semidirect product	31
12.3 Worked example: the Euclidean group	32
13 Supplementary Exercises	33
 II Lie Groups: Continuous Symmetries	 34
14 From Discrete to Continuous Symmetries	34
15 Lie Groups: Definition and Basic Structure	35
15.1 The Lie Algebra as the Tangent Space at the Identity	36
15.2 Group Multiplication Near the Identity: Structure Constants	37
16 Key Examples of Lie Groups	37
16.1 Discrete Symmetries and the Four Components of the Lorentz Group	39
17 Connected Components	40
18 Compactness	41
19 Representations of Groups	42
19.1 Reducibility	43
20 Unitary Representations	43
21 Schur's Lemmas	46
21.1 Real, Complex, and Pseudo-real Representations	48
22 Orthogonality Relations and Characters	48
22.1 Characters	49
 III Lie Algebras: From Groups to Algebras and Back	 50
23 Introduction: Why Linearise a Lie Group?	50
24 The Lie Algebra: Abstract Definition	51
25 From the Group to the Algebra: Tangent Vectors and the Exponential Map	52
25.1 Tangent Vectors to Curves	53
25.2 The Exponential Map and One-Parameter Subgroups	54
25.3 A General Recipe: Linearising the Defining Condition of a Matrix Group	55
25.4 Example Revisited: $\mathfrak{su}(2)$	57
26 Left- and Right-Invariant (Haar) Measure	57
26.1 Construction and the Defining Functional Equation	57
26.2 Compact Groups Are Unimodular	58
27 The Lie Group–Lie Algebra Correspondence	60
27.1 Subalgebras and Ideals	60
27.2 Homomorphisms and Isomorphisms	61

28 Direct Products, Direct Sums, and Tensor Products	63
28.1 Three Constructions	63
28.2 Representations Built From These Constructions	65
29 The Adjoint Representation	67
30 Simply Connected Groups and the Universal Covering Group	69
30.1 Homotopy of Paths and Loops	69
30.2 The Universal Covering Group	70
31 Classification of the Irreducible Representations of $\mathfrak{su}(2) \cong \mathfrak{so}(3)$	71
31.1 Ladder Operators and the Casimir Invariant	72
31.2 The Highest-Weight Construction	72
32 The Clebsch–Gordan Decomposition	74
32.1 Counting States: Which j Occur?	74
32.2 Clebsch–Gordan Coefficients	75
33 Application: Spherical Harmonics as Representation Functions of $SO(3)$	77
IV Structure Theory and Classification of Semisimple Lie Algebras	78
34 Introduction: From a Single Algebra to a Classification	79
35 Simple and Semisimple Lie Algebras	80
35.1 Ideals, Simplicity, and Semisimplicity	80
35.2 The Killing Form	80
35.3 Cartan’s Criteria	81
35.4 Examples	82
36 Complexification and the Lorentz Algebra	83
36.1 Complexification and Real Forms	83
36.2 Example: $\mathfrak{sl}(2, \mathbb{C})$ and Its Two Real Forms	84
36.3 Physical Application: The Lorentz Algebra $\mathfrak{so}(3, 1)$	86
37 Cartan Subalgebras and the Root Space Decomposition	87
37.1 Cartan Subalgebras	87
37.2 The Killing Form on $\mathfrak{h}$, and the Induced Form on Roots	88
37.3 Worked Example: $\mathfrak{su}(2)_{\mathbb{C}} \cong \mathfrak{sl}(2, \mathbb{C})$	89
37.4 Worked Example: $\mathfrak{su}(3)_{\mathbb{C}}$ and the Gell-Mann Matrices	89
38 Root Strings and the Master Formula	90
38.1 The Chevalley Basis Attached to a Root	90
38.2 Root Spaces Are One-Dimensional	90
38.3 The α -String Through β , and the Master Formula	91
38.4 The Angle and Length Constraints	92
39 Simple Roots	93
39.1 Positive and Simple Roots	93
39.2 The Key Rigidity Theorem	93
40 The Cartan Matrix, Dynkin Diagrams, and the Classification	95
40.1 The Cartan Matrix	95

40.2 Dynkin Diagrams	96
40.3 Small Examples	97
41 Worked Example: The Root System of $SU(3)$	97
V Weights, the Classical Lie Algebras, and the Lorentz Group	98
42 Introduction: Completing the Programme	99
43 Weights, the Weyl Group, and the Highest-Weight Classification	99
43.1 Weights of a Representation	99
43.2 The Weyl Group	101
43.3 Fundamental Weights and Highest Weight	101
43.4 Reality, Compactness, and Weyl's Unitarity Theorem	102
43.5 Complex Conjugate Representations	103
44 Worked Example: The Representations of $\mathfrak{su}(3)$	103
44.1 Setup: Simple Roots and Fundamental Weights	103
44.2 The Fundamental and Antifundamental Representations	104
44.3 The Adjoint Representation	104
44.4 The Six-Dimensional Representation	105
44.5 The Decuplet and a Physical Application	105
45 Explicit Construction and Classification of the Classical Lie Algebras	106
45.1 $\mathfrak{su}(N) \leftrightarrow A_{N-1}$	106
45.2 $\mathfrak{so}(2n+1) \leftrightarrow B_n$	108
45.3 $\mathfrak{usp}(2n) \leftrightarrow C_n$	108
45.4 $\mathfrak{so}(2n) \leftrightarrow D_n$	109
45.5 Classical Groups as Rotation Groups over $\mathbb{R}, \mathbb{C}, \mathbb{H}$	109
46 Invariant Tensors and Casimir Operators	110
47 Physical Application: The Lorentz Group and Its Representations	111
47.1 The Lorentz Group	111
47.2 The Lie Algebra $\mathfrak{so}(3,1)$	111
47.3 Complexification and the (j, j') Classification	112
47.4 The Weyl Spinor Representations	112

Part I

Group Theory: Discrete and Finite Groups

1 Introduction: Groups in Physics

In physics, group theory is the theory of *transformations of physical objects*. Whenever we ask “does this system look the same after I do something to it?” — rotate it, translate it, exchange two identical particles, boost to another inertial frame, relabel a gauge field — we are asking a question about symmetry, and the collection of all transformations that leave the system looking the same (or map it to another admissible configuration in a controlled way) is almost always a *group*.

Example 1.1 (Rotations). Consider rotations of a rigid body, or of the coordinate axes, about a fixed point in three-dimensional space. Two features of rotations are immediately noteworthy, and both turn out to be central to the algebraic structure we are about to define.

- (i) **Multiplication.** Two rotations performed one after another produce a third rotation: “rotate by R_1 , then by R_2 ” is itself a rotation, which we may call $R_2 R_1$ (conventionally the transformation written on the right acts first). Composition of transformations is the *group multiplication*.
- (ii) **Non-commutativity.** Unlike multiplication of ordinary numbers, the order in which we perform two rotations matters: $R_2 R_1 \neq R_1 R_2$ in general. A vivid way to see this is to rotate a book by 90° about the x -axis and then 90° about the y -axis, and compare with doing this in the opposite order — the book ends up in a different orientation. This failure of commutativity is not a pathology; it is the generic behaviour of rotations in three (or more) dimensions, and it is the mathematical origin of phenomena such as the non-commuting components of angular momentum in quantum mechanics, $[\hat{L}_x, \hat{L}_y] = i\hbar \hat{L}_z$.

A more subtle feature of the rotation group is its *topology*: the set of rotations is not just an abstract set with a multiplication rule, it is also a curved space (a manifold) on which the group elements live, and this manifold is not simply connected. This is best illustrated by the *belt trick* (also called the Dirac string trick or plate trick): if one end of a belt is rotated by 360° while the other end is held fixed, the belt is left with an irreducible half-twist, and it takes a further 360° rotation (i.e. 720° in total) to untwist it completely. This demonstrates that a 2π rotation is not, topologically, the same as doing nothing, whereas a 4π rotation is — a fact that is directly responsible for the existence of spin- $\frac{1}{2}$ particles and the double-valuedness of spinor representations that will be discussed later in the course, once we come to the Lorentz group.

From the mathematical point of view, a group is an *algebraic structure*: a set together with certain rules for combining (“multiplying”) its elements, rules that are abstracted from very concrete examples such as the rotations above, but which turn out to capture an enormous range of phenomena — from the symmetries of crystals and molecules, through the conservation laws of Noether’s theorem, to the gauge symmetries of the Standard Model. The purpose of the present notes is to develop this algebraic structure — groups, subgroups, homomorphisms, cosets, quotient groups, and products of groups — carefully and with an eye to the physical examples that make it useful. Once this purely algebraic scaffolding is in place, the course continues with *Lie groups* (continuous groups that are also smooth manifolds), *Lie algebras* (the linearisation of Lie groups at the identity, which is where most practical computations in physics actually take place), the classification of simple Lie algebras, and finally the Lorentz group and spinors.

Physical perspective

Symmetry considerations pervade essentially every corner of theoretical physics. A few examples, roughly in order of increasing sophistication: crystals are classified by their finite point and space groups; angular momentum in quantum mechanics is organised by representations of $SO(3)$ and its double cover $SU(2)$; isospin and flavour symmetries of the strong interaction are described by $SU(2)$ and $SU(3)$; the Standard Model is a gauge theory with symmetry group $SU(3) \times SU(2) \times U(1)$ (or rather a suitable global quotient thereof); and special relativity is the statement that the laws of physics are invariant under the Poincaré group, the semidirect product of the Lorentz group with spacetime translations that we will meet in miniature, in its Euclidean cousin, at the end of these notes (Section 12).

2 The Definition of a Group

2.1 Axioms

A group is a set equipped with one binary operation, subject to four axioms. These axioms are chosen so that they are satisfied by the “obvious” examples of transformations (rotations, permutations, translations, ...) while being abstract enough to apply uniformly to sets of numbers, matrices, functions, and symmetry operations alike.

Definition 2.1 (Group). A set G together with a multiplication rule

$$\circ : G \times G \rightarrow G, \quad (g_1, g_2) \mapsto g_1 \circ g_2,$$

is called a *group* $(G, \circ)$ if the following four axioms hold:

- (a) **Closure.** $\forall g_1, g_2 \in G : g_1 \circ g_2 \in G$.
- (b) **Associativity.** $(g_1 \circ g_2) \circ g_3 = g_1 \circ (g_2 \circ g_3)$ for all $g_1, g_2, g_3 \in G$.
- (c) **Identity.** There exists an element $e \in G$ (the *identity* or *unit element*) such that

$$e \circ g = g \circ e = g \quad \forall g \in G,$$

and this element is unique.

- (d) **Inverse.** For every $g \in G$ there exists an element $g^{-1} \in G$ (the *inverse* of g) such that

$$g \circ g^{-1} = g^{-1} \circ g = e.$$

Remark. Only one of the two equalities $g \circ g^{-1} = e$ and $g^{-1} \circ g = e$ needs to be *postulated*: if a “left inverse” and a “left identity” exist for every element, the corresponding right versions can be derived from them (and conversely). This is a standard exercise in axiomatics that we will not dwell on, but it is worth remembering, because in some applications (e.g. semigroups with only partially defined inverses) the distinction matters. The same remark applies to the identity element: postulating $e \circ g = g$ for all g , together with the existence of inverses, is already enough to force $g \circ e = g$ as well.

Remark (Notation). When the group operation is written multiplicatively we normally drop the symbol $\circ$ and simply write $g_1 g_2$ for $g_1 \circ g_2$; the identity is then often denoted 1 or e , and the inverse of g is g^{-1} . When the operation is addition (as for $(\mathbb{Z}, +)$ below) we write $g_1 + g_2$, the identity is 0, and the inverse of g is $-g$. Physicists often reserve the word “group” loosely for the set G itself, leaving the multiplication rule implicit; strictly speaking

the group is the *pair* $(G, \circ)$, since a single underlying set of numbers can carry more than one natural group structure: the real numbers form a group $(\mathbb{R}, +)$ under addition, and, once 0 is removed, a different group $(\mathbb{R} \setminus \{0\}, \cdot)$ under multiplication (compare Examples 2.1 and 4.2 below).

Remark (Why exactly these axioms?). Each axiom encodes a property that is indispensable if G is to describe *transformations that can be undone and composed*. Closure says that performing two allowed transformations in succession gives another allowed transformation. Associativity says that when we compose three transformations, it does not matter whether we first combine the first two and then apply the third, or first combine the last two and then apply the first — the net transformation is the same; this is automatic when $\circ$ literally means “do one transformation after another”, so it may look like a triviality, but for an *abstract* set with an abstract rule $\circ$ it is a genuine constraint (indeed, dropping associativity gives the much larger and wilder class of *loops* and *quasigroups*). The identity is the “do nothing” transformation, and existence of inverses says that every transformation is *reversible*: if you can rotate a system, you can rotate it back.

2.2 First examples

Example 2.1 (Positive reals under multiplication). Let $G = \mathbb{R}^+ = (0, \infty)$, with $\circ$ the ordinary multiplication of real numbers, $r_1 \circ r_2 = r_1 r_2 \in \mathbb{R}^+$. We check the axioms:

- Closure: the product of two positive numbers is positive. ✓
- Associativity: $(r_1 r_2) r_3 = r_1 (r_2 r_3)$ for all reals. ✓
- Identity: $e = 1$, since $1 \cdot r = r \cdot 1 = r$. ✓
- Inverse: $r^{-1} = 1/r \in \mathbb{R}^+$ satisfies $r \cdot (1/r) = 1$. ✓

So $(\mathbb{R}^+, \cdot)$ is a group. Notice that it is automatically *abelian* (commutative), $r_1 r_2 = r_2 r_1$, since ordinary multiplication of real numbers commutes — this will not be true of groups in general, as Section 3 makes clear.

Example 2.2 (Integers under addition). Let $G = \mathbb{Z}$, with $\circ$ ordinary addition, $n_1 \circ n_2 = n_1 + n_2$.

- Closure: $n_1 + n_2 \in \mathbb{Z}$. ✓
- Associativity: $(n_1 + n_2) + n_3 = n_1 + (n_2 + n_3)$. ✓
- Identity: $e = 0$. ✓
- Inverse: for $n \in \mathbb{Z}$, the inverse is $-n$, since $n + (-n) = 0$. ✓

So $(\mathbb{Z}, +)$ is a group (again abelian). Note the identity element is 0, *not* 1: which element plays the role of “ e ” is a property of the operation, not an absolute fact about the number 1.

Example 2.3 (A non-example: integers under subtraction). Let $G = \mathbb{Z}$ with $g_1 \circ g_2 := g_1 - g_2$. This is *not* a group, because associativity fails:

$$(n_1 - n_2) - n_3 \neq n_1 - (n_2 - n_3)$$

in general (the left side equals $n_1 - n_2 - n_3$, the right side equals $n_1 - n_2 + n_3$; they agree only when $n_3 = 0$). This is a useful cautionary example: closure, an identity-like element (0, since $n - 0 = n$), and even a naive notion of “inverse” can all superficially seem to be present, yet the structure fails to be a group because of a single missing axiom. Always check *all four* axioms.

Physical perspective

The two examples $(\mathbb{R}^+, \cdot)$ and $(\mathbb{Z}, +)$ already illustrate the two broad classes of groups that will occupy us: $(\mathbb{R}^+, \cdot)$ is a *continuous* (indeed one-parameter, Lie) group, isomorphic as we will see later to $(\mathbb{R}, +)$ via the exponential map; $(\mathbb{Z}, +)$ is a *discrete, infinite, countable* group. Physics uses both kinds constantly: continuous groups appear as symmetries of continuous systems (rotations, translations, gauge transformations), while discrete groups appear as symmetries of discrete systems (crystal lattices, molecules, identical particle exchange).

3 Finite and Continuous, Abelian and Non-Abelian Groups

The two examples of Section 2 were both abelian and both had infinitely many elements. Physics needs much more: finite groups, infinite discrete groups, and continuous groups, some commutative and some not. This section collects the basic examples that will recur throughout the course.

3.1 A continuous group: $SO(2)$

Example 3.1 (Rotations of the plane). Let G be the set of 2×2 real matrices of the form

$$O(\varphi) = \begin{pmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{pmatrix}, \quad \varphi \in [0, 2\pi),$$

with group multiplication given by ordinary matrix multiplication. Each $O(\varphi)$ satisfies

$$O(\varphi)^T O(\varphi) = \mathbb{I}_2, \quad \det O(\varphi) = +1,$$

i.e. $O(\varphi)$ is a (real) *orthogonal* matrix of unit determinant; the set of all such matrices is denoted $SO(2)$, the *special orthogonal group* in two dimensions. Geometrically, $O(\varphi)$ rotates a vector in the plane counterclockwise by the angle φ , as illustrated in Fig. 1.

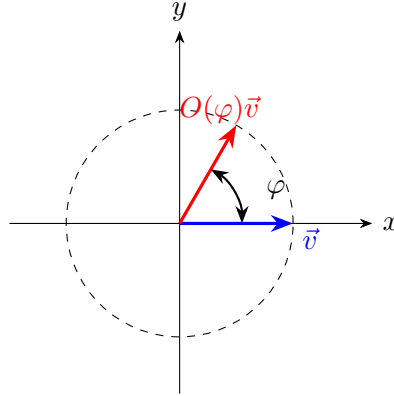


Figure 1: The action of $O(\varphi) \in SO(2)$ on a vector $\vec{v}$ in the plane: rotation counterclockwise through the angle φ .

Composing two rotations gives $O(\varphi_2)O(\varphi_1) = O(\varphi_1 + \varphi_2)$ (rotate by φ_1 , then by φ_2 , is the same as rotating once by $\varphi_1 + \varphi_2$), so $SO(2)$ is abelian — rotations in the *plane* always commute, unlike rotations in three or more dimensions (Example 1.1). $SO(2)$ is our first example of a *continuous group*: its elements are labelled by a continuous parameter φ , and nearby group elements (nearby values of φ) are “close” to one another in a sense that can be made precise

once a topology is put on the group — this is the beginning of the theory of Lie groups, taken up later in the course.

3.2 Discrete groups

Example 3.2 (The group $\mathbb{Z}_2$). $\mathbb{Z}_2 = \{1, -1\}$ with ordinary multiplication is a discrete group with two elements: $1 \cdot 1 = 1$, $1 \cdot (-1) = -1$, $(-1) \cdot (-1) = 1$; the identity is 1 and every element is its own inverse. $\mathbb{Z}_2$ is the simplest non-trivial group and it appears everywhere in physics: parity ($P^2 = 1$), the centre of many continuous groups (Section 11), $\mathbb{Z}_2$ gauge theories, Ising-model spins, ...

Example 3.3 (The cyclic group $\mathbb{Z}_N$). More generally, for $N \in \mathbb{N}$ let

$$\mathbb{Z}_N = \left\{ 1, e^{2\pi i/N}, e^{2 \cdot 2\pi i/N}, \dots, e^{(N-1) \cdot 2\pi i/N} \right\} \subset \mathbb{C},$$

the N -th roots of unity, with ordinary complex multiplication. Geometrically these are N equally-spaced points on the unit circle (Fig. 2); algebraically, $\mathbb{Z}_N$ is generated by a single element $g = e^{2\pi i/N}$ with $g^N = 1$, so that $\mathbb{Z}_N = \{1, g, g^2, \dots, g^{N-1}\}$. This is our first example of a *cyclic group*, and (like Example 3.2) $\mathbb{Z}_N$ is abelian and *discrete* but, unlike $\mathbb{Z}_2$, it has more than two elements: it interpolates, as $N \rightarrow \infty$, towards the continuous group $SO(2)$ of Example 3.1 (indeed $\mathbb{Z}_N$ is precisely the subgroup of $SO(2)$ consisting of rotations by multiples of $2\pi/N$).

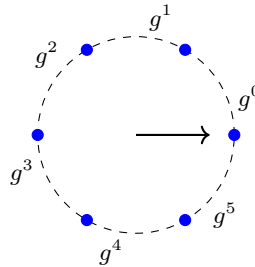


Figure 2: The cyclic group $\mathbb{Z}_6$: the six sixth roots of unity, equally spaced on the unit circle, with $g = e^{2\pi i/6}$.

3.3 A non-abelian discrete group: symmetries of the square

All the groups discussed so far are abelian. To see non-commutativity in a finite group we look at the symmetries of a square (this is the dihedral group D_4 , of *order* 8 — see Definition 3.1 below for the general definition of the order of a group).

Example 3.4 (Symmetries of the square). Label the corners of a square 1, 2, 3, 4 as in Fig. 3. The rigid motions of the plane that map the square onto itself (permuting its corners but preserving distances) are: the identity, three non-trivial rotations about the centre (by 90° , 180° , 270°), and four reflections (across the two diagonals and across the two lines joining opposite edge midpoints). Represented as 2×2 orthogonal matrices acting on the plane,

these eight symmetries are

$$\begin{aligned} M_1 &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, & M_2 &= \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, & M_3 &= \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, & M_4 &= \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \\ M_5 &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, & M_6 &= \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, & M_7 &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, & M_8 &= \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}. \end{aligned}$$

Here M_1 is the identity; M_2, M_3, M_4 are rotations by $90^\circ, 180^\circ, 270^\circ$; M_5, M_6 are reflections across the horizontal and vertical axes; and M_7, M_8 are reflections across the two diagonals. The group multiplication is matrix multiplication, and $G = \{M_1, \dots, M_8\}$ is closed under it (composing two symmetries of the square gives another symmetry of the square) — this set of eight matrices is a concrete realisation of the abstract group D_4 .

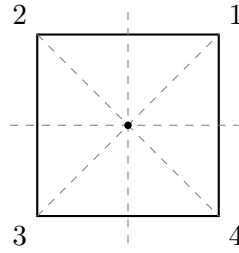


Figure 3: The square used to define D_4 in Example 3.4: the four corners are exchanged by the rotations M_2, M_3, M_4 and by the four reflections $M_5, \dots, M_8$ across the axes and diagonals shown dashed.

This group is **non-abelian**: for instance $M_2M_5 \neq M_5M_2$ (recalling our convention that the rightmost transformation acts first, M_2M_5 means “reflect across the horizontal axis, then rotate by 90° ”, which is not the same operation as M_5M_2 , “rotate by 90° first, then reflect” — physically, flipping the square over and then rotating the result is visibly different from rotating first and then flipping). This is checked directly:

$$\begin{aligned} M_2M_5 &= \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = M_7, \\ M_5M_2 &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} = M_8 \neq M_7. \end{aligned}$$

Exercise 3.1. Verify that $\{M_1, \dots, M_8\}$ of Example 3.4 is closed under matrix multiplication by completing the multiplication table (the “Cayley table”) of the group; check the group axioms directly; identify which pairs of elements do *not* commute.

Definition 3.1 (Order of a group). The number of elements of a group G (finite or infinite) is called the *order* of G , written $|G|$ or $\text{ord}(G)$. Thus $|\mathbb{Z}_2| = 2$, $|\mathbb{Z}_N| = N$, and the symmetry group of the square in Example 3.4 has order 8. Continuous groups such as $SO(2)$ have infinite order (indeed uncountably infinite, one element for each $\varphi \in [0, 2\pi)$).

3.4 Continuous non-abelian groups: $SO(N)$ and $SU(N)$

Example 3.5 (Special orthogonal and special unitary groups). Generalising Example 3.1, define

$$SO(N) = \{ O \in \mathbb{R}^{N \times N} : O^T O = \mathbb{1}_N, \det O = 1 \},$$

$$SU(N) = \{ U \in \mathbb{C}^{N \times N} : U^\dagger U = \mathbb{1}_N, \det U = 1 \},$$

both with matrix multiplication as the group operation. For $N \geq 3$ ($SO(N)$) and $N \geq 2$ ($SU(N)$) these groups are **non-abelian**: generic rotations in three or more dimensions do not commute (Example 1.1), and generic unitary matrices do not commute either.

It is instructive to verify explicitly that $SU(N)$ satisfies the group axioms, since the argument is a template used again and again in physics whenever a set of matrices defined by algebraic conditions (unitarity, orthogonality, unit determinant, tracelessness, ...) is claimed to form a group.

Proposition 3.2. $SU(N)$ is a group under matrix multiplication.

Proof. Let $U_1, U_2 \in SU(N)$, i.e. $U_1^\dagger U_1 = U_2^\dagger U_2 = \mathbb{1}_N$ and $\det U_1 = \det U_2 = 1$.

- **Closure.** We must check $U_1 U_2 \in SU(N)$:

$$(U_1 U_2)^\dagger (U_1 U_2) = U_2^\dagger U_1^\dagger U_1 U_2 = U_2^\dagger \mathbb{1}_N U_2 = U_2^\dagger U_2 = \mathbb{1}_N,$$

and using the multiplicativity of the determinant,

$$\det(U_1 U_2) = \det(U_1) \det(U_2) = 1 \cdot 1 = 1.$$

So $U_1 U_2 \in SU(N)$.

- **Associativity.** Matrix multiplication is associative for *all* matrices, so in particular for elements of $SU(N)$; nothing needs to be checked here beyond quoting this general fact.
- **Identity.** $\mathbb{1}_N$ is unitary with unit determinant, so $\mathbb{1}_N \in SU(N)$, and it acts as the identity element.
- **Inverse.** Since $U^\dagger U = \mathbb{1}_N$, the inverse of U is $U^{-1} = U^\dagger$; we must check $U^\dagger \in SU(N)$: $(U^\dagger)^\dagger U^\dagger = U U^\dagger = \mathbb{1}_N$ (using unitarity again, now the other way round) and $\det(U^\dagger) = \overline{\det U} = \overline{1} = 1$. So $U^{-1} \in SU(N)$.

All four axioms hold, so $SU(N)$ is a group. □

Exercise 3.3. Adapt the proof above to show that $SO(N)$, $O(N) = \{ O : O^T O = \mathbb{1}_N \}$ (orthogonal matrices, determinant ± 1), and $U(N) = \{ U : U^\dagger U = \mathbb{1}_N \}$ (unitary matrices, arbitrary phase of the determinant) are all groups under matrix multiplication. Which of these four groups ($SO(N)$, $O(N)$, $SU(N)$, $U(N)$) are abelian, and for which values of N ?

Physical perspective

$SO(3)$ is the group of rotations of ordinary three-dimensional space and organises orbital angular momentum in quantum mechanics; $SU(2)$, its double cover (two elements of $SU(2)$ map to each element of $SO(3)$, reflecting the belt-trick topology of Example 1.1), is needed as soon as spin is included. $SU(3)$ is the colour gauge group of the strong interaction (quantum

chromodynamics) and, historically, was first used (before gauge theories) to classify hadrons by their approximate flavour symmetry (the “Eightfold Way”). $U(1)$ is the gauge group of electromagnetism.

4 Homomorphisms and Isomorphisms

Groups become genuinely useful once we compare them with one another. The right notion of “a map between groups that respects the group structure” is that of a homomorphism.

Definition 4.1 (Homomorphism). Let $(G_1, \circ)$ and $(G_2, *)$ be groups. A map

$$\phi : G_1 \longrightarrow G_2$$

is called a *homomorphism* if it preserves the group structure, i.e.

$$\phi(g_1 \circ g_2) = \phi(g_1) * \phi(g_2) \quad \forall g_1, g_2 \in G_1.$$

Definition 4.2 (Isomorphism). If a homomorphism $\phi : G_1 \rightarrow G_2$ is *bijective* (one-to-one and onto), it is called an *isomorphism*, and G_1 and G_2 are said to be *isomorphic*, written $G_1 \cong G_2$.

In words: a homomorphism is a translation dictionary between two groups that turns products in G_1 into products in G_2 — it does not matter whether we multiply first in G_1 and then translate, or translate first and then multiply in G_2 ; the diagram

$$\begin{array}{ccc} (g_1, g_2) & \xrightarrow{\circ} & g_1 \circ g_2 \\ \downarrow \phi \times \phi & & \downarrow \phi \\ (\phi(g_1), \phi(g_2)) & \xrightarrow{*} & \phi(g_1) * \phi(g_2) = \phi(g_1 \circ g_2) \end{array}$$

commutes. An isomorphism is a homomorphism that, in addition, matches up the elements of G_1 and G_2 one-to-one: isomorphic groups are, from the point of view of group theory, “the same group” presented in two different guises (different labels for the elements, perhaps a different-looking multiplication rule, but identical abstract structure). This is why, in physics, it is common to say things like “the symmetry group of the equilateral triangle *is* S_3 ” rather than “... is isomorphic to S_3 ”: once two groups are known to be isomorphic, one may as well identify them.

Example 4.1 (The exponential map is a homomorphism, but not an isomorphism). Let $G_1 = (\mathbb{R}, +)$ and $G_2 = (\mathbb{R} \setminus \{0\}, \cdot)$, and consider

$$\phi : \mathbb{R} \rightarrow \mathbb{R} \setminus \{0\}, \quad \phi(x) = e^x.$$

This is a homomorphism, because the defining functional equation of the exponential is precisely the homomorphism property:

$$\phi(x + y) = e^{x+y} = e^x e^y = \phi(x) \phi(y).$$

However ϕ is *not* an isomorphism, because it is not *surjective*: $e^x > 0$ for every real x , so the negative numbers are never in the image of ϕ (nor is 0). Thus ϕ maps $(\mathbb{R}, +)$ homomorphically, but only *into*, not *onto*, $(\mathbb{R} \setminus \{0\}, \cdot)$.

Example 4.2 (A genuine isomorphism). Restricting the codomain in Example 4.1 to exactly the image of ϕ repairs the surjectivity failure. Let $G'_2 = (\mathbb{R}^+, \cdot)$ (the group of Example 2.1) and consider again

$$\phi : (\mathbb{R}, +) \longrightarrow (\mathbb{R}^+, \cdot), \quad \phi(x) = e^x.$$

Now ϕ is a bijection (its inverse is $\phi^{-1}(y) = \ln y$), so ϕ is an isomorphism:

$$(\mathbb{R}, +) \cong (\mathbb{R}^+, \cdot).$$

This identifies the additive group $(\mathbb{R}, +)$ with the multiplicative group $(\mathbb{R}^+, \cdot)$ of Example 2.1 — two of our earliest examples — as secretly the same group. Physically, this is exactly the statement used when one converts additive quantities (e.g. actions, or log-probabilities) into multiplicative ones (amplitudes, probabilities) via the exponential, or vice versa via the logarithm.

Example 4.3 (A homomorphism onto the circle group). Let $G_1 = (\mathbb{R}, +)$ and let $S^1 = \{z \in \mathbb{C} : |z| = 1\}$ be the unit circle in the complex plane, with ordinary complex multiplication. The map

$$\phi : (\mathbb{R}, +) \longrightarrow (S^1, \cdot), \quad \phi(x) = e^{ix}$$

is a homomorphism, since $\phi(x + y) = e^{i(x+y)} = e^{ix}e^{iy} = \phi(x)\phi(y)$. It is surjective (every point of the unit circle is e^{ix} for some real x) but *not* injective: $\phi(x) = \phi(x + 2\pi n)$ for every integer n , so infinitely many elements of $\mathbb{R}$ are mapped to the same element of S^1 . This is our first hint of the idea, made precise in Section 9, that $S^1 \cong \mathbb{R}/2\pi\mathbb{Z}$: the circle is the real line “with all multiples of 2π identified”. Physically, ϕ is the statement that a phase (an element of S^1 , e.g. the phase of a quantum-mechanical wavefunction or of a $U(1)$ gauge transformation) is only ever defined up to an additive shift by 2π .

Remark (The trivial homomorphism). For *any* pair of groups G_1, G_2 , the map sending every element of G_1 to the identity $e_2 \in G_2$,

$$\phi : G_1 \rightarrow G_2, \quad \phi(g) = e_2 \quad \text{for all } g \in G_1,$$

is a homomorphism (trivially: $\phi(g_1 \circ g_2) = e_2 = e_2 * e_2 = \phi(g_1) * \phi(g_2)$). It carries no information about G_1 whatsoever, which is why it is called the *trivial homomorphism*; nonetheless it is worth recording because “the trivial map is always a valid homomorphism” is a useful sanity check, and because the trivial homomorphism is the extreme case (largest possible kernel, smallest possible image) of the general theory developed in Section 10.

Exercise 4.1. For each of the following maps, decide whether it is a homomorphism, and if so whether it is injective, surjective, or bijective (an isomorphism):

- (1) $\phi : (\mathbb{R}, +) \rightarrow (\mathbb{R}^+, \cdot), \phi(x) = e^x$;
- (2) $\phi : (\mathbb{R}, +) \rightarrow (S^1, \cdot), \phi(x) = e^{ix}$;
- (3) $\phi : (\mathbb{Z}, +) \rightarrow (\mathbb{Z}_N, \cdot), \phi(n) = g^n$ where $g = e^{2\pi i/N}$ generates $\mathbb{Z}_N$ as in Example 3.3;

- (4) $\phi : SU(2) \rightarrow SO(3)$, sending a unitary matrix to the rotation it induces on the Pauli vector $\vec{\sigma} \cdot \vec{x} \mapsto U(\vec{\sigma} \cdot \vec{x})U^\dagger$ (you are not asked to prove this is a homomorphism from scratch, only to determine, given that it is one, whether it can possibly be injective in view of the belt-trick discussion of Example 1.1).

(Exercise 4.1 is best attempted again after Section 10, once the notions of kernel and image have been made precise; we record it here because the maps themselves belong naturally with the present examples.)

5 Subgroups, the Rearrangement Theorem, and Cayley's Theorem

5.1 Subgroups

Definition 5.1 (Subgroup). A subset $S \subseteq G$ is called a *subgroup* of G if S , equipped with the multiplication inherited from G , is itself a group.

Every group G has two subgroups that carry no information: G itself, and the one-element set $\{e\}$ consisting of the identity alone (for the trivial group $G = \{e\}$ these two coincide). Any subgroup $S \subsetneq G$ (i.e. $S \neq G$) is called a *proper* subgroup — note that $\{e\}$ itself *is* a proper subgroup of any non-trivial G under this standard usage; when we specifically want to exclude $\{e\}$ as well, we will say so explicitly, as in “proper, non-trivial subgroup”.

Checking all four group axioms every time we want to know whether a given subset is a subgroup is wasteful, since S automatically inherits associativity from G . The following criterion reduces the check to a single condition.

Theorem 5.1 (Subgroup criterion). A non-empty subset $S \subseteq G$ is a subgroup of G if and only if

$$s, s' \in S \implies s \circ (s')^{-1} \in S.$$

Proof. ($\Rightarrow$) If S is a subgroup this is immediate: $s' \in S \Rightarrow (s')^{-1} \in S$ (inverses exist in S), and then $s \circ (s')^{-1} \in S$ by closure.

($\Leftarrow$) Suppose $s \circ (s')^{-1} \in S$ whenever $s, s' \in S$. We recover the remaining axioms:

- **Identity.** Take $s' = s$ (possible since $S \neq \emptyset$): $s \circ s^{-1} = e \in S$.
- **Inverse.** Take $s' = s$ and use the newly established $e \in S$: setting the first element to e and the second to s gives $e \circ s^{-1} = s^{-1} \in S$.
- **Closure.** Given $s, s' \in S$, we already know $(s')^{-1} \in S$; applying the hypothesis to the pair $(s, (s')^{-1})$ gives $s \circ ((s')^{-1})^{-1} = s \circ s' \in S$.

Associativity is inherited automatically from G . □

Example 5.1. $\mathbb{Z}_2 = \{1, -1\}$ is a subgroup of $(\mathbb{R} \setminus \{0\}, \cdot)$: for $s, s' \in \{1, -1\}$ we always have $s(s')^{-1} \in \{1, -1\}$ (e.g. $(-1) \cdot (-1)^{-1} = 1$, $1 \cdot (-1)^{-1} = -1$). More generally $\mathbb{Z}_N$ of Example 3.3 is a subgroup of $SO(2)$ (identifying $e^{i\varphi}$ with the rotation matrix $O(\varphi)$): the criterion says that the product of one N -th root of unity with the inverse of another is again an N -th root of unity, which is clear since both live on the same N equally spaced points of

the circle.

5.2 The Rearrangement Theorem

The following simple but far-reaching fact says that multiplying every element of a group by a fixed element merely *permutes* the group — it never creates repeats and never misses anything.

Theorem 5.2 (Rearrangement theorem). Let $g' \in G$ be any fixed element. As g ranges over all of G , the products $g' \circ g$ (and likewise $g \circ g'$) run through every element of G exactly once.

Proof. Existence. Given any $g'' \in G$, set $g = (g')^{-1} \circ g''$. Then

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

so g'' does appear among the products $g' \circ g$.

Uniqueness. Suppose $g' \circ g = g' \circ g^\#$ for two elements $g, g^\# \in G$. Multiplying on the left by $(g')^{-1}$,

$$g = (g')^{-1} \circ g' \circ g = (g')^{-1} \circ g' \circ g^\# = g^\#,$$

so distinct g 's give distinct products $g' \circ g$: no repeats occur. Combining existence and uniqueness, every element of G occurs exactly once among $\{g' \circ g : g \in G\}$. $\square$

For a finite group, the rearrangement theorem is exactly the statement that in the *multiplication table* (or *Cayley table*) of G — the $|G| \times |G|$ array whose (i, j) entry is $g_i \circ g_j$ — every row and every column contains each group element exactly once, just like a completed Sudoku puzzle or a Latin square.

Example 5.2 (Cayley table of $\mathbb{Z}_4$). Let $\mathbb{Z}_4 = \{e, g, g^2, g^3\}$ with $g^4 = e$ (this is Example 3.3 for $N = 4$). Its multiplication table is

$\circ$	e	g	g^2	g^3
e	e	g	g^2	g^3
g	g	g^2	g^3	e
g^2	g^2	g^3	e	g
g^3	g^3	e	g	g^2

Each row and each column is a rearrangement of $\{e, g, g^2, g^3\}$, exactly as Theorem 5.2 guarantees; this table will be used again in Example 9.2 to build the quotient group $\mathbb{Z}_4/\{e, g^2\}$.

5.3 Cayley's theorem: every group is a permutation group

The rearrangement theorem has a striking consequence, due to Cayley: it shows that left multiplication by a fixed group element g' defines a *permutation* of the set G (a bijective map from G to itself), and that this assignment $g' \mapsto$ (permutation of G) is itself a homomorphism (indeed an injective one) from G into the group of all permutations of the set G .

Theorem 5.3 (Cayley). Every group G is isomorphic to a subgroup of the symmetric group $S_{|G|}$ (the group of all permutations of a set of $|G|$ elements), via the map

$$g' \mapsto \sigma_{g'}, \quad \sigma_{g'}(g) := g' \circ g.$$

Sketch. By Theorem 5.2, $\sigma_{g'}$ is a bijection $G \rightarrow G$, i.e. a permutation of the (labelled) elements of G . The map $g' \mapsto \sigma_{g'}$ is a homomorphism into the permutation group of G , since $\sigma_{g_1} \circ \sigma_{g_2}(g) = g_1 \circ (g_2 \circ g) = (g_1 \circ g_2) \circ g = \sigma_{g_1 \circ g_2}(g)$ for all g , i.e. $\sigma_{g_1} \sigma_{g_2} = \sigma_{g_1 g_2}$. It is injective because $\sigma_{g'} = \sigma_e = \text{id}$ forces $g' \circ e = e$, i.e. $g' = e$ (only the identity of G induces the identity permutation); this will be recognised in Section 10 as the statement that the *kernel* of $g' \mapsto \sigma_{g'}$ is trivial. An injective homomorphism identifies G with its image, a subgroup of $S_{|G|}$. $\square$

Example 5.3 (Cayley’s theorem for $\mathbb{Z}_3$). Let $\mathbb{Z}_3 = \{e, g, g^2\}$, $g^3 = e$. Label the three elements e, g, g^2 as $1, 2, 3$. Left multiplication by g sends $e \mapsto g, g \mapsto g^2, g^2 \mapsto e$, i.e. in the labelling $1 \rightarrow 2 \rightarrow 3 \rightarrow 1$: this is the 3-cycle $(1\ 2\ 3) \in S_3$. Similarly left multiplication by g^2 gives the permutation $(1\ 3\ 2)$. Thus $\mathbb{Z}_3$ is realised as the subgroup $\{\text{id}, (1\ 2\ 3), (1\ 3\ 2)\}$ of S_3 — the subgroup of *even* permutations (cyclic rotations) of three objects, i.e. the alternating group $A_3 \subset S_3$.

Physical perspective

Cayley’s theorem is conceptually important — it says that “abstract” groups are, up to isomorphism, nothing more exotic than sets of permutations — but for continuous groups it is not the most economical description: representing $SO(3)$ (an uncountably infinite group) as permutations of an uncountably infinite set is not as useful as representing it directly by 3×3 orthogonal matrices, as we already did in Example 3.5. The theory of group *representations* in Section 10 generalises this matrix idea to arbitrary groups and is, in practice, how physicists use group theory.

6 Conjugacy Classes

In a non-abelian group, “the same kind of element seen from a different point of view” is captured by the notion of conjugation. This idea will turn out, in Section 7, to be exactly what is needed to define normal subgroups, and later in the course it underlies the classification of representations by conjugacy class (e.g. characters of finite groups are constant on conjugacy classes).

Definition 6.1 (Conjugate elements). Two elements $g, g' \in G$ are called *conjugate* if there exists an element $x \in G$ such that

$$g' = x \circ g \circ x^{-1}.$$

Conjugacy is a *mutual* relation: if $g' = x \circ g \circ x^{-1}$ then also $g = x^{-1} \circ g' \circ x$, so “ g' is conjugate to g ” is the same statement as “ g is conjugate to g' ”. It is also *transitive*: if g is conjugate to g' (via x) and g' is conjugate to g'' (via y), then

$$g'' = y \circ g' \circ y^{-1} = y \circ (x \circ g \circ x^{-1}) \circ y^{-1} = (yx) \circ g \circ (yx)^{-1},$$

so g is conjugate to g'' (via yx). Together with the trivial fact that every element is conjugate to itself (via $x = e$), this shows that conjugacy is an *equivalence relation*, and therefore partitions G into disjoint equivalence classes.

Definition 6.2 (Conjugacy class). The *conjugacy class* of $h \in G$ is the set of all elements conjugate to h ,

$$[h] := \{x \circ h \circ x^{-1} : x \in G\}.$$

Example 6.1. If $h = e$, then $x \circ e \circ x^{-1} = x \circ x^{-1} = e$ for every x , so the conjugacy class of the identity is $[e] = \{e\}$: the identity is always alone in its own class. If G is *abelian*, then $x \circ h \circ x^{-1} = x \circ x^{-1} \circ h = h$ for every x , so *every* element is alone in its own conjugacy class — conjugation is only interesting in non-abelian groups.

Theorem 6.1. For any group G :

- (a) $\{e\}$ is a conjugacy class by itself.
- (b) Every element of G is a member of some conjugacy class.
- (c) No element of G can be a member of two different conjugacy classes.

In other words, the conjugacy classes of G form a *partition* of G into disjoint, non-empty subsets.

Proof. (a) is Example 6.1.

(b) Given any $h \in G$, take $x = e$: $e \circ h \circ e^{-1} = h$, so h belongs to its own conjugacy class $[h]$.

(c) Suppose g belongs to both $[h_1]$ and $[h_2]$, i.e.

$$g = x \circ h_1 \circ x^{-1} \quad \text{and} \quad g = y \circ h_2 \circ y^{-1}$$

for some $x, y \in G$. We must show $[h_1] = [h_2]$. From $g \in [h_1]$ it follows (by the mutuality noted above) that $h_1 \in [g]$, and likewise $h_2 \in [g]$. Equating the two expressions for g ,

$$x \circ h_1 \circ x^{-1} = y \circ h_2 \circ y^{-1} \implies h_2 = (y^{-1}x) \circ h_1 \circ (y^{-1}x)^{-1},$$

which exhibits h_2 as conjugate to h_1 . Since conjugacy is an equivalence relation, conjugate elements have identical conjugacy classes, so $[h_1] = [h_2]$. $\square$

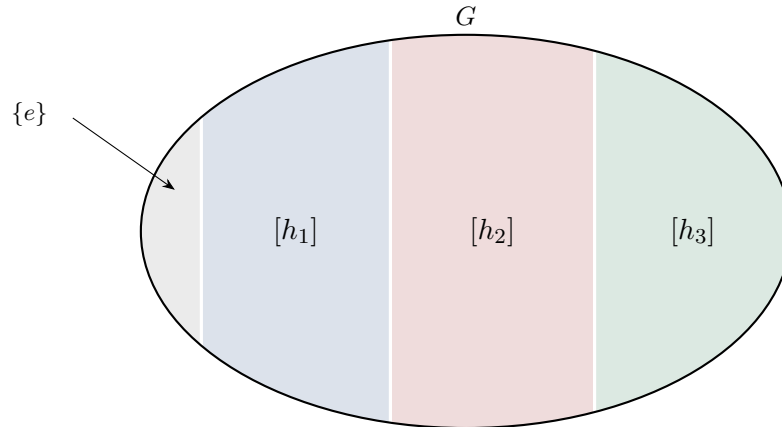


Figure 4: The conjugacy classes of a group G partition G into disjoint subsets that together cover all of G (Theorem 6.1): the identity is always its own (here, thinnest) class, and every other element belongs to exactly one of the remaining classes. (Compare the analogous picture for cosets, Fig. 5.)

Example 6.2 (Conjugacy classes of S_3). Let $S_3 = \{e, (12), (13), (23), (123), (132)\}$ be the group of permutations of three objects (equivalently, via its faithful action on the three vertices, the symmetry group of an equilateral triangle). A direct computation (conjugating each element by every group element) shows that S_3 splits into exactly three conjugacy

classes:

$$[e] = \{e\}, \quad [(12)] = \{(12), (13), (23)\}, \quad [(123)] = \{(123), (132)\}.$$

Geometrically, on the triangle, the class $\{(12), (13), (23)\}$ consists of the three reflections (each fixes one vertex and swaps the other two), and the class $\{(123), (132)\}$ consists of the two non-trivial rotations (by 120° and 240°). This illustrates a general and very useful heuristic: elements in the same conjugacy class of a symmetry group are typically “the same kind of geometric operation” (all reflections together, all rotations of a given order together), because conjugation by x amounts to “do the same operation, but as seen from a frame rotated/reflected by x ”.

Physical perspective

The physical importance of conjugacy classes is that any function of a group element which is built purely out of the group structure (e.g. the trace of a matrix representation, or any quantity invariant under a change of basis/frame) can only depend on the conjugacy class, not on the particular representative. This is why, for example, the character $\chi(g) = \text{tr } \rho(g)$ of a representation ρ is constant on conjugacy classes, a fact used throughout representation theory and particle physics (classifying multiplets by “Casimir invariants”, which are class functions).

7 Normal (Invariant) Subgroups and Simple Groups

Conjugacy classes give us exactly the tool needed to single out the subgroups that behave well with respect to conjugation — these are the subgroups that can be used to build quotient groups in Section 9, and they play the role, among groups, that ideals play among rings.

Definition 7.1 (Invariant / normal subgroup). A subgroup $S \subseteq G$ is called an *invariant subgroup* (or *normal subgroup*) if it is mapped to itself by conjugation with every element of G :

$$\forall s \in S \text{ and } \forall g \in G : \quad g \circ s \circ g^{-1} \in S.$$

Every group has (at least) two normal subgroups, called the *universal* normal subgroups: G itself, and the trivial subgroup $\{e\}$ (conjugating any element of G stays in G ; conjugating e always gives e). Whether G possesses any *other* normal subgroups is a highly non-trivial structural question about G .

Theorem 7.1. A subgroup $S \subseteq G$ is normal if and only if S is a union of complete conjugacy classes of G (i.e. whenever $s \in S$, the entire class $[s]$ is contained in S).

Proof. ($\Rightarrow$) If S is normal, then for every $s \in S$ and every $x \in G$ we have $x \circ s \circ x^{-1} \in S$ by definition, i.e. the whole conjugacy class $[s] = \{x s x^{-1} : x \in G\}$ lies in S . Hence S is made up of complete conjugacy classes.

($\Leftarrow$) Suppose S consists entirely of complete conjugacy classes: for every $s \in S$, $[s] \subseteq S$. Then in particular $x \circ s \circ x^{-1} \in [s] \subseteq S$ for every $x \in G$, which is precisely the statement that S is invariant. $\square$

Theorem 7.1 gives an efficient way to search for normal subgroups of a finite group: list the

conjugacy classes, and check which unions of classes (always including $\{e\}$, and closed under the group operation) form subgroups.

Example 7.1 (Normal subgroups of S_3). Recall from Example 6.2 that S_3 has three conjugacy classes, $\{e\}$ (size 1), $\{(12), (13), (23)\}$ (size 3), and $\{(123), (132)\}$ (size 2). A normal subgroup must be a union of whole classes that is also closed under multiplication and contains inverses. The only possibilities besides the universal ones $\{e\}$ and S_3 itself is $\{e\} \cup \{(123), (132)\} = \{e, (123), (132)\}$, and this set is indeed a subgroup (it is $A_3 \cong \mathbb{Z}_3$, the rotations of the triangle). The set $\{e\} \cup \{(12), (13), (23)\}$, by contrast, is *not* even a subgroup: for instance $(12) \cdot (13) = (132)$ is not in the set. So S_3 has exactly one non-trivial proper normal subgroup, $A_3 \cong \mathbb{Z}_3$.

Definition 7.2 (Simple group). A group G is called *simple* if its only normal subgroups are the universal ones, $\{e\}$ and G itself.

Simple groups are the “atoms” out of which all finite groups can be built by repeated extension (this is the content of the Jordan–Hölder theorem, which is beyond the scope of these notes but is the finite-group analogue of the statement, familiar from atomic structure or nuclear shell models, that complex structures decompose into elementary building blocks). S_3 , having the non-trivial normal subgroup A_3 , is *not* simple; by contrast $\mathbb{Z}_p$ for p prime *is* simple, because (as we will see as an immediate consequence of Lagrange’s theorem in Section 8) a group of prime order has no proper subgroups at all, let alone proper normal ones.

7.1 Cosets: a first look

To go further — in particular, to understand *how big* a normal subgroup’s conjugacy-class decomposition can be, and to build the quotient group G/S in Section 9 — we need the notion of a coset, which we introduce here and study in detail in the next section.

Definition 7.3 (Coset). Let H be a subgroup of G and let $g \in G$ be any fixed element. The sets

$$gH := \{g \circ h : h \in H\}, \quad Hg := \{h \circ g : h \in H\}$$

are called, respectively, the *left coset* and the *right coset* of H with respect to g .

Note immediately that $eH = He = H$ itself is one of the (left and right) cosets of H — namely the coset associated with $g = e$. The full theory of cosets, including the key fact that distinct cosets are always either identical or completely disjoint, and Lagrange’s theorem on the order of subgroups of finite groups, is developed in Section 8.

8 Cosets and Lagrange’s Theorem

We now develop the theory of cosets introduced at the end of Section 7, culminating in Lagrange’s theorem: one of the most useful elementary facts in all of finite group theory, since it constrains, from pure counting, which subgroups a given finite group can possibly have.

Throughout this section $H \leq G$ is a fixed subgroup (not necessarily normal), and we study its left cosets gH , $g \in G$; everything applies equally, mutatis mutandis, to right cosets Hg .

Theorem 8.1. Let $H \leq G$ be a subgroup. Then:

- (a) If $h \in H$, then $Hh = hH = H$ (a special case of the rearrangement theorem, Theorem 5.2, applied within H).
- (b) If $g \notin H$, then gH is *not* a subgroup of G .
- (c) Every $g \in G$ is a member of some left coset of H (namely of gH itself).
- (d) If H has finite order $s = |H|$, then every left coset gH contains exactly s distinct elements.
- (e) Two left cosets g_1H and g_2H are either identical or completely disjoint (they share no common element).

Proof. (a) This is Theorem 5.2 applied to the group H itself: left (or right) multiplication by $h \in H$ merely rearranges the elements of H .

(b) If gH were a subgroup it would have to contain the identity: $e = g \circ h'$ for some $h' \in H$, i.e. $g = (h')^{-1} \in H$. Contrapositively, if $g \notin H$ then gH cannot be a subgroup.

(c) Since $e \in H$ (as H is a subgroup), $g = g \circ e \in gH$.

(d) The map $h \mapsto g \circ h$ from H to gH is surjective by definition of gH , and it is injective because $g \circ h_1 = g \circ h_2$ implies $h_1 = h_2$ (multiply both sides on the left by g^{-1}). So it is a bijection $H \rightarrow gH$, and in particular $|gH| = |H| = s$.

(e) Suppose g_1H and g_2H share a common element, i.e. $g_1 \circ h_1 = g_2 \circ h_2$ for some $h_1, h_2 \in H$. Then

$$g_2^{-1} \circ g_1 = h_2 \circ h_1^{-1} =: h' \in H \quad \implies \quad g_2^{-1} \circ g_1 \circ H = h' \circ H = H$$

(using part (a)). Multiplying on the left by g_2 ,

$$g_2H = g_2 \circ (g_2^{-1}g_1H) = (g_2g_2^{-1})g_1H = g_1H.$$

So sharing even a single element forces the two cosets to be *completely* equal; equivalently, two cosets that are not identical must be disjoint. $\square$

Corollary 8.2. If $g' \in gH$, then $g'H = gH$: every element of a coset is an equally good “representative” for the whole coset.

Proof. If $g' = g \circ h$ for some $h \in H$, then $g'H = (gh)H = g(hH) = gH$ using part (a) of Theorem 8.1. Concretely, if $g' \in gH$ then $g' \circ h' = g \circ h \circ h' = g \circ h''$ for $h, h', h'' \in H$, so every element of $g'H$ already lies in gH , and vice versa. $\square$

Part (e) of Theorem 8.1, together with part (c), says exactly that the left cosets of H *partition* G into disjoint blocks, each of size $|H|$ (by part (d), when H is finite): see Fig. 5. This is the promised generalisation of the partition into conjugacy classes of Fig. 4 — indeed $H = \{e\}$ recovers a (trivial) partition of G into singletons, while $H = G$ gives the (equally trivial) partition of G into a single block.

Theorem 8.3 (Lagrange’s theorem). If G and H are finite, with $|G| = r$ and $H \leq G$ of order $|H| = s$, then s divides r , and the number of distinct left cosets of H in G is

$$[G : H] := \frac{r}{s}.$$

(The number $[G : H]$ is called the *index* of H in G .)

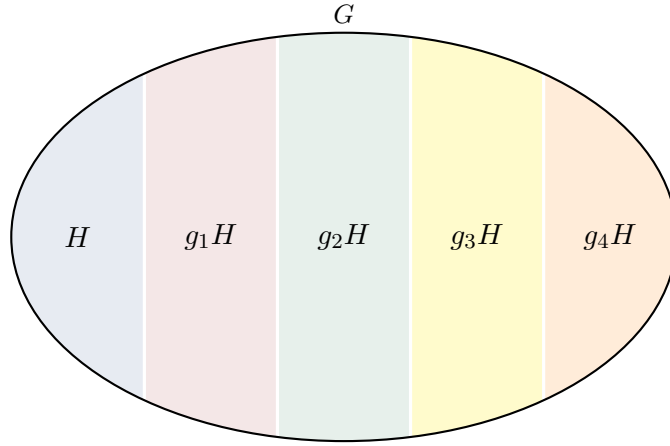


Figure 5: The left cosets of $H \leq G$ partition G into disjoint blocks that together cover all of G , each in one-to-one correspondence with H itself (Theorem 8.1).

Proof. By Theorem 8.1(c,e), the left cosets of H partition G into disjoint blocks; by part (d), each block has exactly s elements. If there are n such blocks, then counting elements of G block by block gives $r = |G| = n \cdot s$, so $s \mid r$ and $n = r/s$. $\square$

Corollary 8.4. If $|G| = r$ with r prime, then G has no proper (non-trivial) subgroups: its only subgroups are $\{e\}$ and G itself. Consequently $\mathbb{Z}_p$ (p prime) is a simple group (Section 7).

Proof. By Lagrange's theorem, the order s of any subgroup $H \leq G$ must divide r . If r is prime, its only divisors are 1 and r itself, so $s = 1$ (forcing $H = \{e\}$) or $s = r$ (forcing $H = G$). $\square$

Example 8.1. $|S_3| = 6 = 2 \cdot 3$. Lagrange's theorem allows subgroups of order 1, 2, 3, 6 (and forbids, say, a subgroup of order 4 or 5): indeed S_3 has subgroups $\{e\}$ (order 1), $\{e, (12)\}$, $\{e, (13)\}$, $\{e, (23)\}$ (three subgroups of order 2, generated by each reflection), $\{e, (123), (132)\} = A_3$ (order 3), and S_3 itself (order 6) — and no others.

8.1 When left and right cosets coincide

We can now sharpen Section 7's definition of a normal subgroup: normality is exactly the condition under which left and right cosets agree.

Theorem 8.5. Left and right cosets of H are identical, $gH = Hg$ for every $g \in G$, if and only if H is normal (invariant) in G .

Proof. Suppose $gH = Hg$ for all g . Take any $h_1 \in H$; since $g \circ h_1 \in gH = Hg$, we can write $g \circ h_1 = h_2 \circ g$ for some $h_2 \in H$, i.e.

$$g \circ h_1 \circ g^{-1} = h_2 \in H.$$

As this holds for every $h_1 \in H$ and every $g \in G$, H is normal. Conversely, if H is normal, then for every $h_1 \in H$ and $g \in G$ there is $h_2 = gh_1g^{-1} \in H$ with $gh_1 = h_2g \in Hg$; so $gH \subseteq Hg$, and the reverse inclusion follows in the same way using g^{-1} in place of g . Hence $gH = Hg$. $\square$

Remark. For a normal subgroup, $gH = Hg$ does *not* mean that g commutes with every individual element of H (i.e. it does not mean $g \circ h = h \circ g$ for each h); it only means that the two *sets* gH and Hg coincide as sets — multiplying H by g on the left produces the same coset as multiplying by g on the right, but individual elements may well be permuted amongst themselves in the process. This distinction is exactly what makes normal subgroups useful for building quotient *groups* in Section 9, rather than merely quotient sets.

9 Quotient (Factor) Groups

We now come to one of the central constructions of group theory: given a normal subgroup $H \trianglelefteq G$, the cosets of H can themselves be organised into a group, called the quotient (or factor) group G/H . This is the precise mathematical incarnation of the informal idea, already glimpsed in Example 4.3 ($S^1 \cong \mathbb{R}/2\pi\mathbb{Z}$), of “ G with the elements of H identified with the identity”.

Definition 9.1 (Quotient group). Let $H \trianglelefteq G$ be a normal subgroup. The *quotient group* (or *factor group*) G/H is the set of left cosets of H in G ,

$$G/H = \{ gH : g \in G \},$$

equipped with the multiplication

$$(g_1H) \circ (g_2H) := (g_1 \circ g_2)H.$$

Proposition 9.1 (Well-definedness). The multiplication rule above does not depend on which representatives g_1, g_2 are chosen for the cosets g_1H, g_2H (this is where normality of H is essential).

Proof. Let g'_1 be another representative of g_1H and g'_2 another representative of g_2H , so $g'_1 = g_1 \circ h_1$ and $g'_2 = g_2 \circ h_2$ for some $h_1, h_2 \in H$ (Corollary 8.2). Then

$$g'_1 \circ g'_2 = g_1 \circ h_1 \circ g_2 \circ h_2 = g_1 \circ g_2 \circ (g_2^{-1}h_1g_2) \circ h_2.$$

Because H is *normal*, $g_2^{-1}h_1g_2 =: h'_1 \in H$, so

$$g'_1 \circ g'_2 = g_1 \circ g_2 \circ \underbrace{h'_1}_{\in H} \circ h_2 \in (g_1 \circ g_2)H.$$

Hence $g'_1H \circ g'_2H := (g'_1g'_2)H = (g_1g_2)H = g_1H \circ g_2H$: the product coset does not depend on the choice of representatives. (This computation is exactly why the definition of G/H requires H to be normal: for a non-normal subgroup, moving h_1 past g_2 would not, in general, land back inside H , and the product of cosets would not even be well defined as a single coset.) $\square$

Proposition 9.2. $(G/H, \circ)$ is a group.

Proof. *Closure* is built into the definition (the product of two cosets is again a coset, by Proposition 9.1). *Associativity* is inherited from G :

$$\begin{aligned} (g_1H) \circ [(g_2H) \circ (g_3H)] &= g_1H \circ (g_2g_3)H = (g_1(g_2g_3))H \\ &= ((g_1g_2)g_3)H = (g_1g_2)H \circ g_3H = [(g_1H) \circ (g_2H)] \circ (g_3H). \end{aligned}$$

Identity: $eH = H$ acts as the identity element of G/H , since $gH \circ H = (ge)H = gH$ (and likewise $H \circ gH = gH$). *Inverse:* for any $gH \in G/H$, the coset $g^{-1}H$ is a two-sided inverse, since $g^{-1}H \circ gH = (g^{-1}g)H = eH = H$ (and likewise the other way round). $\square$

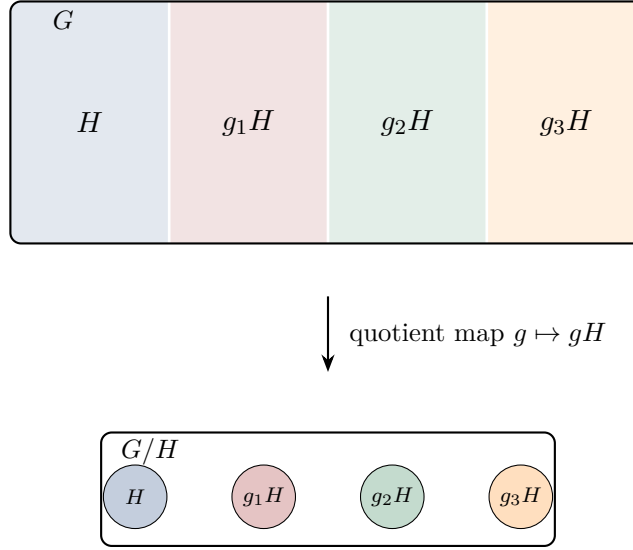


Figure 6: The quotient map $G \rightarrow G/H$ collapses each coset of H to a single point; the group G/H has one element for each coset, i.e. $|G/H| = [G : H]$ when G is finite (Theorem 8.3).

Example 9.1 (Quotienting $\mathbb{R}^N$ by a coordinate subspace). Let $G = (\mathbb{R}^N, +)$, the additive group of N -component real vectors $\vec{v} = (v_1, \dots, v_N)$, and let $H = (\mathbb{R}^k, +)$ ($k < N$) be the subgroup of vectors whose last $N - k$ components vanish, $\vec{v} = (v_1, \dots, v_k, 0, \dots, 0)$. Since G is abelian, every subgroup is automatically normal (conjugation is trivial, $ghg^{-1} = h$), so the quotient G/H is defined. Two vectors $\vec{v}, \vec{v}'$ lie in the same coset of H iff they agree in their last $N - k$ components, $v_{k+1} = v'_{k+1}, \dots, v_N = v'_N$ (they may differ arbitrarily in the first k components, which is exactly the freedom to add an element of H). So each coset is uniquely labelled by the tuple $(v_{k+1}, \dots, v_N) \in \mathbb{R}^{N-k}$, and

$$\mathbb{R}^N / \mathbb{R}^k \cong \mathbb{R}^{N-k}.$$

In this example one can also find a subgroup $K = (\{0\}^k \times \mathbb{R}^{N-k}, +) \leq G$, isomorphic to G/H , such that $G \cong H \times K$ literally (as a direct product, Section 12) — every vector splits uniquely as $\vec{v} = (v_1, \dots, v_k, 0, \dots, 0) + (0, \dots, 0, v_{k+1}, \dots, v_N)$, the first summand in H and the second in K . This illustrates the general principle, discussed further below, that $G/H \times H \cong G$ precisely when the quotient G/H can be found, in addition, as an honest (here: normal) subgroup $K \leq G$ complementary to H .

Example 9.2 ($\mathbb{Z}_4 / \{e, g^2\} \cong \mathbb{Z}_2$). Let $G = \mathbb{Z}_4 = \{e, g, g^2, g^3\}$, $g^4 = e$, with g represented concretely as $e^{i\pi/2}$ (Example 5.2). The subset $H = \{e, g^2\}$ is a subgroup, isomorphic to $\mathbb{Z}_2$ (check: $g^2 \circ g^2 = g^4 = e$), and since $\mathbb{Z}_4$ is abelian, H is automatically normal. The left cosets of H are

$$eH = \{e, g^2\}, \quad gH = \{g, g^3\},$$

and by Theorem 8.3 these are *all* the cosets, since $[\mathbb{Z}_4 : H] = 4/2 = 2$. The quotient group $\mathbb{Z}_4/H = \{eH, gH\}$ has multiplication table

$$eH \circ eH = eH, \quad eH \circ gH = gH, \quad gH \circ gH = (g \cdot g)H = g^2H = eH,$$

(the last equality holds because $g^2 \in H$, so $g^2H = H = eH$ by Theorem 8.1(a)). This is exactly the multiplication table of $\mathbb{Z}_2$, so

$$\mathbb{Z}_4/\{e, g^2\} \cong \mathbb{Z}_2.$$

Remark (A subtlety: $G/H \times H$ need not reproduce G). It is tempting to think that a group G can always be reconstructed from a normal subgroup H together with the quotient G/H , via a direct product $G/H \times H$ (direct products are defined precisely in Section 12, but the reader can for now take $A \times B = \{(a, b) : a \in A, b \in B\}$ with componentwise multiplication). This is *false* in general. In Example 9.2 we found $\mathbb{Z}_4/\{e, g^2\} \cong \mathbb{Z}_2$ and the subgroup itself is $\{e, g^2\} \cong \mathbb{Z}_2$, so naively one might guess $\mathbb{Z}_4 \cong \mathbb{Z}_2 \times \mathbb{Z}_2$. But this is wrong: in $\mathbb{Z}_2 \times \mathbb{Z}_2 = \{(e, e), (e, g), (g, e), (g, g)\}$, *every* non-identity element squares to the identity,

$$(e, g)^2 = (e, e), \quad (g, e)^2 = (e, e), \quad (g, g)^2 = (g^2, g^2) = (e, e),$$

whereas in $\mathbb{Z}_4$ the element g satisfies $g^2 \neq e$ (indeed g has order 4, not 2). So $\mathbb{Z}_2 \times \mathbb{Z}_2 \not\cong \mathbb{Z}_4$, even though both are abelian groups of order 4 built out of “the same” subgroup and quotient. The resolution of this apparent paradox is that $G \cong G/H \times H$ holds only under an extra condition:

$$\begin{aligned} G/H \times H \cong G \quad &\iff \quad \exists \text{ a subgroup } K \leq G \text{ with } K \cong G/H, \\ &K \text{ normal in } G, \quad K \cap H = \{e\}, \quad KH = G. \end{aligned}$$

i.e. the quotient G/H must be realisable not just abstractly, but as an actual normal subgroup K of G that meets H only in the identity and, together with H , generates all of G (compare Example 9.1, where such a K does exist). All four conditions on the right are needed: merely finding a normal $K \cong G/H$ is not enough — e.g. in $G = \mathbb{Z}_8 = \langle g \rangle$ with $H = \{e, g^4\} \cong \mathbb{Z}_2$, the subgroup $K = \langle g^2 \rangle \cong \mathbb{Z}_4$ is normal (abelian G) and isomorphic to $G/H \cong \mathbb{Z}_4$, yet $\mathbb{Z}_8 \not\cong \mathbb{Z}_2 \times \mathbb{Z}_4$ (a direct check of element orders shows $\mathbb{Z}_2 \times \mathbb{Z}_4$ has no element of order 8); what fails is $K \cap H = \{e\}$, since $g^4 \in K$. In $\mathbb{Z}_4$, no subgroup isomorphic to $\mathbb{Z}_2$ other than $\{e, g^2\}$ itself exists, so no such complementary K can be found either, and the direct product reconstruction fails. The precise statement of when G *does* split as a direct (or, more generally, semidirect) product of two of its subgroups is the subject of Section 12.

10 Representations, Kernel, Image, and the First Isomorphism Theorem

We return to homomorphisms (Section 4), now specialised to the case that is by far the most useful for physics: homomorphisms of an abstract group G into groups of matrices.

10.1 Representations

Definition 10.1 (Representation). A *representation* of a group G is a homomorphism

$$\rho : G \longrightarrow \text{GL}(d, \mathbb{C})$$

from G into the group of non-singular $d \times d$ matrices (i.e. onto a subgroup of it, not necessarily onto all of it), with matrix multiplication as the group product in the target. The integer d is called the *dimension* of the representation.

Physicists usually care about representations rather than about abstract groups *per se*, because physical quantities (wavefunctions, field multiplets, tensors) live in vector spaces, and symmetries act on these vector spaces linearly — that is exactly what a representation encodes.

Example 10.1 (Representations of $\mathbb{Z}_2$). Let $\mathbb{Z}_2 = \{e, g\}$, $g^2 = e$ (Example 3.2).

- A *one-dimensional* representation: $\rho_1(e) = 1$, $\rho_1(g) = -1$ (both 1×1 “matrices”).
Check: $\rho_1(g)\rho_1(g) = (-1)(-1) = 1 = \rho_1(g^2) = \rho_1(e)$. ✓
- A *two-dimensional* representation:

$$\rho_2(e) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \rho_2(g) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

$$\text{Check: } \rho_2(g)^2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \rho_2(e). \quad \checkmark$$

Both are legitimate representations of the same group; a single abstract group has, in general, infinitely many inequivalent representations *in total* (summed over all dimensions). For a *finite* group, however, there are only finitely many irreducible representations (the theory of irreducible representations, central to applications, is developed later in the full course), and by Maschke’s theorem every finite-dimensional representation decomposes into irreducibles; consequently there are also only finitely many inequivalent representations of any one *fixed* dimension d (as many as there are ways to build up d as an ordered sum of irreducible dimensions). For instance, $\mathbb{Z}_2$ has exactly two inequivalent one-dimensional representations (ρ_1 above and the trivial one) and exactly three inequivalent two-dimensional ones (trivial $\oplus$ trivial, trivial $\oplus\rho_1$, and $\rho_1 \oplus \rho_1$, of which ρ_2 above is an example of the last, in a non-diagonal basis) — not infinitely many.

Definition 10.2 (Faithful representation). If $\rho : G \rightarrow \text{GL}(d, \mathbb{C})$ is *injective* (equivalently, bijective onto its image, i.e. an isomorphism onto a matrix group), it is called an *isomorphism*, and the representation is said to be *faithful*: distinct group elements are represented by distinct matrices, so no information about G is lost.

ρ_1 of Example 10.1 is faithful (both elements give different 1×1 matrices); so is ρ_2 . By contrast, the trivial representation $\rho(g) = 1$ for *every* $g \in G$ (the matrix version of the trivial homomorphism of Section 4) is never faithful unless G itself is trivial.

Remark (Isomorphic groups have “the same” representations). If $G_1 \cong G_2$ via an isomorphism $\phi : G_1 \rightarrow G_2$ (with inverse $\phi^{-1} : G_2 \rightarrow G_1$), then every representation of one group immediately produces a representation of the other, simply by pre-composing with ϕ or ϕ^{-1} :

- if Ψ is a representation of G_2 , then $\Psi \circ \phi$ is a representation of G_1 (check: $(\Psi \circ \phi)(g_1 g'_1) = \Psi(\phi(g_1)\phi(g'_1)) = \Psi(\phi(g_1))\Psi(\phi(g'_1))$, using first that ϕ is a homomorphism and then that Ψ is);
- likewise, if X is a representation of G_1 , then $X \circ \phi^{-1}$ is a representation of G_2 .

So isomorphic groups have, in a precise sense, identical representation theories — another illustration of why isomorphic groups are treated as “the same group” in practice.

10.2 Kernel and image

Definition 10.3 (Kernel). Let $\phi : G_1 \rightarrow G_2$ be a homomorphism. The *kernel* of ϕ is

$$\ker(\phi) := \{ g_1 \in G_1 : \phi(g_1) = e_2 \},$$

the set of elements of G_1 sent to the identity of G_2 .

Definition 10.4 (Image). The *image* of ϕ is

$$\text{im}(\phi) := \{ g_2 \in G_2 : \exists g_1 \in G_1 \text{ with } \phi(g_1) = g_2 \},$$

i.e. the set of elements of G_2 that are actually hit by ϕ .

Proposition 10.1. $\ker(\phi)$ is a normal subgroup of G_1 , and $\text{im}(\phi)$ is a subgroup of G_2 .

Proof. $\ker(\phi)$ is a subgroup: let $k_1, k_2 \in \ker \phi$, so $\phi(k_1) = \phi(k_2) = e_2$. Then $\phi(k_1 k_2) = \phi(k_1)\phi(k_2) = e_2 e_2 = e_2$, so $k_1 k_2 \in \ker \phi$ (closure). First, applying ϕ to $e_1 \cdot k = k$ (with $k \in \ker \phi$ fixed) gives $\phi(e_1)\phi(k) = \phi(k)$, i.e. $\phi(e_1)e_2 = e_2$, forcing $\phi(e_1) = e_2$, i.e. $e_1 \in \ker \phi$ (identity). Now, using $\phi(e_1) = e_2$ just established, applying ϕ to $e_1 = k^{-1}k$ gives $e_2 = \phi(k^{-1})\phi(k) = \phi(k^{-1})e_2$, so $\phi(k^{-1}) = e_2$, i.e. $k^{-1} \in \ker \phi$ (inverses).

$\ker(\phi)$ is normal: for $g \in G_1$ and $k \in \ker \phi$,

$$\phi(g \circ k \circ g^{-1}) = \phi(g)\phi(k)\phi(g^{-1}) = \phi(g)e_2\phi(g)^{-1} = e_2,$$

so $g \circ k \circ g^{-1} \in \ker \phi$ for every $g \in G_1$: $\ker \phi$ is invariant under conjugation by all of G_1 , i.e. normal.

$\text{im}(\phi)$ is a subgroup: for $\phi(g_1), \phi(g_2) \in \text{im} \phi$, $\phi(g_1)\phi(g_2) = \phi(g_1 g_2) \in \text{im} \phi$ (closure); $e_2 = \phi(e_1) \in \text{im} \phi$ (identity); and $\phi(g_1)^{-1} = \phi(g_1^{-1}) \in \text{im} \phi$ (inverse, using $\phi(g_1)\phi(g_1^{-1}) = \phi(g_1 g_1^{-1}) = \phi(e_1) = e_2$). $\square$

The kernel measures the failure of ϕ to be injective: two elements $g_1, g'_1 \in G_1$ are sent to the same element of G_2 , $\phi(g_1) = \phi(g'_1)$, exactly when $g_1^{-1}g'_1 \in \ker \phi$, i.e. exactly when g_1 and g'_1 lie in the same coset of $\ker \phi$. This observation is the heart of the following fundamental theorem.

10.3 The First Homomorphism (Isomorphism) Theorem

Theorem 10.2 (First Homomorphism Theorem). Let $\phi : G \rightarrow G'$ be a homomorphism with kernel $K = \ker(\phi)$. Then:

- (a) K is a normal subgroup of G .
- (b) Every element of a left coset gK is mapped by ϕ to the same element of G' .
- (c) The map $\Theta : G/K \rightarrow \text{im}(\phi)$ defined by $\Theta(gK) := \phi(g)$ is well defined and is an isomorphism:

$$G/K \cong \text{im}(\phi) \subseteq G'.$$

Proof. (a) This is Proposition 10.1.

(b) If $g' \in gK$, then $g' = g \circ k$ for some $k \in K$, so $\phi(g') = \phi(g)\phi(k) = \phi(g)e' = \phi(g)$: every representative of the coset gK gives the same value of ϕ .

(c) By (b), $\Theta(gK) := \phi(g)$ does not depend on the choice of representative g of the coset gK , so Θ is a well-defined map $G/K \rightarrow G'$; its image is visibly $\text{im}(\phi)$ by definition of the image. Θ is a homomorphism:

$$\Theta(g_1K)\Theta(g_2K) = \phi(g_1)\phi(g_2) = \phi(g_1g_2) = \Theta((g_1g_2)K) = \Theta(g_1K \circ g_2K).$$

Θ is injective: if $\Theta(g_1K) = \Theta(g_2K)$, i.e. $\phi(g_1) = \phi(g_2)$, then $\phi(g_2^{-1}g_1) = \phi(g_2)^{-1}\phi(g_1) = e'$, so $g_2^{-1}g_1 \in K$, i.e. $g_1 \in g_2K$, i.e. $g_1K = g_2K$ (using Corollary 8.2): distinct cosets are sent to distinct elements of G' . And Θ is (by construction) surjective onto $\text{im}(\phi)$. A bijective homomorphism is an isomorphism, so $G/K \cong \text{im}(\phi)$. $\square$

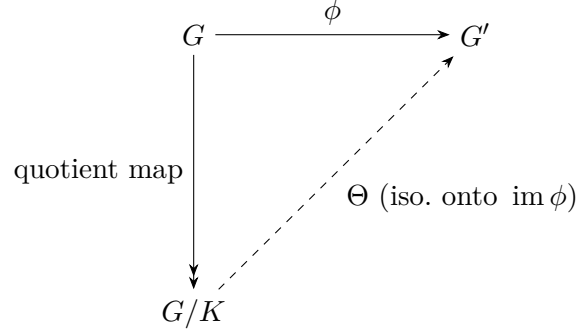


Figure 7: The First Homomorphism Theorem: every homomorphism ϕ factors as an (onto) quotient map followed by an isomorphism Θ onto its image, $G/\ker \phi \cong \text{im} \phi$.

Corollary 10.3. A homomorphism $\phi : G \rightarrow G'$ is injective (hence an isomorphism onto its image) if and only if $\ker(\phi) = \{e\}$.

Proof. If $\ker \phi = \{e\}$, the cosets of $K = \{e\}$ are singletons $\{g\}$, so $G/K \cong G$ literally, and Theorem 10.2(c) becomes $G \cong \text{im}(\phi)$ via ϕ itself, i.e. ϕ is injective. Conversely, if ϕ is injective, then $\phi(k) = e'$ for $k \in \ker \phi$ forces $k = e$ (since also $\phi(e) = e'$ and ϕ is injective), so $\ker \phi = \{e\}$. $\square$

Example 10.2 (Cayley's theorem, revisited). The injective homomorphism $g' \mapsto \sigma_{g'}$ of Theorem 5.3 has trivial kernel by Corollary 10.3 (we checked this directly in the proof of Theorem 5.3), consistent with the conclusion $G \cong \text{im}(\phi) \leq S_{|G|}$.

Example 10.3 (The determinant homomorphism). The determinant $\det : \text{GL}(N, \mathbb{C}) \rightarrow (\mathbb{C} \setminus \{0\}, \cdot)$ is a homomorphism, since $\det(AB) = \det(A)\det(B)$. Its kernel is exactly $SU(N)$ -like: restricted to $U(N)$, $\ker(\det) = SU(N)$; and $\det$ maps $U(N)$ *onto* the unit circle $S^1 \subset \mathbb{C} \setminus \{0\}$ (since every unitary matrix has $|\det U| = 1$, and every phase is achieved, e.g. by $\text{diag}(e^{i\theta}, 1, \dots, 1)$). The First Homomorphism Theorem therefore gives

$$U(N)/SU(N) \cong S^1 \cong U(1),$$

a fact that will resurface, sharpened, when we discuss direct and semidirect products of $U(1)$ with $SU(N)$ in Section 12.

Physical perspective

The First Homomorphism Theorem is the abstract statement behind a practical slogan used throughout physics: “two configurations related by a kernel-element are physically indistinguishable”. If $\phi : G \rightarrow G'$ describes how the group G of some redundant description (e.g. gauge transformations, or field redefinitions) acts on physical observables in G' , then $\ker \phi$ is the subgroup of *purely redundant* transformations (gauge transformations that act trivially on all observables), and the theorem guarantees that the physically distinguishable transformations form the honest group $G/\ker \phi$, isomorphic to the image $\text{im } \phi \subseteq G'$ that is actually realised on observables.

11 Automorphisms and the Centre

Definition 11.1 (Automorphism). An *automorphism* of G is an isomorphism $\phi : G \rightarrow G$ from G to itself. If ϕ is conjugation by a fixed element $x \in G$,

$$\phi_x : G \rightarrow G, \quad \phi_x(g) = x \circ g \circ x^{-1},$$

it is called an *inner automorphism*; any automorphism not of this form is called an *outer automorphism*.

The set $\text{Aut}(G)$ of all automorphisms of G is itself a group under composition (composing two structure-preserving bijections $G \rightarrow G$ gives another one), and the inner automorphisms form a normal subgroup $\text{Inn}(G) \trianglelefteq \text{Aut}(G)$ (an exercise the reader may enjoy attempting using the tools of Section 7). Conjugation by x is, in effect, “relabelling the group from the point of view of x ”, which is exactly the operation used to define conjugacy classes in Section 6; the conjugacy class of h is simply the orbit of h under all inner automorphisms.

Example 11.1. In an abelian group, $\phi_x(g) = xgx^{-1} = xx^{-1}g = g$ for every x , so the *only* inner automorphism is the identity map; any non-trivial symmetry of an abelian group’s multiplication table (e.g. $g \mapsto g^{-1}$ for $\mathbb{Z}_N$, which is a genuine automorphism whenever it is not the identity) is necessarily outer.

11.1 The centre of a group

Definition 11.2 (Centre). The *centre* of G is the set of elements that commute with *everything* in G :

$$Z(G) := \{ g \in G : g \circ h = h \circ g \ \forall h \in G \}.$$

Proposition 11.1. $Z(G)$ is a normal subgroup of G .

Proof. Subgroup: if $g_1, g_2 \in Z(G)$ then for any $h \in G$, $(g_1g_2)h = g_1(g_2h) = g_1(hg_2) = (g_1h)g_2 = (hg_1)g_2 = h(g_1g_2)$, so $g_1g_2 \in Z(G)$; $e \in Z(G)$ trivially; and if $g \in Z(G)$ then from $gh = hg$ we get $h = g^{-1}hg$, i.e. $hg^{-1} = g^{-1}h$ for all h , so $g^{-1} \in Z(G)$. *Normal:* for $g \in Z(G)$ and any $x \in G$, $x \circ g \circ x^{-1} = g \circ x \circ x^{-1} = g \in Z(G)$ (using that g commutes with x), so conjugating an element of the centre simply returns that same element: every element of $Z(G)$ is a conjugacy class all by itself, so in particular $Z(G)$ is trivially a union of complete conjugacy classes, and Theorem 7.1 shows it is normal. $\square$

Example 11.2 (Centre of $SU(N)$). A matrix that commutes with *every* element of $SU(N)$ must, by Schur’s lemma (a fact usually proven once irreducible representations have been discussed, but easy to verify directly here for the defining representation), be proportional to the identity, $g = \lambda \mathbb{1}_N$. Unitarity forces $|\lambda| = 1$, and $\det g = 1$ forces $\lambda^N = 1$. Hence

$$Z(SU(N)) = \left\{ e^{2\pi i k/N} \mathbb{1}_N : k = 0, 1, \dots, N-1 \right\} \cong \mathbb{Z}_N.$$

For example $Z(SU(2)) = \{+\mathbb{1}_2, -\mathbb{1}_2\} \cong \mathbb{Z}_2$: this is the group-theoretic origin of the “sign ambiguity” relating $SU(2)$ to $SO(3)$ mentioned in Example 3.5 and in the belt-trick discussion of Example 1.1 — indeed one can show $SO(3) \cong SU(2)/Z(SU(2))$, an instance of the First Homomorphism Theorem (Theorem 10.2) applied to the two-to-one homomorphism $SU(2) \rightarrow SO(3)$ alluded to in Exercise 4.1.

Example 11.3 (Centre of D_4). For the symmetry group of the square (Example 3.4), the centre is $Z(D_4) = \{M_1, M_3\} = \{\mathbb{1}, -\mathbb{1}\} \cong \mathbb{Z}_2$: the identity and the 180° rotation are the only symmetries that commute with *all* eight elements (a 180° rotation of the plane commutes with every linear map, since $-\mathbb{1}$ is a scalar matrix), whereas a 90° rotation does not commute with a reflection, as we saw explicitly at the end of Section 3.3.

Physical perspective

The centre of the gauge group often has direct physical significance: for instance $Z(SU(N)) \cong \mathbb{Z}_N$ controls the possible “electric” and “magnetic” charges (via the $\mathfrak{k}$ Hooft flux and confinement criteria) of $SU(N)$ gauge theories, and more mundanely, the centre of the Standard Model gauge group $SU(3) \times SU(2) \times U(1)$ (quotiented appropriately) is responsible for the precise global structure of the gauge group actually realised in nature, distinguishing, e.g., quarks and leptons by how they transform under it.

12 Direct and Semidirect Products

We finally address the question raised in Section 9: under what conditions can a group G be reconstructed, as a product, from two of its subgroups? There are two answers, of increasing generality: the *direct product* (when both subgroups are normal and otherwise independent) and the *semidirect product* (when only one of them is normal). The semidirect product is, for physics, arguably the single most important construction in this entire set of notes, because the Euclidean group and the Poincaré group — the symmetry groups of non-relativistic and relativistic space-time, respectively — are both semidirect products of a rotation (or Lorentz) group with a translation group.

12.1 The direct product

Definition 12.1 (Direct product). Given two groups G_1, G_2 , their *direct product* $G_1 \times G_2$ is the set of ordered pairs (g_1, g_2) , $g_1 \in G_1$, $g_2 \in G_2$, with componentwise multiplication:

$$(g_1, g_2) \circ (g'_1, g'_2) := (g_1 \circ g'_1, g_2 \circ g'_2).$$

It is straightforward to check that this defines a group structure on $G_1 \times G_2$: associativity and closure are immediate componentwise; the identity is (e_1, e_2) ; and the inverse of (g_1, g_2) is (g_1^{-1}, g_2^{-1}) .

Inside $G_1 \times G_2$ we find two natural subgroups,

$$\tilde{G}_1 := \{(g_1, e_2) : g_1 \in G_1\} \cong G_1, \quad \tilde{G}_2 := \{(e_1, g_2) : g_2 \in G_2\} \cong G_2,$$

and two properties that will be the template for the general definition below: *every* element of $\tilde{G}_1$ commutes with *every* element of $\tilde{G}_2$,

$$(g_1, e_2) \circ (e_1, g_2) = (g_1, g_2) = (e_1, g_2) \circ (g_1, e_2),$$

and $\tilde{G}_1$ and $\tilde{G}_2$ share only the identity element, $\tilde{G}_1 \cap \tilde{G}_2 = \{(e_1, e_2)\}$.

Example 12.1. $\mathbb{Z}_2 \times \mathbb{Z}_2 = \{(e, e), (e, g), (g, e), (g, g)\}$ is the group already used, as a cautionary counter-example, in Section 9: it is abelian, has order 4, and every non-identity element has order 2 (squares to the identity) — in sharp contrast to $\mathbb{Z}_4$, whose generator has order 4. This is the smallest example showing that two groups of the same (finite) order need not be isomorphic.

The definition above builds a new group $G_1 \times G_2$ out of two given groups (an *external* direct product). It is at least as useful to run the construction backwards: given a single group G that we suspect decomposes as a product of two of its own subgroups (an *internal* direct product), when can we conclude $G \cong G_1 \times G_2$?

Definition 12.2 (Direct product group). If $G \cong G_1 \times G_2$ for some subgroups $G_1, G_2 \leq G$, G is called a *direct product group*.

Theorem 12.1. Suppose G possesses two subgroups G_1, G_2 such that:

- (i) every element of G_1 commutes with every element of G_2 ;
- (ii) G_1 and G_2 have only the identity element in common, $G_1 \cap G_2 = \{e\}$;
- (iii) every $g \in G$ can be written as $g = g_1 \circ g_2$ for some $g_1 \in G_1, g_2 \in G_2$ (necessarily uniquely, given (ii)).

Then G is a direct product group, $G \cong G_1 \times G_2$; in particular, both G_1 and G_2 are automatically normal subgroups of G .

Sketch. Uniqueness of the decomposition in (iii) follows from (ii): if $g_1 g_2 = g'_1 g'_2$ then $g_1^{-1} g'_1 = g'_2 g_2^{-1} \in G_1 \cap G_2 = \{e\}$, forcing $g_1 = g'_1$ and $g_2 = g'_2$. The map $(g_1, g_2) \mapsto g_1 \circ g_2$ is then a well-defined bijection $G_1 \times G_2 \rightarrow G$, and it is a homomorphism precisely because of (i): $(g_1 g_2)(g'_1 g'_2) = g_1(g_2 g'_1)g'_2 = g_1(g'_1 g_2)g'_2 = (g_1 g'_1)(g_2 g'_2)$, which matches the direct-product multiplication rule $(g_1, g_2) \circ (g'_1, g'_2) = (g_1 g'_1, g_2 g'_2)$. Normality of G_1 (and likewise G_2) follows since conjugating $g_1 \in G_1$ by an arbitrary $g = g'_1 g'_2 \in G$ gives, using (i) to move g'_2 past g_1 ,

$$g g_1 g^{-1} = g'_1 g'_2 g_1 (g'_2)^{-1} (g'_1)^{-1} = g'_1 g_1 g'_2 (g'_2)^{-1} (g'_1)^{-1} = g'_1 g_1 (g'_1)^{-1} \in G_1. \quad \square$$

Example 12.2 ($O(3) = \mathbb{Z}_2 \times SO(3)$). The full orthogonal group $O(3) = \{O \in \mathbb{R}^{3 \times 3} : O^T O = \mathbb{1}_3\}$ contains matrices of determinant +1 (rotations, forming $SO(3)$) and determinant -1 (improper rotations, i.e. a rotation combined with a spatial inversion). Set

$$G_1 = SO(3), \quad G_2 = \{\mathbb{1}_3, -\mathbb{1}_3\} \cong \mathbb{Z}_2.$$

Since $-\mathbb{1}_3$ is a scalar matrix, it commutes with *every* 3×3 matrix, so condition (i) of Theorem 12.1 holds; clearly $SO(3) \cap \{\pm \mathbb{1}_3\} = \{\mathbb{1}_3\}$, giving (ii); and every $O \in O(3)$ can be written as $O = (\pm \mathbb{1}_3) \cdot (\pm O)$ with $\pm O \in SO(3)$ chosen to have determinant +1, giving (iii).

Hence

$$O(3) \cong \mathbb{Z}_2 \times SO(3).$$

Physically, this says that “parity” (spatial inversion, $\mathbb{Z}_2$) and proper rotations ($SO(3)$) act completely independently on physical systems — exactly the statement used when one classifies particle states by both spin (an $SO(3)$, or rather $SU(2)$, representation) and intrinsic parity (± 1).

Remark (A near-miss: $U(N)$ is *not* $U(1) \times SU(N)$). One might expect, by analogy with Example 12.2, that factoring out the overall phase of a unitary matrix would similarly split $U(N)$ as a direct product of $U(1)$ (the phase) and $SU(N)$ (the “rest”). This fails:

$$U(N) \not\cong U(1) \times SU(N).$$

The subtlety is condition (ii) of Theorem 12.1: a natural candidate for the $U(1)$ subgroup is $\{e^{i\theta} \mathbb{1}_N : \theta \in [0, 2\pi)\}$, and this does commute with all of $SU(N)$ (scalar matrices commute with everything, exactly as in Example 12.2), so condition (i) holds. But this candidate $U(1)$ subgroup *intersects* $SU(N)$ non-trivially: $e^{i\theta} \mathbb{1}_N \in SU(N)$ whenever $e^{iN\theta} = 1$, i.e. for the N values $\theta = 2\pi k/N$, $k = 0, \dots, N-1$ — exactly the centre $Z(SU(N)) \cong \mathbb{Z}_N$ of Example 11.2. So condition (ii) fails: $U(1) \cap SU(N) \cong \mathbb{Z}_N \neq \{e\}$. What *is* true is the weaker (but still very useful) statement, obtained from the surjective homomorphism $U(1) \times SU(N) \rightarrow U(N)$, $(\lambda, S) \mapsto \lambda S$, whose kernel is exactly the diagonally embedded $\mathbb{Z}_N$ (pairs $(\lambda, S) = (\omega, \omega^{-1} \mathbb{1}_N)$ with $\omega^N = 1$), via the First Homomorphism Theorem (Theorem 10.2):

$$U(N) \cong (U(1) \times SU(N)) / \mathbb{Z}_N.$$

This is exactly the kind of global-structure subtlety that matters, for instance, in identifying the precise gauge group of the Standard Model.

12.2 The semidirect product

Condition (i) of Theorem 12.1 — that *every* element of G_1 commutes with *every* element of G_2 — is very restrictive. If we only ask G_1 to be normal (so that conjugation by G_2 maps G_1 to itself, without requiring the conjugation to act trivially), we obtain a strictly more general, and for physics much more important, construction.

Definition 12.3 (Semidirect product). Suppose G has two subgroups G_1, G_2 such that:

- (i) G_1 is a normal subgroup of G ;
 - (ii) G_1 and G_2 have only the identity in common, $G_1 \cap G_2 = \{e\}$;
 - (iii) every $g \in G$ can be written (necessarily uniquely) as $g = g_1 \circ g_2$ with $g_1 \in G_1$, $g_2 \in G_2$.
- Then G is called the *semidirect product* of G_1 and G_2 , written $G = G_1 \rtimes G_2$.

Notice that this is exactly Theorem 12.1 with condition (i) weakened from “ G_1, G_2 commute elementwise” to “ G_1 is normal”; a direct product is the special case of a semidirect product in which, in addition, G_2 also happens to be normal (and then G_1, G_2 automatically commute elementwise, as one can check). In a semidirect product, G_2 acts on G_1 by conjugation — $g_2 \circ g_1 \circ g_2^{-1} \in G_1$ for $g_1 \in G_1, g_2 \in G_2$, since G_1 is normal — and this action need not be trivial, which is precisely what happened in Section 9 and is what we now see concretely in the group of rigid motions of space.

12.3 Worked example: the Euclidean group

Example 12.3 (The Euclidean group in $\mathbb{R}^3$). A general rigid motion of $\mathbb{R}^3$ (one that preserves all distances) is a rotation followed by a translation:

$$\vec{x} \mapsto \vec{x}' = O\vec{x} + \vec{t}, \quad O \in O(3), \vec{t} \in \mathbb{R}^3.$$

We write such a transformation as a pair $(O, \vec{t})$. Composing two such motions, first $(O_1, \vec{t}_1)$ and then $(O_2, \vec{t}_2)$,

$$\vec{x}'' = O_2\vec{x}' + \vec{t}_2 = O_2(O_1\vec{x} + \vec{t}_1) + \vec{t}_2 = (O_2O_1)\vec{x} + (O_2\vec{t}_1 + \vec{t}_2),$$

shows that the composition law of the *Euclidean group* $E(3) = O(3) \ltimes \mathbb{R}^3$ is

$$(O_2, \vec{t}_2) \circ (O_1, \vec{t}_1) = (O_2O_1, O_2\vec{t}_1 + \vec{t}_2),$$

with identity $(\mathbb{1}_3, \vec{0})$ and inverse (checked directly)

$$(O, \vec{t})^{-1} = (O^T, -O^T\vec{t}), \quad \text{since} \quad (O^T, -O^T\vec{t}) \circ (O, \vec{t}) = (O^TO, O^T\vec{t} - O^T\vec{t}) = (\mathbb{1}_3, \vec{0}),$$

and likewise $(O, \vec{t}) \circ (O^T, -O^T\vec{t}) = (OO^T, -OO^T\vec{t} + \vec{t}) = (\mathbb{1}_3, \vec{0})$.

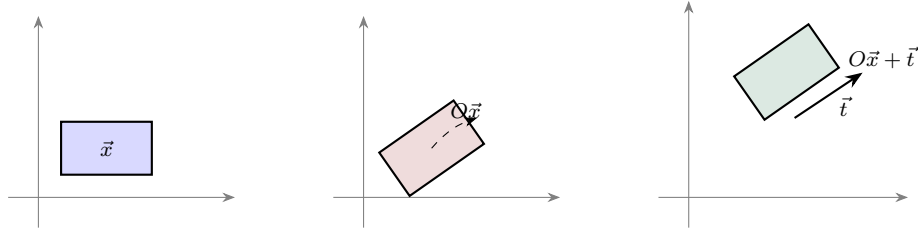


Figure 8: A general Euclidean transformation $(O, \vec{t})$: a rotation O followed by a translation by $\vec{t}$ (Example 12.3).

Example 12.4 ($E(3)$ as a semidirect product). Inside $E(3)$ we identify two natural subgroups,

$$G_1 = \{(\mathbb{1}_3, \vec{t}) : \vec{t} \in \mathbb{R}^3\} \quad (\text{pure translations}), \quad G_2 = \{(O, \vec{0}) : O \in O(3)\} \quad (\text{pure rotations}),$$

each isomorphic to $(\mathbb{R}^3, +)$ and $O(3)$ respectively, intersecting only in the identity $(\mathbb{1}_3, \vec{0})$, and such that every element $(O, \vec{t}) = (\mathbb{1}_3, \vec{t}) \circ (O, \vec{0})$ decomposes uniquely as required. But only G_1 is normal, *not* G_2 : conjugating a translation by a general group element keeps it a translation,

$$(O, \vec{t}) \circ (\mathbb{1}_3, \vec{a}) \circ (O, \vec{t})^{-1} = (O, O\vec{a} + \vec{t}) \circ (O^T, -O^T\vec{t}) = (OO^T, -OO^T\vec{t} + O\vec{a} + \vec{t}) = (\mathbb{1}_3, O\vec{a}) \in G_1,$$

so G_1 is invariant (with the rotation part of the conjugating element acting on the translation vector $\vec{a} \mapsto O\vec{a}$ — exactly the non-trivial conjugation action that distinguishes a semidirect from a direct product). Conjugating a rotation, on the other hand, does *not* stay in G_2 whenever $\vec{t} \neq \vec{0}$:

$$(O, \vec{t}) \circ (R, \vec{0}) \circ (O, \vec{t})^{-1} = (OR, \vec{t}) \circ (O^T, -O^T\vec{t}) = (ORO^T, -ORO^T\vec{t} + \vec{t}),$$

which is *not* a pure rotation (its translation part $-ORO^T\vec{t} + \vec{t}$ is generally non-zero) — physically, “rotating about a different point” is the composition of a rotation with a translation,

so the subgroup of rotations *about the origin specifically* is not preserved when we change reference point by conjugation. Hence

$$E(3) = \mathbb{R}^3 \rtimes O(3),$$

a genuine semidirect, not direct, product: this is the reason rigid motions of space combine less simply than, say, the direct product $O(3) \cong \mathbb{Z}_2 \times SO(3)$ of Example 12.2.

Remark (Matrix representation of $E(3)$). The Euclidean group is conveniently represented by 4×4 matrices acting on $\mathbb{R}^3$ in homogeneous coordinates,

$$g = \begin{pmatrix} O & \vec{t} \\ 0 & 1 \end{pmatrix}, \quad e = \begin{pmatrix} \mathbb{1}_3 & \vec{0} \\ 0 & 1 \end{pmatrix}, \quad g^{-1} = \begin{pmatrix} O^T & -O^T \vec{t} \\ 0 & 1 \end{pmatrix},$$

acting on the augmented vector $(\vec{x}, 1)^T$ by ordinary matrix multiplication, $g \cdot (\vec{x}, 1)^T = (O\vec{x} + \vec{t}, 1)^T$. This matrix form makes the composition law of Example 12.3 completely transparent, since it is now just matrix multiplication, and it is the representation used in practice (e.g. in robotics and computer graphics, as well as in the physics of rigid bodies).

Physical perspective

The Poincaré group of special relativity, $\mathbb{R}^{1,3} \rtimes O(1,3)$ — spacetime translations semidirect with Lorentz transformations — has exactly the same semidirect-product structure as $E(3)$ above, with $O(3)$ replaced by the Lorentz group $O(1,3)$ and $\mathbb{R}^3$ replaced by Minkowski space-time $\mathbb{R}^{1,3}$. Wigner’s classification of relativistic particles by mass and spin is, technically, the classification of the unitary irreducible representations of the Poincaré group, obtained using the semidirect-product structure via what is called the method of induced representations (or “Wigner’s little group” method) — a beautiful and directly physical application of exactly the group-theoretic machinery assembled in this section. This is one of the main destinations of the continuation of this course, once Lie groups and Lie algebras have been introduced.

13 Supplementary Exercises

The exercises below collect and extend the ones marked in the original lecture notes at various points (Examples 3.4 and 3.5 and Exercise 4.1 above already contain several); they are gathered here for convenient reference and practice.

Exercise 13.1. Complete the multiplication (Cayley) table of $D_4 = \{M_1, \dots, M_8\}$ from Example 3.4. Use it to answer: (a) which elements commute with everything (i.e. what is $Z(D_4)$)? (b) which elements are their own inverse? (c) find all subgroups of order 2 and order 4, and check that their orders divide $|D_4| = 8$, consistent with Lagrange’s theorem (Theorem 8.3).

Exercise 13.2. Show directly (i.e. without simply citing Example 3.5) that $SO(N)$ is a group under matrix multiplication, following the pattern of the proof given there for $SU(N)$. Do the same for $O(N)$ and $U(N)$. Which of $SO(N)$, $O(N)$, $SU(N)$, $U(N)$ are abelian, and for which N ?

Exercise 13.3. Let $G = \mathbb{R}^+$ under multiplication and consider the map $\phi(r) = \ln r$. Verify explicitly that this is the inverse isomorphism to the one constructed in Example 4.2, and use it to translate the group axioms of $(\mathbb{R}^+, \cdot)$ into the group axioms of $(\mathbb{R}, +)$ term by term.

Exercise 13.4. Determine the conjugacy classes of D_4 (Example 3.4), and check, as in Theorem 7.1, which unions of conjugacy classes form (normal) subgroups. How many normal subgroups does D_4 have?

Exercise 13.5. Let $H = \{e, (12)\} \leq S_3$. List the left cosets and the right cosets of H in S_3 explicitly, and verify that they do *not* coincide (consistent with Theorem 8.5, since H is not normal in S_3 — compare with the unique normal subgroup A_3 found in Section 7).

Exercise 13.6. Redo the computation of Example 9.2 for $G = \mathbb{Z}_6$ and $H = \{e, g^3\} \cong \mathbb{Z}_2$: list the cosets of H , write the multiplication table of $\mathbb{Z}_6/H$, and identify this quotient group among the groups of order 3.

Exercise 13.7. Verify the two conjugation computations of Example 12.4 using the 4×4 matrix representation of the Euclidean group given at the end of Section 12.3, and confirm that the translation subgroup is preserved while the rotation subgroup is not (unless $\vec{t} = \vec{0}$).

Exercise 13.8. Show that $\text{Inn}(G)$, the group of inner automorphisms of G (Section 11), is isomorphic to $G/Z(G)$. (Hint: consider the homomorphism $G \rightarrow \text{Aut}(G)$, $x \mapsto \phi_x$, and identify its kernel and image using the First Homomorphism Theorem, Theorem 10.2.)

Part II

Lie Groups: Continuous Symmetries

14 From Discrete to Continuous Symmetries

In Part A we studied groups whose elements could be listed, or at least separated from one another: finite groups such as S_3 or $\mathbb{Z}_n$, and infinite but *discrete* groups. Many of the most important symmetries in physics, however, are *continuous*: a rotation can be by any angle $\varphi \in (-\pi, \pi]$, a boost can be by any rapidity, and the elements of the group are labelled by one or more continuously varying real parameters. Such groups are called **Lie groups**, after the Norwegian mathematician Sophus Lie.

Physical perspective

Continuous symmetries are the backbone of modern theoretical physics. The rotation group $SO(3)$ underlies angular momentum in quantum mechanics; the Lorentz group $SO(3, 1)$ is the symmetry group of special relativity; and internal symmetry groups such as $SU(2)$ and $SU(3)$ organise the particles of the electroweak and strong interactions. Understanding the structure of Lie groups — and, crucially, their *representations* — is therefore indispensable

for a physicist.

A Lie group combines two structures that, at first sight, seem unrelated:

- an *algebraic* structure (it is a group, with an associative multiplication, identity, and inverses), and
- a *topological/analytic* structure (its elements form a manifold, so that one can speak of nearby elements, continuity, and differentiability).

The interplay between these two structures is what makes Lie theory both rich and tractable: because the group multiplication is analytic, the entire (local) structure of the group near the identity is captured by a finite-dimensional *linear* object, the **Lie algebra**, obtained by linearising the group multiplication at the identity element.

15 Lie Groups: Definition and Basic Structure

Definition 15.1. A group G is called a (real) **Lie group of dimension n** if it satisfies the following four conditions.

- (L1) **Faithful representation.** G possesses at least one *faithful* representation by $d \times d$ matrices, i.e. an injective homomorphism $g \mapsto \Gamma(g)$ into $GL(d, \mathbb{C})$ (or $GL(d, \mathbb{R})$). We may then measure the “size” of a group element using the matrix norm

$$\|\Gamma\| = \left(\sum_{i,j=1}^d |\Gamma_{ij}|^2 \right)^{1/2}.$$

- (L2) **Parametrization.** There is a subset $M_G \subseteq \mathbb{R}^n$ (an open neighbourhood of the origin) and a map $\xi = (\xi_1, \dots, \xi_n) \in M_G \mapsto \Gamma(\xi) \in G$ that parametrizes the group elements near the identity by n continuously varying real parameters.
- (L3) **One-to-one near the identity.** For every $\xi \in M_G$ there is exactly one group element $\Gamma(\xi)$, and $\Gamma(\xi) = \mathbb{1}$ (the identity) if and only if $\xi = 0$: the parametrization is one-to-one in a neighbourhood of the identity, with $\xi = 0$ corresponding to e .
- (L4) **Analyticity.** Each matrix entry $\Gamma_{ij}(\xi)$ is an *analytic* function of ξ , i.e. it can be expanded in a convergent power series around $\xi = 0$.

Remark. Condition (L1) allows us to think of G concretely as a set of matrices (even if the “abstract” group was defined differently), while (L2)–(L4) say that, at least near the identity, G looks like an ordinary piece of $\mathbb{R}^n$ with analytic transition functions — precisely the definition of an (analytic) manifold. The number n is called the *dimension* of the Lie group.

Physical perspective

On this definition of a Lie group. The definition (L1)–(L4) is deliberately built around the existence of a chosen *faithful matrix representation* $\Gamma(\xi)$. This is exactly the situation a physicist meets in practice: the Lie groups that appear in mechanics, quantum mechanics, and particle physics — $SO(3)$, $SU(2)$, $SU(3)$, the Lorentz group, and so on — are handed to us *already* as groups of matrices acting on some physical space (configuration space, spin space, an internal “flavour” space). It therefore costs us nothing, and buys a great deal

of concreteness, to build the definition around a matrix representation from the outset: the norm $\|\Gamma\|$, the parametrization $\Gamma(\xi)$, and the analyticity of $\Gamma_{ij}(\xi)$ are all immediately meaningful, computable objects.

This is *not*, however, the definition a mathematician would start from, and it is worth flagging the difference clearly so that it does not cause confusion later. The intrinsic, representation-independent definition treats a Lie group as a set G that is simultaneously:

- an abstract *smooth (differentiable) manifold*, i.e. a space that locally looks like $\mathbb{R}^n$ via smooth coordinate charts, glued together by smooth transition functions — with no reference to any ambient matrix space at all; and
- a *group*, such that the multiplication map $G \times G \rightarrow G$, $(g, g') \mapsto g \circ g'$, and the inversion map $G \rightarrow G$, $g \mapsto g^{-1}$, are both *smooth* maps between manifolds.

This intrinsic definition is more general: it does not presuppose that G can be realised as a group of matrices of some fixed finite size, whereas our definition (L1)–(L4) does (via condition (L1)). For the groups of interest to physicists this distinction rarely bites — a theorem of Ado guarantees that every finite-dimensional (real or complex) Lie algebra admits a faithful matrix representation, and the corresponding statement is true for the connected Lie groups built from them in all the cases we shall meet — but conceptually the manifold-based definition is the correct starting point, since it makes clear that the Lie group structure is intrinsic to G and does not depend on an arbitrary choice of matrices used to describe it. We adopt the matrix-based definition here because it lets us get to concrete computations (Lie algebras, representations, characters) immediately, with the tools of Part A; the intrinsic, manifold-based definition, together with the associated notions of tangent bundles, vector fields, and the exponential map, will be developed properly in the second part of the course.

15.1 The Lie Algebra as the Tangent Space at the Identity

Because $\Gamma(\xi)$ is analytic in ξ , we may Taylor-expand around the identity:

$$\Gamma(\xi) = \mathbb{1} + \sum_{\rho=1}^n \xi_{\rho} \Theta_{\rho} + O(\xi^2), \quad \Theta_{\rho} := \left. \frac{\partial \Gamma}{\partial \xi_{\rho}} \right|_{\xi=0}.$$

The matrices $\Theta_1, \dots, \Theta_n$ are called the **generators** of the Lie group. They span a real, n -dimensional *linear* vector space (the tangent space to G at the identity), called the **Lie algebra** $\mathfrak{g}$ of G .

Definition 15.2. Let G be an n -dimensional Lie group with parametrization $\Gamma(\xi)$ as above. The **Lie algebra** of G is the real vector space

$$\mathfrak{g} = \text{span}_{\mathbb{R}}\{\Theta_1, \dots, \Theta_n\}, \quad \Theta_{\rho} = \left. \frac{\partial \Gamma}{\partial \xi_{\rho}} \right|_{\xi=0}.$$

We tacitly assume, as part of what it means for $\xi \mapsto \Gamma(\xi)$ to be a genuine (as opposed to merely continuous and one-to-one) parametrization, that this map is an *immersion* at the identity, i.e. that its derivative there has full rank n — equivalently, that $\{\Theta_{\rho}\}$ are linearly independent, so that $\mathfrak{g}$ really is n -dimensional, matching the dimension of G . (This regularity condition does not actually follow from one-to-one-ness of Γ alone: a map can be injective and analytic near 0 while still having a singular derivative there, e.g. $\xi \mapsto \xi^3$ on $\mathbb{R}$. It is, however, automatically satisfied by every parametrization one meets in practice, such as $\Gamma(\xi) = \exp(\sum_{\rho} \xi_{\rho} \Theta_{\rho})$ for a fixed basis $\{\Theta_{\rho}\}$.)

15.2 Group Multiplication Near the Identity: Structure Constants

Let $g = \Gamma(\xi)$ and $g' = \Gamma(\xi')$ be two group elements near the identity, and let $g'' = g' \circ g = \Gamma(\xi'')$. Since the parametrization is one-to-one near e , the new parameters ξ'' must be some function of the old ones,

$$\xi''_i = f_i(\xi_1, \dots, \xi_n; \xi'_1, \dots, \xi'_n), \quad i = 1, \dots, n,$$

and by (L4) the functions f_i are analytic. Because $g \circ e = g$ and $e \circ g = g$, we must have the boundary conditions

$$f_i(\xi; 0) = \xi_i, \quad f_i(0; \xi') = \xi'_i.$$

Moreover, since g^{-1} also lies near the identity when g does, the equation $f_i(\xi; \xi') = 0$ (expressing $g \circ g^{-1} = e$) has, by the implicit function theorem (using analyticity of f_i), a unique analytic solution $\xi'(\xi)$, whose linearisation follows directly from the boundary conditions above ($f_i(\xi; \xi') = \xi_i + \xi'_i + O(\xi\xi')$, so setting this to zero gives, to leading order):

$$\xi'_i = -\xi_i + O(\xi^2),$$

i.e. the parameter of g^{-1} is, to leading order, simply $-\xi$, as one would expect. Expanding f_i to second order in ξ, ξ' around the origin generates the famous *structure constants* of the Lie algebra, which encode the (non-commutative, in general) multiplication law $[\Theta_a, \Theta_b] = \sum_c f_{ab}^c \Theta_c$ — the subject of a more detailed treatment in the Lie algebra chapter proper. Here we only stress the key conceptual point:

Remark. The entire local structure of the Lie group multiplication — an a priori complicated analytic function of $2n$ variables — is determined, to leading non-trivial order, by the finite-dimensional linear data of the Lie algebra $\mathfrak{g}$. This is the sense in which “the Lie algebra linearises the Lie group”.

16 Key Examples of Lie Groups

Example 16.1. Example 1: The positive reals $(\mathbb{R}^+, \cdot)$

Take $G = (\mathbb{R}^+, \cdot)$, with identity $e = 1$. A faithful (here, 1×1) representation is $\Gamma(x) = x$ itself, or equivalently we may parametrize $x = e^\xi$ with $\xi \in \mathbb{R}$, so that $\Gamma(\xi) = e^\xi$ and group multiplication becomes ordinary addition of the parameter ξ : $\Gamma(\xi)\Gamma(\xi') = e^{\xi+\xi'} = \Gamma(\xi + \xi')$. This is a 1-dimensional, abelian Lie group; its Lie algebra is the 1-dimensional space spanned by $\Theta = \left. \frac{d}{d\xi} e^\xi \right|_{\xi=0} = 1$.

Example 16.2 (Rotations in the plane, $O(2)$ and $SO(2)$). The orthogonal group $O(2)$ consists of real 2×2 matrices D satisfying $D^T D = \mathbb{1}_2$; equivalently, writing $D = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, the conditions are $a^2 + b^2 = c^2 + d^2 = 1$ and $ac + bd = 0$. Solving these, one finds two families:

$$D(\varphi) = \begin{pmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{pmatrix}, \det D = +1 \quad \text{or} \quad D'(\varphi) = \begin{pmatrix} \cos \varphi & \sin \varphi \\ \sin \varphi & -\cos \varphi \end{pmatrix}, \det D = -1.$$

The determinant-+1 matrices form the subgroup $SO(2) = \{D(\varphi) : \varphi \in (-\pi, \pi]\}$, a 1-dimensional, abelian, *compact* Lie group (parametrized by a single angle ranging over a bounded, periodic interval). Every element of $O(2)$ can be written as an element of $SO(2)$ multiplied by the fixed matrix $A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ (a reflection), so $O(2) = SO(2) \cup A \cdot SO(2)$.

The Lie algebra of $SO(2)$ is spanned by the single generator

$$\Theta = \left. \frac{dD}{d\varphi} \right|_{\varphi=0} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

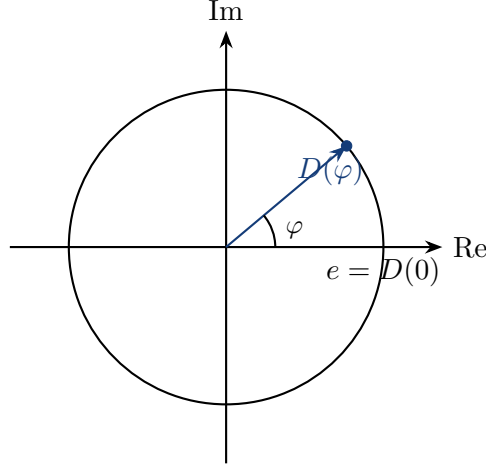


Figure 9: $SO(2)$ is the group of rotations of the plane, parametrized by the angle $\varphi \in (-\pi, \pi]$. Geometrically it is a circle: a compact, connected, 1-dimensional Lie group.

Example 16.3 (The special unitary group $SU(2)$). $SU(2)$ consists of 2×2 complex matrices u with $u^\dagger u = \mathbb{1}_2$ and $\det u = +1$. Writing

$$u = \begin{pmatrix} \alpha & \beta \\ -\beta^* & \alpha^* \end{pmatrix}, \quad \alpha, \beta \in \mathbb{C},$$

the unitarity condition $u^\dagger u = \mathbb{1}_2$ reduces to the single real condition

$$|\alpha|^2 + |\beta|^2 = 1.$$

Writing $\alpha = a_0 + ia_3$ and $\beta = a_2 + ia_1$ with $a_0, a_1, a_2, a_3 \in \mathbb{R}$, this condition becomes $a_0^2 + a_1^2 + a_2^2 + a_3^2 = 1$: *topologically, $SU(2)$ is the 3-sphere S^3 sitting inside $\mathbb{R}^4$.* In particular $SU(2)$ is a compact, connected, 3-dimensional Lie group.

The Lie algebra $\mathfrak{su}(2)$ itself consists of the *anti-Hermitian, traceless* 2×2 matrices $\Theta_a = \frac{i}{2}\sigma_a$, built from the three Hermitian, traceless Pauli matrices,

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

via $u(\vec{a}) \approx \mathbb{1}_2 + \vec{a} \cdot \vec{\Theta} = \mathbb{1}_2 + \frac{i}{2}\vec{a} \cdot \vec{\sigma}$ near the identity: it is the anti-Hermitian combinations $\Theta_a = \frac{i}{2}\sigma_a$, not the Hermitian σ_a themselves, that form a real vector space closed under commutation (as any Lie algebra must). In the “physics convention” one instead works with the Hermitian generators $T_a = \frac{1}{2}\sigma_a = -i\Theta_a$ and writes $u(\vec{a}) \approx \mathbb{1}_2 + i\vec{a} \cdot \vec{T}$, trading the anti-Hermitian Θ_a (which exponentiate directly to unitary matrices) for Hermitian T_a (which are the “observables” physicists want, at the price of an explicit factor of i in the exponent).

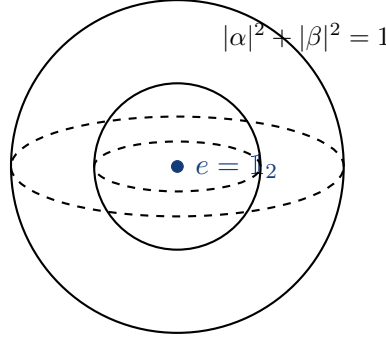


Figure 10: Schematic depiction of $SU(2) \cong S^3$: every group element $u(\alpha, \beta)$ corresponds to a point on the unit 3-sphere $|\alpha|^2 + |\beta|^2 = 1$ in $\mathbb{C}^2 \cong \mathbb{R}^4$. The picture shows the usual reduced-dimension cross-section used to visualise S^3 (concentric circles represent circles of latitude on the sphere).

Example 16.4 (The Lorentz group in 1 + 1 dimensions). Consider linear transformations Λ of 2-dimensional Minkowski space that preserve the metric $\eta = \text{diag}(1, -1)$: $\Lambda^T \eta \Lambda = \eta$. Writing $\Lambda = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, this condition becomes $a^2 - c^2 = 1$, $b^2 - d^2 = -1$, $ab - cd = 0$, which is solved by

$$a = \cosh \chi, \quad b = \sinh \chi, \quad c = \sinh \chi, \quad d = \cosh \chi,$$

i.e. the familiar Lorentz boost

$$\Lambda(\chi) = \begin{pmatrix} \cosh \chi & \sinh \chi \\ \sinh \chi & \cosh \chi \end{pmatrix}, \quad \begin{pmatrix} t' \\ x' \end{pmatrix} = \Lambda(\chi) \begin{pmatrix} t \\ x \end{pmatrix},$$

(with the metric $\eta = \text{diag}(1, -1)$ acting on components ordered (t, x) , so that the preserved quadratic form reads $t^2 - x^2$, matching $a^2 - c^2 = 1$ etc. above), with $\chi \in \mathbb{R}$ the *rapidity*. Unlike $SO(2)$, this parameter ranges over the whole real line rather than a bounded interval: the Lorentz group (even in this simple 1 + 1-dimensional form) is a **non-compact** Lie group.

Physical perspective

Compactness has an important physical consequence, spelled out in Theorem 20.2 below: compact Lie groups (like $SO(3)$, $SU(2)$, $SU(3)$) always admit finite-dimensional *unitary* representations, which is why we can represent angular momentum or isospin by finite-dimensional unitary matrices. Non-compactness alone does not imply the opposite — e.g. the non-compact abelian group $(\mathbb{R}, +)$ still has plenty of finite-dimensional (one-dimensional) unitary representations $\chi \mapsto e^{ik\chi}$ — but for a connected *simple* non-compact Lie group such as $SO^+(3, 1)$, the full (non-abelian, semisimple) Lorentz group, one can show there is *no* non-trivial finite-dimensional *unitary* representation: one reason relativistic quantum field theory must be formulated more subtly (using infinite-dimensional unitary representations, or finite-dimensional but non-unitary ones, such as those acting on spinor or tensor fields).

16.1 Discrete Symmetries and the Four Components of the Lorentz Group

If we also allow the determinant to be -1 and reversals of time or space orientation, we obtain additional disconnected pieces of the full Lorentz group in 1 + 1 dimensions:

$$\Lambda(\chi) = \begin{pmatrix} \cosh \chi & \sinh \chi \\ \sinh \chi & \cosh \chi \end{pmatrix}, \quad P\Lambda(\chi) = \begin{pmatrix} \cosh \chi & \sinh \chi \\ -\sinh \chi & -\cosh \chi \end{pmatrix}, \quad T\Lambda(\chi), \quad PT\Lambda(\chi),$$

where $P = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ is a parity (space) reflection and $T = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$ is time reversal (recall components are ordered (t, x) , so P flips x and T flips t , as their names require). These four families are distinguished by the sign of $\det \Lambda$ and the sign of Λ^0_0 (whether time orientation is preserved), giving four disjoint pieces:

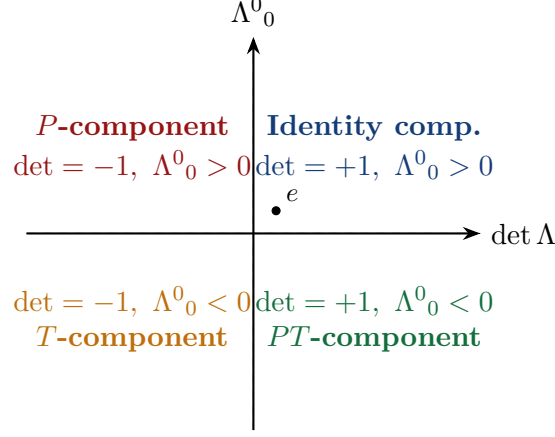


Figure 11: The full Lorentz group (even in $1 + 1$ dimensions) has four connected components, distinguished by $\det \Lambda = \pm 1$ and the sign of Λ^0_0 . Only the component containing the identity (upper right) is itself a subgroup (indeed a normal subgroup, Theorem 17.1); the other three are its cosets, obtained by multiplying by P , T , or PT .

17 Connected Components

Definition 17.1. Two elements g, g' of a Lie group G lie in the same **connected component** if g' can be reached from g by continuously varying the parameters of a faithful representation $\Gamma_{ij}(\xi)$ (i.e. there is a continuous path of group elements joining them). The connected component containing the identity is denoted G_0 (or S).

Example 17.1. $(\mathbb{R} \setminus \{0\}, \cdot)$ has two connected components: the positive reals ($r > 0$) and the negative reals ($r < 0$); the value $r = 0$ is simply excluded, since it does not lie in the group at all. Likewise $O(2) = SO(2) \cup A \cdot SO(2)$ has two connected components, distinguished by $\det D = \pm 1$; and, as we saw above, the full Lorentz group in $1+1$ dimensions has four connected components.

Theorem 17.1. (The identity component is a normal subgroup.) Let S be the connected component of a Lie group G containing the identity e . Then S is itself a normal (invariant) subgroup of G .

Proof. S is a subgroup. We check the subgroup criterion directly: S is non-empty (it contains e), and we must show that $s, s' \in S$ implies $s \circ s'^{-1} \in S$.

By definition of S , there exist continuous paths $\gamma, \gamma' : [0, 1] \rightarrow G$ with $\gamma(0) = \gamma'(0) = e$ and $\gamma(1) = s, \gamma'(1) = s'$, lying entirely in G (i.e. each $\gamma(t), \gamma'(t)$ is itself a group element depending continuously on t). Since group multiplication $(g, g') \mapsto g \circ g'$ and inversion $g \mapsto g^{-1}$ are continuous (indeed analytic, by (L4) together with the inverse-function argument of Section 15),

the composite path

$$\eta(t) := \gamma(t) \circ \gamma'(t)^{-1}, \quad t \in [0, 1],$$

is itself a continuous path of group elements, with $\eta(0) = e \circ e^{-1} = e$ and $\eta(1) = s \circ s'^{-1}$. Hence η is a continuous path joining e to $s \circ s'^{-1}$, which by definition means $s \circ s'^{-1} \in S$. This proves S is a subgroup.

S is normal. Let $x \in G$ be arbitrary, and consider the conjugation map $c_x : G \rightarrow G$, $c_x(g) = xgx^{-1}$, which is continuous (indeed analytic) since it is built by composing the (analytic) group multiplication with itself. If $\gamma : [0, 1] \rightarrow G$ is a continuous path with $\gamma(0) = e$, $\gamma(1) = s \in S$, then $c_x \circ \gamma$ is again a continuous path, now with

$$(c_x \circ \gamma)(0) = x e x^{-1} = e, \quad (c_x \circ \gamma)(1) = x s x^{-1}.$$

So $c_x \circ \gamma$ is a continuous path joining e to $x s x^{-1}$, showing $x s x^{-1} \in S$ for every $s \in S$. As $x \in G$ was arbitrary, $x S x^{-1} \subseteq S$ for every x ; applying the same argument with x^{-1} in place of x gives $x^{-1} S x \subseteq S$, i.e. $S \subseteq x S x^{-1}$, and combining the two inclusions gives $x S x^{-1} = S$ for every $x \in G$. Hence S is normal. $\square$

Corollary 17.2. Every connected component of G is a (left) coset of S : either S itself, or gS for some $g \notin S$. In particular, every connected component is, geometrically, a copy of the Lie group S (translated by g), and all connected components of G have the same dimension as G .

Proof. Let C be any connected component of G , and pick some $g \in C$. We claim $C = gS$.

($gS \subseteq C$.) For any $s \in S$, there is a continuous path γ from e to s lying in G ; then $g \circ \gamma$ (left-translating the whole path by g , again continuous since group multiplication is continuous) is a continuous path from $g \circ e = g$ to $g \circ s$, lying in G . Hence g and gs lie in the same connected component, which is C (since $g \in C$ by assumption); so $gs \in C$. As $s \in S$ was arbitrary, $gS \subseteq C$.

($C \subseteq gS$.) Conversely, if $h \in C$, then g and h lie in the same connected component, so there is a continuous path δ from g to h . Left-translating by g^{-1} , the path $g^{-1} \circ \delta$ is continuous and runs from $g^{-1} \circ g = e$ to $g^{-1} \circ h$, showing $g^{-1}h \in S$, i.e. $h \in gS$.

Combining both inclusions, $C = gS$. Finally, left translation by g , i.e. the map $s \mapsto g \circ s$, is a bijection $S \rightarrow gS$ that is continuous with continuous inverse $h \mapsto g^{-1} \circ h$ (both built from the analytic group operations); it is therefore a homeomorphism, so gS is topologically just a translated copy of S , with the same dimension. $\square$

Definition 17.2. A Lie group is called **connected** if it has only one connected component (i.e. $G = S$). A connected, **non-abelian** Lie group is called **simple** if it has no proper connected normal subgroup other than $\{e\}$ and G itself (the non-abelian requirement excludes groups like $U(1)$ or $(\mathbb{R}, +)$, which vacuously satisfy the subgroup condition — their only proper subgroups are discrete, hence disconnected — without being what one means by “simple”).

18 Compactness

Definition 18.1 (Elementary definition used in this part.). A matrix Lie group G is called **compact** if it is a closed and bounded subset of the finite-dimensional space of matrices in which it is represented. Equivalently, in the parametrised language used here, the group can

be covered by finitely many coordinate regions with bounded parameter ranges, including all their limiting group elements.

Remark. Intrinsic definition, for the later topology and differential-geometry part. A Lie group is compact if its underlying topological space, equivalently its underlying manifold, is compact. In general topology this means that every open cover has a finite subcover. For the finite-dimensional matrix groups considered here, the Heine–Borel theorem reduces this intrinsic definition to the elementary closed-and-bounded criterion above. We use the elementary formulation in the present group-theory part and postpone the topological formulation and its consequences until the relevant geometric language has been developed.

Example 18.1. $SO(2)$ is compact: φ ranges over the bounded, “periodic” interval $(-\pi, \pi]$. $SU(2)$ is compact: it is (topologically) the 3-sphere, a closed and bounded subset of $\mathbb{R}^4$. In contrast, $(\mathbb{R}^+, \cdot)$ and the Lorentz group (parametrized by an unbounded rapidity $\chi \in \mathbb{R}$) are non-compact.

19 Representations of Groups

We now turn to the theory of *representations*, which is how group theory becomes a genuinely powerful computational tool in physics: instead of working with an abstract group, we work with concrete matrices acting on a vector space, e.g. the Hilbert space of a quantum-mechanical system.

Definition 19.1. A **representation** of a group G is a homomorphism $\Phi : G \rightarrow \{\Gamma\}$ from G to a group of $d \times d$ (real or complex) matrices, under which group multiplication maps to matrix multiplication:

$$\Phi(g_1) \Phi(g_2) = \Phi(g_1 \circ g_2) \quad \text{for all } g_1, g_2 \in G.$$

The integer d is called the **dimension** of the representation, and the vector space $V \cong \mathbb{C}^d$ (or $\mathbb{R}^d$) on which the matrices act is called the **carrier space** (or *module*) of the representation.

Example 19.1. Every group G has the **trivial representation** $\Phi(g) = 1$ (the 1×1 identity matrix) for all $g \in G$. A less trivial example: for $G = (\mathbb{R}, +)$, the map $\Phi(x) = e^x$ is a (1-dimensional) representation, since $\Phi(x + y) = e^{x+y} = e^x e^y = \Phi(x)\Phi(y)$; here $\Phi : (\mathbb{R}, +) \rightarrow (\mathbb{R}^+, \cdot)$.

Definition 19.2. Two representations Γ and Γ' of G (of the same dimension d) are called **equivalent** if there exists an invertible matrix S such that

$$\Gamma'(g) = S \Gamma(g) S^{-1} \quad \text{for all } g \in G.$$

Equivalent representations describe “the same” representation, merely expressed in a different basis of the carrier space.

19.1 Reducibility

Definition 19.3. A representation Γ acting on V is called **reducible** if there is a proper, non-trivial subspace $W \subset V$ that is *invariant*, i.e. $\Gamma(g)W \subseteq W$ for all $g \in G$. Choosing a basis adapted to W , a reducible representation takes the block-triangular form

$$\Gamma(g) = \begin{pmatrix} \Gamma_{11}(g) & \Gamma_{12}(g) \\ 0 & \Gamma_{22}(g) \end{pmatrix}.$$

If, furthermore, one can choose the basis so that $\Gamma_{12} = 0$ as well (i.e. both W and a complementary subspace are invariant), the representation is called **completely reducible** (or **decomposable**):

$$\Gamma(g) = \begin{pmatrix} \Gamma_{11}(g) & 0 \\ 0 & \Gamma_{22}(g) \end{pmatrix}.$$

A representation with no non-trivial invariant subspace at all is called **irreducible**.

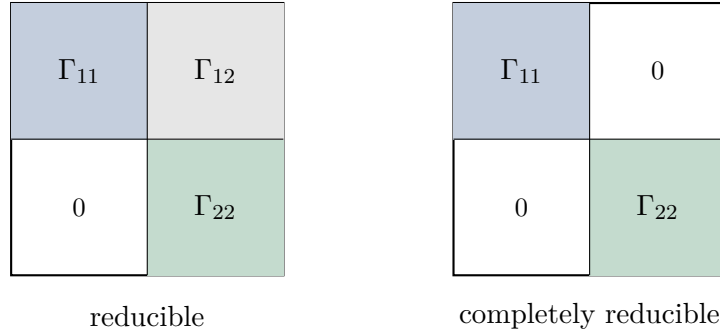


Figure 12: A reducible representation can always be brought to block upper-triangular form (left): the two *diagonal* blocks Γ_{11} , Γ_{22} are themselves representations (shaded), the block below the diagonal is forced to be zero by invariance of W' , but the block Γ_{12} above the diagonal is generic and need not vanish (shown in neutral grey, *not* the same colour as an actual representation block). If the representation is *completely* reducible (right), the remaining off-diagonal block Γ_{12} vanishes as well — both off-diagonal blocks are now white — and $\Gamma = \Gamma_{11} \oplus \Gamma_{22}$ decomposes into an honest direct sum of the two (unchanged, hence identically coloured) smaller representations.

20 Unitary Representations

For physical applications — where representation matrices typically act on a Hilbert space and must preserve probabilities — *unitary* representations ($\Gamma(g)^\dagger \Gamma(g) = \mathbb{1}$ for all g) are particularly important. Remarkably, for finite groups and compact Lie groups, unitarity comes essentially for free.

Theorem 20.1. (Unitarity of representations of finite groups.) Every representation of a finite group is equivalent to a unitary representation.

Proof. Let $\Gamma(g)$, $g \in G$, $|G| = N$, be a (not necessarily unitary) matrix representation. Define

$$S = \sum_{g \in G} \Gamma(g)^\dagger \Gamma(g).$$

Step 1: S is Hermitian and positive definite. Since $(\Gamma(g)^\dagger \Gamma(g))^\dagger = \Gamma(g)^\dagger \Gamma(g)$, each summand is

Hermitian, hence so is S . For any vector $\psi \neq 0$, $\psi^\dagger S \psi = \sum_g \|\Gamma(g)\psi\|^2 > 0$ (strictly, because the term $g = e$ alone already contributes $\|\psi\|^2 > 0$, as $\Gamma(e) = \mathbb{I}$).

Step 2: the key invariance identity $\Gamma(g_0)^\dagger S \Gamma(g_0) = S$ for every fixed $g_0 \in G$. This is where the rearrangement theorem is used — exactly once, and applied directly to S itself, rather than mixed into the later change-of-basis algebra. Insert $\Gamma(g_0)^\dagger(\cdot)\Gamma(g_0)$ around the whole sum defining S and use that Γ is a homomorphism, $\Gamma(h)\Gamma(g_0) = \Gamma(hg_0)$:

$$\begin{aligned} \Gamma(g_0)^\dagger S \Gamma(g_0) &= \sum_{h \in G} \Gamma(g_0)^\dagger \Gamma(h)^\dagger \Gamma(h) \Gamma(g_0) \\ &= \sum_{h \in G} (\Gamma(h)\Gamma(g_0))^\dagger (\Gamma(h)\Gamma(g_0)) = \sum_{h \in G} \Gamma(hg_0)^\dagger \Gamma(hg_0). \end{aligned}$$

By the rearrangement theorem, right multiplication by the fixed element g_0 permutes G (as h ranges over G , so does $h' := hg_0$), so relabelling the sum,

$$\Gamma(g_0)^\dagger S \Gamma(g_0) = \sum_{h' \in G} \Gamma(h')^\dagger \Gamma(h') = S.$$

Since $g_0 \in G$ was arbitrary, we have shown

$$\Gamma(g)^\dagger S \Gamma(g) = S \quad \text{for every } g \in G. \quad (1)$$

Step 3: constructing the unitary equivalent. Because S is Hermitian and positive definite (Step 1), it can be diagonalised by a unitary matrix U : $S = U^\dagger d U$ with $d = \text{diag}(d_1, \dots, d_n)$, $d_i > 0$. Define

$$X := U^\dagger \sqrt{d} U, \quad \text{so that } X = X^\dagger > 0, \quad X^2 = U^\dagger d U = S, \quad X^{-1} \text{ exists.}$$

Now define the new representation $\Gamma'(g) := X \Gamma(g) X^{-1}$, equivalent to Γ by construction. Since $X = X^\dagger$, we have $(X^{-1})^\dagger = (X^\dagger)^{-1} = X^{-1}$, so

$$\Gamma'(g)^\dagger = (X \Gamma(g) X^{-1})^\dagger = (X^{-1})^\dagger \Gamma(g)^\dagger X^\dagger = X^{-1} \Gamma(g)^\dagger X.$$

Therefore, using $X^2 = S$ and then the invariance identity Eq. (1) from Step 2:

$$\begin{aligned} \Gamma'(g)^\dagger \Gamma'(g) &= (X^{-1} \Gamma(g)^\dagger X) (X \Gamma(g) X^{-1}) = X^{-1} \Gamma(g)^\dagger X^2 \Gamma(g) X^{-1} \\ &= X^{-1} (\Gamma(g)^\dagger S \Gamma(g)) X^{-1} = X^{-1} S X^{-1} = X^{-1} X^2 X^{-1} = \mathbb{1}. \end{aligned}$$

So Γ' is unitary. □

Remark. For a *compact* Lie group the same strategy works, replacing the finite sum $\sum_{g \in G}$ by an integral $\int_G dg$ over the group with respect to an invariant (“Haar”) measure, $S = \int_G dg \Gamma(g)^\dagger \Gamma(g)$ — the compactness guarantees this integral is finite (finite “volume”), which is exactly what is needed for S to be well defined and positive definite as in Step 1 above. The invariance identity of Step 2, $\Gamma(g_0)^\dagger S \Gamma(g_0) = S$, now follows from the substitution $h' = hg_0$ inside the integral, i.e. from the property $\int_G dg F(gg_0) = \int_G dg F(g)$ for any (integrable) function F and any fixed $g_0 \in G$ — the continuous analogue of the rearrangement theorem used in Step 2 — which is precisely what the (right-invariant, and for compact groups also left-invariant, i.e. bi-invariant) Haar measure provides. Steps 1 and 3 then go through exactly as before.

Theorem 20.2. If G is finite or a compact Lie group, then every representation of G is equivalent to a unitary representation.

Lemma 20.3 (Equivalences between unitary representations can be chosen unitary). If Γ and Γ' are both *unitary* representations of G and are equivalent, i.e. $\Gamma'(g) = S \Gamma(g) S^{-1}$ for some invertible S , then S can be chosen to be unitary as well.

Proof. From $\Gamma'(g) = S \Gamma(g) S^{-1}$ we get $\Gamma'(g)^{-1} = S \Gamma(g)^{-1} S^{-1}$; since both representations are unitary, $\Gamma(g)^{-1} = \Gamma(g)^\dagger$ and $\Gamma'(g)^{-1} = \Gamma'(g)^\dagger$, and also $\Gamma'(g)^\dagger = (S^{-1})^\dagger \Gamma(g)^\dagger S^\dagger$ by taking the Hermitian conjugate of the defining relation. Comparing the two expressions for $\Gamma'(g)^{-1} = \Gamma'(g)^\dagger$,

$$S \Gamma(g)^{-1} S^{-1} = (S^{-1})^\dagger \Gamma(g)^{-1} S^\dagger \implies (S^\dagger S) \Gamma(g)^{-1} = \Gamma(g)^{-1} (S^\dagger S) \quad \text{for all } g,$$

i.e. $B := S^\dagger S$ commutes with every $\Gamma(g)^{-1}$, equivalently with every $\Gamma(g)$. If Γ is irreducible, Schur's Lemma II (Theorem 21.2, proved below) immediately gives $B = \lambda \mathbb{1}$ with $\lambda > 0$ (since $B = S^\dagger S$ is positive definite), so $\tilde{S} := S/\sqrt{\lambda}$ still implements the equivalence $\Gamma' = \tilde{S} \Gamma \tilde{S}^{-1}$ and is now unitary, $\tilde{S}^\dagger \tilde{S} = \mathbb{1}$. (For a general, reducible Γ the same conclusion holds after decomposing into irreducible blocks and applying this argument block by block, using Corollary 20.4 below.) $\square$

Physical perspective

It is important not to over-generalise Theorem 20.2: it is compactness (or finiteness) that guarantees a finite-dimensional *unitary* representation exists at all. A *non-compact* Lie group can fail badly to have any non-trivial finite-dimensional unitary representation — indeed, this is exactly the situation for the Lorentz group discussed above (Section 16): its non-compactness (the rapidity χ ranges over the unbounded line $\mathbb{R}$, rather than a bounded interval) is precisely why the “volume” argument of the remark above breaks down ($\int_{-\infty}^{\infty} d\chi$ diverges), and correspondingly why all of its non-trivial finite-dimensional representations (built from spinor and tensor indices) are necessarily *non-unitary*.

Corollary 20.4. Every reducible representation of a finite or compact Lie group is **completely reducible**.

Proof. By Theorems 20.1 and 20.2 we may replace Γ by an equivalent unitary representation (this does not affect reducibility, since equivalence is just a change of basis). So assume $\Gamma(g)^\dagger \Gamma(g) = \mathbb{1}$ for all g . Let $W \subset V$ be a proper, non-trivial invariant subspace, $\Gamma(g)W \subseteq W$ for all $g \in G$, and let P be the orthogonal projector of V onto W (so $P^2 = P$, $P^\dagger = P$, and $\Gamma(g)P$ is simply $\Gamma(g)$ restricted to act on W and then re-embedded in V).

Step 1: invariance of W is equivalent to $\Gamma(g)P = P\Gamma(g)P$ for all g . Indeed, $\Gamma(g)W \subseteq W$ says precisely that applying P after $\Gamma(g)$ has no effect when we start from a vector already in W , i.e. $P\Gamma(g)Pv = \Gamma(g)Pv$ for every $v \in V$ (not just $v \in W$), which is the stated matrix identity $P\Gamma(g)P = \Gamma(g)P$.

Step 2: the complementary subspace $W^\perp = (\mathbb{1} - P)V$ is invariant too. Take the Hermitian conjugate of the identity from Step 1:

$$(\Gamma(g)P)^\dagger = (P\Gamma(g)P)^\dagger \implies P^\dagger \Gamma(g)^\dagger = P^\dagger \Gamma(g)^\dagger P^\dagger \implies P \Gamma(g)^\dagger = P \Gamma(g)^\dagger P,$$

using $P^\dagger = P$. Since g was an arbitrary element of G , we may replace g by g^{-1} in this identity; using unitarity, $\Gamma(g^{-1}) = \Gamma(g)^{-1} = \Gamma(g)^\dagger$, so the identity becomes exactly $P\Gamma(g) = P\Gamma(g)P$.

— i.e. the *same* identity as in Step 1, but with P and $\Gamma(g)$ swapped on the left. Combining $\Gamma(g)P = P\Gamma(g)P$ (Step 1) with $P\Gamma(g) = P\Gamma(g)P$ (just derived), the right-hand sides agree, so in fact

$$\Gamma(g)P = P\Gamma(g) \quad \text{for all } g \in G,$$

i.e. P commutes with the whole representation. It follows immediately that $\mathbb{1} - P$ commutes with $\Gamma(g)$ too, and then

$$\Gamma(g)(\mathbb{1} - P) = (\mathbb{1} - P)\Gamma(g) = (\mathbb{1} - P)\Gamma(g)(\mathbb{1} - P),$$

which is precisely the invariance condition $\Gamma(g)W^\perp \subseteq W^\perp$ for the complementary subspace $W^\perp = (\mathbb{1} - P)V$.

Step 3: conclusion. Choosing an orthonormal basis adapted to the orthogonal decomposition $V = W \oplus W^\perp$ (first an orthonormal basis of W , then one of $W^\perp$), the invariance of *both* W and $W^\perp$ means every $\Gamma(g)$ is simultaneously block-diagonal in this basis: $\Gamma(g) = \Gamma_{11}(g) \oplus \Gamma_{22}(g)$ with no off-diagonal coupling. Hence Γ is completely reducible. If either block is itself reducible, the same argument applies to it in turn, so by induction on the dimension Γ decomposes entirely into a direct sum of *irreducible* unitary representations. $\square$

Example 20.1. (A representation that is reducible but *not* completely reducible.)

Consider $G = (\mathbb{R}, +)$ (which is neither finite nor compact) and the 2-dimensional representation

$$\Gamma(x) = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}, \quad \Gamma(x)\Gamma(y) = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & y \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & x+y \\ 0 & 1 \end{pmatrix} = \Gamma(x+y).$$

The subspace spanned by $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ is invariant, but there is no complementary invariant subspace: $\Gamma(x)$ cannot be brought to block-diagonal form for all x simultaneously. This illustrates why the hypotheses of Corollary 20.4 (finiteness or compactness) are essential.

21 Schur's Lemmas

Schur's Lemmas are two deceptively simple statements with enormous consequences for representation theory.

Theorem 21.1. (Schur's Lemma I.) Let Γ and Γ' be irreducible representations of G , of dimensions d and d' respectively, and suppose A is a $d' \times d$ matrix such that

$$\Gamma'(g)A = A\Gamma(g) \quad \text{for all } g \in G.$$

Then either $A = 0$, or $d = d'$ and A is invertible (so that Γ, Γ' are in fact equivalent).

Proof. The matrix A defines a linear map $V \rightarrow V'$ between the carrier spaces of Γ (dimension d) and Γ' (dimension d'); the hypothesis $\Gamma'(g)A = A\Gamma(g)$ says that A *intertwines* the two representations. We show separately that $\ker A$ is an invariant subspace of V and that $\text{im } A$ is an invariant subspace of V' .

$\ker A$ is invariant. Suppose $\psi \in \ker A$, i.e. $A\psi = 0$. For any $g \in G$,

$$A(\Gamma(g)\psi) = (A\Gamma(g))\psi = (\Gamma'(g)A)\psi = \Gamma'(g)(A\psi) = \Gamma'(g) \cdot 0 = 0,$$

so $\Gamma(g)\psi \in \ker A$ as well. Hence $\Gamma(g)\ker A \subseteq \ker A$ for every g : $\ker A$ is an invariant subspace of V . Since Γ is irreducible, by definition it has *no* non-trivial invariant subspace, so $\ker A$ must be

either $\{0\}$ or the whole of V . If $\ker A = V$, then $A = 0$ and we are done. So assume from now on $\ker A = \{0\}$, i.e. A is injective.

$\text{im } A$ is invariant. Let $\phi \in \text{im } A$, say $\phi = A\psi$ for some $\psi \in V$. For any $g \in G$,

$$\Gamma'(g)\phi = \Gamma'(g)A\psi = A\Gamma(g)\psi \in \text{im } A,$$

since $\Gamma(g)\psi \in V$ is again some vector that A maps into $\text{im } A$ by definition. Hence $\Gamma'(g)\text{im } A \subseteq \text{im } A$ for every g : $\text{im } A$ is an invariant subspace of V' . Since Γ' is irreducible, $\text{im } A$ is either $\{0\}$ or all of V' . As we are in the case $A \neq 0$ (having excluded $\ker A = V$ above), $\text{im } A \neq \{0\}$, so $\text{im } A = V'$: A is surjective.

Conclusion. A is injective and surjective, hence a linear isomorphism $V \rightarrow V'$; in particular $\dim V = \dim V'$, i.e. $d = d'$, and A (now a square matrix) is invertible. Moreover, from $\Gamma'(g)A = A\Gamma(g)$ we get $\Gamma'(g) = A\Gamma(g)A^{-1}$ for all g , so Γ and Γ' are equivalent representations. $\square$

Theorem 21.2. (Schur's Lemma II.) Let Γ be a d -dimensional *irreducible* representation of G over $\mathbb{C}$, and suppose B is a $d \times d$ matrix commuting with the representation,

$$\Gamma(g)B = B\Gamma(g) \quad \text{for all } g \in G.$$

Then $B = \lambda \mathbb{1}$ for some scalar $\lambda \in \mathbb{C}$.

Proof. This is where working over the *complex* numbers is essential: the characteristic polynomial $\det(B - \lambda \mathbb{1}) = 0$ is a degree- d polynomial equation in λ , which by the fundamental theorem of algebra always has at least one root $\lambda \in \mathbb{C}$ (this can fail over $\mathbb{R}$: e.g. a rotation matrix $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ has no real eigenvalue). Fix such an eigenvalue λ , and set $A := B - \lambda \mathbb{1}$. Then A still commutes with the representation,

$$\Gamma(g)A = \Gamma(g)B - \lambda\Gamma(g) = B\Gamma(g) - \lambda\Gamma(g) = A\Gamma(g) \quad \text{for all } g,$$

i.e. A satisfies the hypothesis of Schur's Lemma I (Theorem 21.1) with $\Gamma' = \Gamma$ (so automatically $d = d'$). That lemma tells us A is either 0 or invertible. But by construction λ is an eigenvalue of B , meaning precisely that $\det(B - \lambda \mathbb{1}) = \det A = 0$: A is *not* invertible. The only remaining possibility is $A = 0$, i.e. $B = \lambda \mathbb{1}$. $\square$

Corollary 21.3. Every irreducible representation of an **abelian** group over $\mathbb{C}$ is one-dimensional.

Proof. Let Γ be an irreducible *complex* representation of an abelian group G , and fix any $g_0 \in G$. Since G is abelian, $\Gamma(g_0)$ commutes with $\Gamma(g)$ for every $g \in G$: $\Gamma(g)\Gamma(g_0) = \Gamma(g_0)\Gamma(g)$. By Schur's Lemma II (Theorem 21.2, which requires the ground field to be algebraically closed — this is exactly where $\mathbb{C}$ is used), $\Gamma(g_0) = \lambda(g_0) \mathbb{1}$ for some scalar $\lambda(g_0)$. As g_0 was arbitrary, *every* representation matrix is a multiple of the identity; but then *every* subspace is invariant, so irreducibility forces $d = 1$. $\square$

Remark. The qualifier is essential: over $\mathbb{R}$, Schur's Lemma II fails (an eigenvalue of a real matrix need not be real), and indeed $SO(2)$ — an abelian Lie group — has, acting on $\mathbb{R}^2$ in the defining (identity) representation, the genuinely 2-dimensional real-irreducible rotation matrices $D(\varphi) = \begin{pmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{pmatrix}$ met above: this representation has no real invariant

line (a would-be eigenvector would need a real eigenvalue of modulus 1 other than ± 1 , impossible for $\varphi \neq 0, \pi$), yet becomes reducible into two 1-dimensional complex pieces $e^{\pm i\varphi}$ once complexified.

Example 21.1. For $G = \mathbb{Z}_2 = \{e, g\}$, an example of a 2-dimensional (reducible!) representation is $\Gamma(e) = \mathbb{1}_2$, $\Gamma(g) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$; this decomposes as the direct sum of the two inequivalent 1-dimensional irreducible representations $\{1, 1\}$ and $\{1, -1\}$, consistent with the corollary above.

21.1 Real, Complex, and Pseudo-real Representations

If $\Gamma(g)$ is a representation, then so is its complex conjugate $\Gamma^*(g)$, since $\Gamma^*(g_1)\Gamma^*(g_2) = (\Gamma(g_1)\Gamma(g_2))^* = \Gamma^*(g_1g_2)$. This leads to a three-way classification of irreducible representations:

Definition 21.1. An irreducible representation Γ is called:

- **real**, if there exists S such that $\Gamma^*(g) = S \Gamma(g) S^{-1}$ for all g , with S symmetric ($S^T = S$) — equivalently, Γ is equivalent to a representation by real matrices;
- **pseudo-real**, if such an S exists but is instead *antisymmetric*, $S^T = -S$;
- **complex**, if no such S exists at all, i.e. Γ and Γ^* are inequivalent representations.

Physical perspective

This classification matters physically: e.g. the spin- $\frac{1}{2}$ representation of $SU(2)$ is pseudo-real (related to the existence of the antisymmetric invariant tensor $\epsilon_{\alpha\beta}$ used to raise/lower $SU(2)$ indices), while representations of $SU(N)$ for $N \geq 3$ are generically complex — which is why, for instance, quarks (in the fundamental of $SU(3)_{\text{color}}$) and antiquarks (in the conjugate representation) are physically distinct.

22 Orthogonality Relations and Characters

Schur's Lemmas have a further, extremely useful consequence: an *orthogonality relation* among the matrix elements of inequivalent irreducible representations, which underlies practically all applications of representation theory to physics (selection rules, decomposition of reducible representations, etc.).

Let Γ_a (of dimension n_a) and Γ_b (of dimension n_b) be two *unitary, irreducible* representations of a finite group G of order $N = |G|$. Fix indices $1 \leq j \leq n_a$, $1 \leq m \leq n_b$, and let $E^{(jm)}$ denote the $n_a \times n_b$ matrix with a single entry 1 in position (j, m) and 0 elsewhere. Define the $n_a \times n_b$ matrix

$$M_{jm} := \sum_{g \in G} \Gamma_a(g) E^{(jm)} \Gamma_b(g)^{-1}.$$

Step 1: M_{jm} intertwines Γ_a and Γ_b . For any $h \in G$, relabelling the summation index $g \rightarrow g' = hg$ (a bijection of G with itself, so the sum over $g' \in G$ is the same as the sum over $g \in G$) and using $\Gamma_b(g)^{-1} = \Gamma_b(h^{-1}g')^{-1} = \Gamma_b(g')^{-1}\Gamma_b(h)$:

$$\Gamma_a(h) M_{jm} = \sum_{g \in G} \Gamma_a(hg) E^{(jm)} \Gamma_b(g)^{-1} = \sum_{g' \in G} \Gamma_a(g') E^{(jm)} \Gamma_b(g')^{-1} \Gamma_b(h) = M_{jm} \Gamma_b(h).$$

This is exactly the hypothesis of Schur's Lemma I (Theorem 21.1), with the roles $(\Gamma', \Gamma, A) = (\Gamma_a, \Gamma_b, M_{jm})$.

Step 2: evaluating M_{jm} via Schur's Lemma. Writing out matrix elements, since $E_{pq}^{(jm)} = \delta_{pj}\delta_{qm}$,

$$(M_{jm})_{ik} = \sum_{g \in G} \sum_{p,q} \Gamma_a(g)_{ip} E_{pq}^{(jm)} (\Gamma_b(g)^{-1})_{qk} = \sum_{g \in G} \Gamma_a(g)_{ij} (\Gamma_b(g)^{-1})_{mk}.$$

If Γ_a and Γ_b are *inequivalent* (in particular if $a \neq b$), Schur's Lemma I forces $M_{jm} = 0$ for every choice of j, m , i.e.

$$\sum_{g \in G} \Gamma_a(g)_{ij} (\Gamma_b(g)^{-1})_{mk} = 0 \quad \text{for all } i, j, k, m.$$

If instead $a = b$ (so $n_a = n_b =: n$ and $\Gamma_a = \Gamma_b$), Schur's Lemma II applies instead (since now M_{jm} commutes with the single irreducible representation Γ_a in the sense $\Gamma_a(h)M_{jm} = M_{jm}\Gamma_a(h)$), giving $M_{jm} = \lambda_{jm} \mathbb{1}_n$ for some scalar λ_{jm} depending on j, m . To pin down λ_{jm} , take the trace (set $i = k$ and sum):

$$\text{tr } M_{jm} = \sum_{g \in G} \sum_i \Gamma_a(g)_{ij} (\Gamma_a(g)^{-1})_{mi} = \sum_{g \in G} [\Gamma_a(g)^{-1} \Gamma_a(g)]_{mj} = \sum_{g \in G} \delta_{mj} = N \delta_{mj},$$

while also $\text{tr } M_{jm} = \text{tr}(\lambda_{jm} \mathbb{1}_n) = n \lambda_{jm}$. Comparing, $\lambda_{jm} = \frac{N}{n} \delta_{mj}$, so

$$\sum_{g \in G} \Gamma_a(g)_{ij} (\Gamma_a(g)^{-1})_{mk} = \frac{N}{n_a} \delta_{mj} \delta_{ik}.$$

Step 3: using unitarity to reach the final form. Since $\Gamma_a(g)$ is unitary, $(\Gamma_a(g)^{-1})_{mk} = (\Gamma_a(g)^\dagger)_{mk} = \Gamma_a(g)^*_{km}$; relabelling $k \rightarrow l$, the $a = b$ result becomes $\sum_g \Gamma_a(g)_{ij} \Gamma_a^*(g)_{lm} = \frac{N}{n_a} \delta_{jm} \delta_{il}$, while the $a \neq b$ result becomes $\sum_g \Gamma_a(g)_{ij} \Gamma_b^*(g)_{lm} = 0$. Combining both cases into a single statement gives the celebrated **(great) orthogonality relation**:

$$\sum_{g \in G} \frac{n_a}{N} \Gamma_a(g)_{ij} \Gamma_b^*(g)_{lm} = \delta_{ab} \delta_{il} \delta_{jm}. \quad (2)$$

In words: viewed as vectors indexed by $g \in G$, the (rescaled) matrix-element functions $g \mapsto \sqrt{n_a/N} \Gamma_a(g)_{ij}$ are mutually orthonormal, for all irreducible representations a and all choices of matrix indices i, j — this is the precise sense in which representation matrix elements generalise the orthonormal Fourier basis $\{e^{in\varphi}\}$, made explicit in Eq. (3) below.

22.1 Characters

Definition 22.1. The **character** of a representation Γ is the function $\chi_\Gamma : G \rightarrow \mathbb{C}$ given by the trace,

$$\chi_\Gamma(g) := \text{tr}(\Gamma(g)).$$

Since the trace is invariant under conjugation, equivalent representations have identical characters, and characters are automatically *class functions* (constant on conjugacy classes) — precisely because conjugate group elements are related by $g' = xgx^{-1}$, and $\text{tr}(\Gamma(x)\Gamma(g)\Gamma(x)^{-1}) = \text{tr } \Gamma(g)$.

To pass from the matrix-element orthogonality relation Eq. (2) to a relation among characters, set $j = i$ and $m = l$ in Eq. (2) and sum over both i (from 1 to n_a) and l (from 1 to n_b):

$$\sum_{i=1}^{n_a} \sum_{l=1}^{n_b} \sum_{g \in G} \frac{n_a}{N} \Gamma_a(g)_{ii} \Gamma_b^*(g)_{ll} = \sum_{i=1}^{n_a} \sum_{l=1}^{n_b} \delta_{ab} \delta_{il} \delta_{il}.$$

On the left, $\sum_i \Gamma_a(g)_{ii} = \text{tr } \Gamma_a(g) = \chi_a(g)$ and likewise $\sum_l \Gamma_b^*(g)_{ll} = \chi_b^*(g)$; on the right, $\sum_{i,l} \delta_{ab} \delta_{il} \delta_{il}$ is only non-zero when $a = b$ (forcing $n_a = n_b =: n$), in which case $\sum_{i,l=1}^n \delta_{il} \delta_{il} = \sum_{i=1}^n 1 = n = n_a$. So the right-hand side collapses to $\delta_{ab} n_a$, and dividing through by n_a yields the **character orthogonality relation**:

$$\frac{1}{N} \sum_{g \in G} \chi_a(g) \chi_b^*(g) = \delta_{ab}. \quad (3)$$

Physical perspective

The orthogonality relation Eq. (3) is a direct generalisation of the orthogonality of Fourier modes: for $G = U(1)$ (or, at finite level, $\mathbb{Z}_N$), the irreducible representations are $\Gamma_n(\varphi) = e^{in\varphi}$, and Eq. (3) becomes precisely

$$\frac{1}{2\pi} \int_0^{2\pi} d\varphi e^{-im\varphi} e^{in\varphi} = \delta_{n,m},$$

the orthogonality of Fourier modes on the circle. In this sense, Fourier analysis is simply the representation theory of the abelian group $U(1)$ — and the character orthogonality relation Eq. (3) is the direct generalisation of Fourier's theorem to arbitrary (finite or compact) groups.

Characters are especially useful in practice because they let us decompose an arbitrary (reducible) representation Γ into irreducibles *without* having to explicitly find the change-of-basis matrix: if $\Gamma = \bigoplus_a m_a \Gamma_a$ (with Γ_a irreducible, appearing with multiplicity m_a), then $\chi_\Gamma = \sum_a m_a \chi_a$, and Eq. (3) gives immediately

$$m_a = \frac{1}{N} \sum_{g \in G} \chi_\Gamma(g) \chi_a^*(g),$$

a simple, purely numerical recipe for extracting the multiplicities.

Part III

Lie Algebras: From Groups to Algebras and Back

23 Introduction: Why Linearise a Lie Group?

In Part B we defined Lie groups as sets of matrices depending analytically on continuous parameters, and we introduced, in passing, the **generators** $\Theta_\rho = \partial\Gamma/\partial\xi_\rho|_{\xi=0}$ obtained by differentiating the group elements at the identity. We saw that these generators do not merely span a vector space: the group multiplication law, expanded to second order in the parameters, forces the commutators $[\Theta_\rho, \Theta_\sigma]$ to close back onto the span of the Θ_ρ , defining the *structure constants* of the group.

This chapter takes that observation seriously and develops it into a full theory. The generators, together with the commutator bracket, form an algebraic structure in their own right — the **Lie algebra** $\mathfrak{g}$ of G — and remarkably, *almost all* of the information contained in the (connected component of the) Lie group G is already encoded in the finite-dimensional, purely algebraic object $\mathfrak{g}$. Since linear algebra is vastly simpler than the analysis of a curved group manifold,

this is an enormous simplification: many questions about Lie groups (What are the subgroups? What are the representations? Which groups are related by covering maps?) can be answered by first solving the corresponding, purely algebraic question for $\mathfrak{g}$, and only afterwards worrying about global (topological) subtleties.

Physical perspective

This is the single most important idea in the application of group theory to physics. A physicist essentially never studies a continuous symmetry group “globally”; instead, one writes down the generators (angular momentum operators, isospin operators, the generators of the Poincaré group, . . .), computes their commutation relations, and works almost exclusively at the Lie algebra level. The present chapter makes precise exactly how much information survives this simplification, and — crucially — exactly what is lost (global topology: is the group simply connected? compact? what are its discrete central subgroups?), so that we know when we must go back and think about the group itself.

24 The Lie Algebra: Abstract Definition

We first isolate, in purely algebraic terms, the structure that a Lie group’s generators satisfy, without reference to any specific group or matrix representation.

Definition 24.1. A **Lie algebra** $\mathfrak{X}$ (over the real numbers) is a (finite-dimensional) real vector space equipped with a bilinear product

$$* : \mathfrak{X} \times \mathfrak{X} \longrightarrow \mathfrak{X}, \quad (a, b) \longmapsto a * b,$$

called the **Lie product** (or *Lie bracket*), satisfying the following axioms for all $a, b, c \in \mathfrak{X}$ and $\alpha, \beta \in \mathbb{R}$:

- (i) **Bilinearity:** $(\alpha a + \beta b) * c = \alpha(a * c) + \beta(b * c)$, and likewise in the second slot.
- (ii) **Antisymmetry:** $a * b = -b * a$ (in particular $a * a = 0$).
- (iii) **Jacobi identity:**

$$a * (b * c) + b * (c * a) + c * (a * b) = 0.$$

Remark. Axioms (i)–(iii) are exactly what allows for a *complete classification* of (finite-dimensional, semisimple) Lie algebras and their representations — one of the great achievements of nineteenth- and twentieth-century mathematics, and the basis for everything from the classification of elementary particle multiplets to the periodic table’s structure via angular momentum. Antisymmetry alone would make $\mathfrak{X}$ merely an anticommutative algebra; it is the Jacobi identity that encodes, in linear-algebraic language, the associativity of the group multiplication from which the bracket ultimately descends (see Theorem 25.1 below), and it is what makes the theory rigid enough to classify.

The definition also makes sense, verbatim, for a *complex* vector space $\mathfrak{X}$ (with $\alpha, \beta \in \mathbb{C}$); one then speaks of a **complex Lie algebra**. Given a real Lie algebra $\mathfrak{X}$, one may always form its **complexification** $\mathfrak{X}_{\mathbb{C}} := \mathfrak{X} \oplus i\mathfrak{X}$ (formal complex linear combinations of elements of $\mathfrak{X}$), extending the bracket complex-bilinearly; this turns out to be an extremely useful trick, allowing us to form complex linear combinations of generators (such as the ladder operators $t_{\pm} = t_1 \pm it_2$ of Section 31 below) while still classifying representations of the original real Lie algebra $\mathfrak{X}$.

Example 24.1. Let $\mathfrak{X} = \{ N \times N \text{ complex matrices} \}$: since it allows complex linear combinations, this is a complex vector space of (complex) dimension N^2 , and hence also a real vector space, of (real) dimension $2N^2$, if one restricts to real linear combinations of a basis. Define

$$a * b := ab - ba = [a, b],$$

the ordinary matrix commutator. Bilinearity and antisymmetry are immediate. The Jacobi identity is a direct (if slightly tedious) computation:

$$\begin{aligned} [a, [b, c]] + [b, [c, a]] + [c, [a, b]] &= a(bc - cb) - (bc - cb)a \\ &\quad + b(ca - ac) - (ca - ac)b \\ &\quad + c(ab - ba) - (ab - ba)c \\ &= (abc - acb - bca + cba) + (bca - bac - cab + acb) \\ &\quad + (cab - cba - abc + bac) \\ &= 0, \end{aligned}$$

since every one of the six triple products $abc, acb, bac, bca, cab, cba$ appears exactly once with a $+$ sign and once with a $-$ sign. Hence *every* associative algebra becomes a Lie algebra under the commutator bracket; this is by far the most important source of examples, and it is the one relevant to matrix Lie groups.

Definition 24.2. A Lie algebra $\mathfrak{X}$ is called **abelian** if $a * b = 0$ for all $a, b \in \mathfrak{X}$.

Choosing a basis $\{t_a\}_{a=1}^n$ of $\mathfrak{X}$, bilinearity of the bracket means that $[t_a, t_b]$ is determined once we know its expansion in the same basis,

$$[t_a, t_b] = f_{ab}^c t_c \quad (\text{sum over } c \text{ understood}),$$

and the numbers $f_{ab}^c \in \mathbb{R}$ are the **structure constants** of $\mathfrak{X}$ in this basis (we already met these in Part B, obtained there directly from the group multiplication law). Antisymmetry of the bracket forces $f_{ab}^c = -f_{ba}^c$, and the Jacobi identity translates into a quadratic constraint on the f_{ab}^c :

Theorem 24.1 (Jacobi identity in components). For any basis $\{t_a\}$ of a Lie algebra with structure constants f_{ab}^c ,

$$f_{ab}^d f_{dc}^e + f_{bc}^d f_{da}^e + f_{ca}^d f_{db}^e = 0.$$

Proof. Expand $[t_a, [t_b, t_c]] + [t_b, [t_c, t_a]] + [t_c, [t_a, t_b]] = 0$ using $[t_b, t_c] = f_{bc}^d t_d$ etc.:

$$[t_a, [t_b, t_c]] = [t_a, f_{bc}^d t_d] = f_{bc}^d [t_a, t_d] = f_{bc}^d f_{ad}^e t_e = -f_{bc}^d f_{da}^e t_e,$$

and cyclically for the other two terms. Summing the three (cyclic in a, b, c) and setting the coefficient of each basis vector t_e to zero gives the stated identity (up to relabelling of the cyclic order, which is immaterial since the sum is already cyclic). $\square$

25 From the Group to the Algebra: Tangent Vectors and the Exponential Map

We now justify, carefully and in full generality, the claim already used informally in Part B: the commutator of two generators of a Lie group is again a generator. This is where the Jacobi

identity — and indeed the whole Lie-algebra structure — is seen to descend directly from the *associativity* of the group multiplication.

25.1 Tangent Vectors to Curves

Definition 25.1. Let G be an n -dimensional Lie group with parametrization $\Gamma(\xi)$, $\xi \in \mathbb{R}^n$, and let $\Gamma(s)$, $s \in (-\varepsilon, \varepsilon)$, be a differentiable curve in G with $\Gamma(0) = e$ (i.e. $\xi(s)$ is a differentiable curve in parameter space with $\xi(0) = 0$). Writing the matrices $\Theta_\rho = \partial\Gamma/\partial\xi_\rho|_{\xi=0}$ for the generators as before, the **tangent vector** to the curve $\Gamma(s)$ at $s = 0$ is

$$\left. \frac{d\Gamma(s)}{ds} \right|_{s=0} = \sum_{\rho=1}^n \left. \frac{d\xi_\rho}{ds} \right|_{s=0} \left. \frac{\partial\Gamma}{\partial\xi_\rho} \right|_{\xi=0} = \sum_{\rho=1}^n a_\rho \Theta_\rho =: a \in \mathfrak{g}, \quad a_\rho := \left. \frac{d\xi_\rho}{ds} \right|_{s=0}.$$

Since the Θ_ρ span the linear vector space $\mathfrak{g}$, *every* tangent vector to a curve through the identity is (by definition) an element of $\mathfrak{g}$; and conversely every element $a = \sum_\rho a_\rho \Theta_\rho \in \mathfrak{g}$ is the tangent vector to *some* curve through e (for instance, the straight line $\xi(s) = sa$ in parameter space, to first order; we will shortly meet a much better choice of curve, the one-parameter subgroup $A(s) = \exp(sa)$). In this sense $\mathfrak{g}$ is exactly the tangent space to the manifold G at the point e , as the notation of Part B already suggested.

Theorem 25.1. (Closure of the bracket / origin of the Jacobi identity.) Let $a, b \in \mathfrak{g}$ be tangent vectors (at $s = 0$) to two curves $A(s), B(s)$ through the identity of G . Then the commutator $[a, b] = ab - ba$ (ordinary matrix commutator, since $A(s), B(s)$ are matrices) is again an element of $\mathfrak{g}$: it is the tangent vector, at $s = 0$, to the curve

$$C(s) := A(\sqrt{s}) B(\sqrt{s}) A(\sqrt{s})^{-1} B(\sqrt{s})^{-1} \quad (s \geq 0),$$

i.e. $C(s) = \mathbb{1} + s[a, b] + O(s^{3/2})$.

Proof. Write the curves to second order in their own parameter,

$$A(s) = \mathbb{1} + sa + \frac{1}{2}s^2a_2 + O(s^3), \quad B(s) = \mathbb{1} + sb + \frac{1}{2}s^2b_2 + O(s^3),$$

where $a_2 := \ddot{A}(0)$, $b_2 := \ddot{B}(0)$ are whatever second derivatives the curves happen to have (they need *not* lie in $\mathfrak{g}$; we shall see that they drop out of the final answer, as they must, since $[a, b] \in \mathfrak{g}$ cannot depend on the arbitrary choice of curve representing a and b). Inverting order by order (multiply out $A(s)A(s)^{-1} = \mathbb{1}$ and match powers of s),

$$A(s)^{-1} = \mathbb{1} - sa + s^2\left(a^2 - \frac{1}{2}a_2\right) + O(s^3), \quad B(s)^{-1} = \mathbb{1} - sb + s^2\left(b^2 - \frac{1}{2}b_2\right) + O(s^3).$$

We now build up the “group commutator” $D(s) := A(s)B(s)A(s)^{-1}B(s)^{-1}$ two factors at a time.

Step 1: $A(s)B(s)$.

$$A(s)B(s) = \mathbb{1} + s(a + b) + s^2\left(\frac{1}{2}a_2 + \frac{1}{2}b_2 + ab\right) + O(s^3).$$

Step 2: multiply by $A(s)^{-1}$. Keeping all terms up to $O(s^2)$,

$$\begin{aligned} A(s)B(s)A(s)^{-1} &= \left[\mathbb{1} + s(a + b) + s^2\left(\frac{1}{2}a_2 + \frac{1}{2}b_2 + ab\right)\right] \left[\mathbb{1} - sa + s^2\left(a^2 - \frac{1}{2}a_2\right)\right] + O(s^3) \\ &= \mathbb{1} + s[(a + b) - a] + s^2\left[\left(a^2 - \frac{1}{2}a_2\right) - (a + b)a + \left(\frac{1}{2}a_2 + \frac{1}{2}b_2 + ab\right)\right] + O(s^3) \\ &= \mathbb{1} + sb + s^2\left[\frac{1}{2}b_2 + ab - ba\right] + O(s^3) = \mathbb{1} + sb + s^2\left(\frac{1}{2}b_2 + [a, b]\right) + O(s^3), \end{aligned}$$

where in the last line the terms $a^2 - \frac{1}{2}a_2 - a^2 - ba + \frac{1}{2}a_2 + \frac{1}{2}b_2 + ab$ collapsed to $\frac{1}{2}b_2 + ab - ba$.

Step 3: multiply by $B(s)^{-1}$.

$$\begin{aligned} D(s) &= A(s)B(s)A(s)^{-1}B(s)^{-1} = \left[\mathbb{1} + sb + s^2\left(\frac{1}{2}b_2 + [a, b]\right) \right] \left[\mathbb{1} - sb + s^2\left(b^2 - \frac{1}{2}b_2\right) \right] + O(s^3) \\ &= \mathbb{1} + s(b - b) + s^2\left[\left(b^2 - \frac{1}{2}b_2\right) - b^2 + \left(\frac{1}{2}b_2 + [a, b]\right)\right] + O(s^3) \\ &= \mathbb{1} + s^2[a, b] + O(s^3). \end{aligned}$$

All dependence on the “unphysical” second derivatives a_2, b_2 has cancelled, exactly as it must. Reparametrizing $s = \sqrt{\cdot}$, i.e. setting $C(s) := D(\sqrt{s}) = \mathbb{1} + s[a, b] + O(s^{3/2})$, we read off

$$\left. \frac{dC(s)}{ds} \right|_{s=0} = [a, b],$$

so $[a, b]$ is indeed the tangent vector at the identity to the curve $C(s) \in G$, i.e. $[a, b] \in \mathfrak{g}$. $\square$

Remark. This is the promised derivation of the Lie-algebra structure “from the group”: the only input was that G is a group of matrices closed under multiplication and inversion (so that $D(s) = A(s)B(s)A(s)^{-1}B(s)^{-1}$ is again a curve in G) and that this multiplication is associative. That the resulting bracket automatically satisfies antisymmetry and the Jacobi identity (Definition 24.1) is therefore not a separate assumption — it is inherited for free from the associativity of matrix multiplication, exactly as in the worked example of Example 24.1. Conceptually: *the Lie bracket measures the extent to which two one-parameter subgroups fail to commute*, at leading (quadratic) order in the group parameters. If G is abelian, $D(s) = \mathbb{1}$ identically and $[a, b] = 0$ for all $a, b \in \mathfrak{g}$: the Lie algebra of an abelian Lie group is itself abelian, consistent with Section 24.

25.2 The Exponential Map and One-Parameter Subgroups

Among all curves through the identity with a given tangent vector $a \in \mathfrak{g}$, one choice is canonical and deserves special attention.

Definition 25.2. For $a \in \mathfrak{g}$ (an $n \times n$ matrix, in our matrix-group setting), the **matrix exponential** is defined by the everywhere-convergent power series

$$\exp(a) := \sum_{k=0}^{\infty} \frac{a^k}{k!} = \mathbb{1} + a + \frac{1}{2}a^2 + \frac{1}{6}a^3 + \cdots.$$

Theorem 25.2. Let $a \in \mathfrak{g}$. Then:

- (a) $A(t) := \exp(ta)$, $t \in \mathbb{R}$, defines a **one-parameter subgroup** of G : a smooth curve lying entirely in G , with

$$A(0) = \mathbb{1}, \quad A(t_1)A(t_2) = A(t_1 + t_2) \quad \forall t_1, t_2 \in \mathbb{R}, \quad \left. \frac{dA}{dt} \right|_{t=0} = a.$$

- (b) Every element of G sufficiently close to the identity can be written, uniquely, as $\exp(a)$ for some (small) $a \in \mathfrak{g}$: the exponential map is a local diffeomorphism between a neighbourhood of $0 \in \mathfrak{g}$ and a neighbourhood of $e \in G$, with local inverse the logarithm.
- (c) If G is, in addition, **compact** and **connected**, then *every* element of G (not just those near e) can be written as $\exp(a)$ for some $a \in \mathfrak{g}$ (though generally not uniquely: e.g. $\exp(2\pi\Theta) = \exp(0) = \mathbb{1}$ for a suitably normalised generator Θ of a compact

one-parameter subgroup).

Proof. (a) That $A(t) \in G$ for every t , and that $A(t_1 + t_2) = A(t_1)A(t_2)$, follows because $a \in \mathfrak{g}$ means, by Definition 25.1, that a is tangent to *some* curve in G through e ; one shows (this requires the fuller apparatus of Lie’s theorems, only sketched here) that the specific curve $\exp(ta)$, built purely from a and the ambient matrix algebra, is guaranteed to stay inside G precisely because $\mathfrak{g}$ is closed under the bracket (Theorem 25.1) — intuitively, G is “curved along the directions in $\mathfrak{g}$ ” to all orders, not just to first order, and the exponential resums the entire Taylor series of the group multiplication law along the direction a . The functional equation $A(t_1)A(t_2) = A(t_1 + t_2)$ is the ordinary law of exponents for commuting matrices (ta and $t'a$ always commute), $e^{t_1a}e^{t_2a} = e^{(t_1+t_2)a}$, proved termwise from the power series exactly as for ordinary numbers. Differentiating the series termwise at $t = 0$ gives $\dot{A}(0) = a$.

(b) This is the inverse function theorem applied to the map $a \mapsto \exp(a)$ near $a = 0$: its derivative at $a = 0$ is the identity map on $\mathfrak{g}$ (from $\exp(a) = \mathbb{1} + a + O(a^2)$), hence invertible, so $\exp$ is a diffeomorphism between a small enough ball around $0 \in \mathfrak{g}$ and a small enough neighbourhood of $\mathbb{1} \in G$.

(c) We only sketch the true statement here, since a fully honest proof needs tools beyond this course (either a bi-invariant Riemannian metric on G , for which $\exp$ becomes the Riemannian exponential map, together with the Hopf–Rinow theorem applied to the compact — hence complete — manifold G ; or the maximal torus theorem, which shows every element of a compact connected G lies in *some* maximal torus $T \subset G$, and $\exp|_{\mathfrak{t}} : \mathfrak{t} \rightarrow T$ is already known to be onto because T is a torus). What part (b) *does* immediately give is the weaker statement that the subgroup H generated by $\exp(\mathfrak{g})$ (i.e. all finite products of exponentials) is all of G : H contains an open neighbourhood of e by part (b), and a subgroup containing an open neighbourhood of e is automatically both open (it is a union of translates hU of that neighbourhood) and closed (its complement is a union of cosets, each open by the same argument, so the complement is open too); since G is connected, $H = G$. (Beware the tempting but *fallacious* shortcut of trying to upgrade this to “ $\exp(\mathfrak{g})$ itself is open” via the Baker–Campbell–Hausdorff series $\exp(a)\exp(b) = \exp(a + b + \frac{1}{2}[a, b] + \dots)$: this series converges, and $b \mapsto \text{BCH}(a, b)$ is a local diffeomorphism in b , only where $\text{ad}(a)$ has no eigenvalue in $2\pi i\mathbb{Z} \setminus \{0\}$ — which can fail, so this does *not* establish that $\exp(\mathfrak{g})$ is open at an arbitrary one of its points $\exp(a)$.) For a *non-compact* connected group even the full statement (c) can fail: e.g. in $SL(2, \mathbb{R})$ there are elements not of the form $\exp(a)$ for any single a in the Lie algebra, although (consistent with the weaker statement just proved) every element is still a *product* of finitely many exponentials. $\square$

Remark. Part (c) is extremely useful in practice: to check whether some claimed identity, homomorphism, or representation property holds for *all* of a compact connected Lie group G , it suffices to check it on elements of the form $\exp(a)$, $a \in \mathfrak{g}$ — and by continuity and the group property, often it is enough to check it to first order in a near the identity, recovering the whole group by exponentiating and using closure. This is precisely the sense in which “the Lie algebra determines the (connected, compact) Lie group”; Section 30 below makes precise exactly how much freedom remains (the choice of *which* connected group with a given Lie algebra one means).

25.3 A General Recipe: Linearising the Defining Condition of a Matrix Group

The exponential map provides an extremely efficient, direct way of writing down the Lie algebra of any matrix Lie group defined by algebraic conditions on its elements: simply substitute $g = \exp(a)$ into the defining condition and expand to first order in a (equivalently, differentiate

the condition at $t = 0$ along the curve $g(t) = \exp(ta)$. We illustrate this for the classical matrix groups.

Example 25.1. (a) $SL(N, \mathbb{C})$ and its Lie algebra $\mathfrak{sl}(N, \mathbb{C})$. The defining condition is $\det(g) = 1$. Using the identity $\det(\exp(a)) = \exp(\operatorname{tr} a)$ (immediate by putting a in triangular form, e.g. over $\mathbb{C}$, so that $\exp(a)$ is triangular with diagonal entries $e^{a_{ii}}$, whence $\det \exp(a) = \prod_i e^{a_{ii}} = e^{\sum_i a_{ii}} = e^{\operatorname{tr} a}$),

$$\det(\exp(a)) = e^{\operatorname{tr}(a)} = 1 \quad \Longleftrightarrow \quad \operatorname{tr}(a) = 0.$$

So $\mathfrak{sl}(N, \mathbb{C}) = \{a \in \mathbb{C}^{N \times N} : \operatorname{tr}(a) = 0\}$, the traceless complex $N \times N$ matrices, a complex vector space of dimension $N^2 - 1$.

(b) $U(N)$ and $SU(N)$. Unitarity, $u^\dagger u = \mathbb{1}$, becomes, with $u = \exp(a)$,

$$\exp(a)^\dagger \exp(a) = \exp(a^\dagger) \exp(a) = \mathbb{1} \quad \Longrightarrow \quad a^\dagger = -a,$$

i.e. a is **anti-Hermitian**. (One can see the implication directly to first order, $\mathbb{1} + a^\dagger + a + O(a^2) = \mathbb{1}$, or note that $a^\dagger = -a$ is exactly the condition under which $\exp(a)^\dagger = \exp(a^\dagger) = \exp(-a) = \exp(a)^{-1}$ holds to *all* orders, so it is both necessary and sufficient.) Together with tracelessness (from (a), since $SU(N) \subset SL(N, \mathbb{C})$), we obtain

$$\mathfrak{su}(N) = \{a \in \mathbb{C}^{N \times N} : a^\dagger = -a, \operatorname{tr}(a) = 0\},$$

a *real* vector space (anti-Hermitian matrices do not close under multiplication by a general complex scalar) of real dimension $N^2 - 1$. Physicists almost always write $a = i\theta$ with $\theta = \theta^\dagger$ Hermitian and traceless, so that group elements are written $\exp(i\vec{\xi} \cdot \vec{t})$ with t_a Hermitian generators, as we did for $SU(2)$ above (there, $N^2 - 1 = 3$, matching the three Pauli matrices). $U(N)$ itself has Lie algebra $\mathfrak{u}(N)$, the anti-Hermitian matrices *without* the trace condition, of real dimension N^2 ; note $\mathfrak{u}(N) = \mathfrak{su}(N) \oplus \mathfrak{u}(1)$, reflecting $U(N) \cong (SU(N) \times U(1))/\mathbb{Z}_N$.

(c) $O(N)$ and $SO(N)$. Orthogonality, $O^T O = \mathbb{1}$, becomes $\exp(a)^T \exp(a) = \mathbb{1}$. Differentiating at the identity gives $a^T + a = 0$; conversely, $a^T = -a$ implies $\exp(a)^T = \exp(-a) = \exp(a)^{-1}$, so the condition is exact, i.e.

$$a^T = -a,$$

a antisymmetric. An antisymmetric real matrix is automatically traceless, so $\det(\exp a) = e^{\operatorname{tr} a} = 1$ automatically: every $a \in \mathfrak{so}(N)$ already satisfies the $SO(N)$ condition, and $\mathfrak{so}(N) = \mathfrak{o}(N)$ (the tangent spaces to $O(N)$ and to $SO(N)$ at the identity coincide, even though $O(N)$ itself has two connected components — reflections are not reached by exponentiating $\mathfrak{so}(N)$, consistent with Theorem 25.2(c), since $O(N)$ is compact but *disconnected*). The real dimension of $\mathfrak{so}(N)$ (the number of independent entries of a real antisymmetric $N \times N$ matrix) is $\binom{N}{2} = \frac{1}{2}N(N-1)$.

(d) The pseudo-orthogonal groups $O(p, q)$. Let $\eta = \operatorname{diag}(\underbrace{+1, \dots, +1}_p, \underbrace{-1, \dots, -1}_q)$ be

the invariant metric ($p + q = N$; for the Lorentz group, $p = 1, q = 3$ or $p = 3, q = 1$ according to signature convention). The defining condition is $O^T \eta O = \eta$, which linearises (by the identical argument) to

$$a^T \eta + \eta a = 0 \quad \Longleftrightarrow \quad \eta a \text{ is antisymmetric} \quad \Longleftrightarrow \quad (\eta a)^T = -\eta a.$$

So $\mathfrak{so}(p, q) = \{a \in \mathbb{R}^{N \times N} : \eta a + a^T \eta = 0\}$; again real dimension $\frac{1}{2}N(N-1)$, matching $\mathfrak{so}(N)$ (the two Lie algebras are, in fact, related by the complexification-then-different-real-form procedure familiar from the Lorentz group $\mathfrak{so}(3, 1)$ versus the Euclidean $\mathfrak{so}(4)$). Explicitly,

for $p = 1, q = 1$ (boosts along one axis) one finds $a = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ satisfies $\eta a + a^T \eta = 0$ with $\eta = \text{diag}(1, -1)$, and $\exp(ta) = \begin{pmatrix} \cosh t & \sinh t \\ \sinh t & \cosh t \end{pmatrix}$, the familiar Lorentz boost of rapidity t — illustrating why the boost generators are *not* anti-symmetric (unlike rotation generators) and why boosts do not form a compact one-parameter subgroup: $\cosh t, \sinh t$ never return to $(1, 0)$ for $t \neq 0$.

25.4 Example Revisited: $\mathfrak{su}(2)$

Recall from Part B the generators of $SU(2)$ built from the Pauli matrices, $\Theta_a = \frac{i}{2}\sigma_a$ (so that Θ_a is anti-Hermitian, as required for $\exp(\xi_a \Theta_a) \in SU(2)$ to be unitary) or, in the physicist's Hermitian convention, $t_a := \frac{1}{2}\sigma_a$ with $\Theta_a = it_a$. A direct computation using $\sigma_a \sigma_b = \delta_{ab} \mathbb{1} + i\epsilon_{abc} \sigma_c$ gives

$$[t_a, t_b] = \frac{1}{4}[\sigma_a, \sigma_b] = \frac{1}{4}(\sigma_a \sigma_b - \sigma_b \sigma_a) = \frac{1}{4}(2i\epsilon_{abc} \sigma_c) = i\epsilon_{abc} t_c,$$

so the structure constants of $\mathfrak{su}(2)$ in the Hermitian basis $\{t_a\}$ are $f_{ab}^c = i\epsilon_{abc}$ (the factor of i is a consequence of using Hermitian, rather than anti-Hermitian, generators; it disappears if one instead uses $\Theta_a = it_a$, for which $[\Theta_a, \Theta_b] = -\epsilon_{abc} \Theta_c$, real structure constants as required by Definition 24.1 for a *real* Lie algebra). We will use exactly this Hermitian convention, standard in physics, throughout Section 31 below.

By Theorem 25.2(a), $U(t) = \exp(it \hat{n} \cdot \vec{t})$ for a unit vector $\hat{n} \in \mathbb{R}^3$ is the one-parameter subgroup of $SU(2)$ describing rotation about the axis $\hat{n}$ by angle t (up to the familiar factor of 2 relating $SU(2)$ and $SO(3)$ rotation angles, discussed in Example 27.3 below); and by Theorem 25.2(c), since $SU(2)$ is compact and connected, *every* element of $SU(2)$ is of the form $\exp(i\vec{\xi} \cdot \vec{t})$ for some $\vec{\xi} \in \mathbb{R}^3$ (though the representation is not unique, since $\hat{n} \cdot \vec{t}$ and its negative both generate the same one-parameter subgroup traversed in opposite directions, and $|\vec{\xi}|$ is only defined modulo 4π along any fixed axis).

26 Left- and Right-Invariant (Haar) Measure

Although this chapter is chiefly concerned with the purely algebraic object $\mathfrak{g}$, one analytic fact about the group G itself is used so pervasively (already in Part B's proof that every representation of a compact group is equivalent to a unitary one, where a finite group's averaging $\frac{1}{|G|} \sum_{g \in G}$ had to be replaced by an integral $\int_G dg$ over the group) that we pause to construct it properly here.

Definition 26.1. Let G be an n -dimensional Lie group with a coordinate patch $x = (x^1, \dots, x^n) \mapsto \Lambda(x) \in G$, $\Lambda(0) = e$. A measure $d\mu(x) = \rho(x) d^n x$ (with $\rho(x) > 0$ smooth) is called **left-invariant** if

$$\int_G f(x_0 \cdot g) d\mu(g) = \int_G f(g) d\mu(g)$$

for every integrable f and every fixed $x_0 \in G$ (writing group elements simply as g, x_0 rather than $\Lambda(x), \Lambda(x_0)$); equivalently, for every measurable $S \subseteq G$, $\mu(x_0 S) = \mu(S)$. A measure is **right-invariant** if instead $\mu(S x_0) = \mu(S)$ for all x_0 .

26.1 Construction and the Defining Functional Equation

Write the group composition law in coordinates as $\Lambda(x_0)\Lambda(x) = \Lambda(A(x_0, x))$ for a smooth function A (so that $x' = A(x_0, x)$ is the coordinate of the left-translate of $\Lambda(x)$ by $\Lambda(x_0)$); note $A(x_0, 0) = x_0$, since $\Lambda(x_0)\Lambda(0) = \Lambda(x_0)e = \Lambda(x_0)$.

Theorem 26.1. A density $\rho_L(x)$ defines a left-invariant measure $d\mu_L = \rho_L(x) d^n x$ if and only if it satisfies, for all x_0, x ,

$$\rho_L(A(x_0, x)) \left| \det \frac{\partial A(x_0, x)}{\partial x} \right| = \rho_L(x). \quad (4)$$

In particular, setting $x = 0$ (and using $A(x_0, 0) = x_0$),

$$\rho_L(x_0) = \rho_L(0) \left| \det \frac{\partial A(x_0, x)}{\partial x} \right|_{x=0}^{-1}, \quad (5)$$

so ρ_L is completely determined, up to the overall normalisation constant $\rho_L(0)$, by the Jacobian of left translation evaluated at the identity. An entirely analogous construction, using $\Lambda(x)\Lambda(x_0) = \Lambda(A'(x, x_0))$ and differentiating with respect to the *first* argument at $x = 0$, produces the right-invariant density ρ_R .

Proof. Left-invariance requires $\int f(x_0 \cdot g) \rho_L(g) d^n g = \int f(g) \rho_L(g) d^n g$ for every f . Substituting $g = \Lambda(x)$ in the left-hand integral and changing variables to $x' = A(x_0, x)$ (so $x = A(x_0, \cdot)^{-1}(x')$, a diffeomorphism for fixed x_0 , with $f(x_0 \cdot \Lambda(x)) = f(\Lambda(x'))$),

$$\int f(\Lambda(x')) \rho_L(x(x')) \left| \det \frac{\partial x}{\partial x'} \right| d^n x' = \int f(\Lambda(x)) \rho_L(x) d^n x \quad \forall f,$$

and since this must hold for *every* test function f , the two integrands, viewed as functions of the (now shared) integration variable x' , must agree pointwise:

$$\rho_L(x(x')) \left| \det \frac{\partial x}{\partial x'} \right| = \rho_L(x') \quad \text{for all } x',$$

where $x = x(x')$ denotes the inverse change of variables $x' = A(x_0, x)$. Substituting $x' = A(x_0, x)$ back in (equivalently, just relabelling the dummy variable $x' \rightarrow x$ and using $\det(\partial x / \partial x') = [\det(\partial x' / \partial x)]^{-1} = [\det(\partial A(x_0, x) / \partial x)]^{-1}$) gives

$$\rho_L(A(x_0, x)) \left| \det \frac{\partial A(x_0, x)}{\partial x} \right| = \rho_L(x),$$

which is Eq. (4). Equation (5) is the special case $x = 0$. The right-invariant case is identical with the roles of the two arguments of the group law exchanged. $\square$

Remark. A fundamental theorem of Lie theory (which we do not prove here) guarantees that a left-invariant measure — and likewise a right-invariant one — always *exists*, and is unique up to an overall positive multiplicative constant; Eq. (5) exhibits it explicitly once the group law $A(x_0, x)$ is known in some coordinate system. This is the **Haar measure** of G . For a *finite* group, the (left- and right-invariant, and in that case automatically equal) Haar measure is simply the uniform counting measure $\rho(g) \equiv 1$ used throughout Part A, and Eq. (4) reduces to the statement underlying the rearrangement theorem: summing $f(x_0 g)$ over all $g \in G$ reproduces $\sum_g f(g)$.

26.2 Compact Groups Are Unimodular

In general, the left- and right-invariant measures of a Lie group need not coincide (a simple non-compact counterexample is the “ $ax + b$ ” group of affine transformations $x \mapsto ax + b$ of $\mathbb{R}$, $a > 0$). For *compact* groups, however, the two always agree, which is exactly the property used in Part B to generalise the finite-group unitarity proof to compact Lie groups.

Theorem 26.2. If G is a compact Lie group, its left-invariant and right-invariant Haar measures coincide (up to the normalisation constant): G is **unimodular**.

Proof. Fix a left-invariant measure μ_L . For each $x_0 \in G$, the right-translate of μ_L , defined by $\mu_L^{x_0}(S) := \mu_L(Sx_0)$, is again left-invariant (since $\mu_L^{x_0}(gS) = \mu_L(gSx_0) = \mu_L(Sx_0) = \mu_L^{x_0}(S)$, using left-invariance of μ_L itself). By the uniqueness (up to scale) of the left-invariant measure, there must exist a positive number $\Delta(x_0)$, called the **modular function**, such that

$$\mu_L(Sx_0) = \Delta(x_0) \mu_L(S) \quad \text{for all measurable } S.$$

Applying this twice,

$$\mu_L(Sx_0x_1) = \Delta(x_1) \mu_L(Sx_0) = \Delta(x_0)\Delta(x_1) \mu_L(S),$$

while directly $\mu_L(S(x_0x_1)) = \Delta(x_0x_1) \mu_L(S)$; comparing, $\Delta(x_0x_1) = \Delta(x_0)\Delta(x_1)$, i.e. $\Delta : G \rightarrow (\mathbb{R}_{>0}, \times)$ is a continuous group homomorphism.

Now use compactness: $\Delta(G)$ is the continuous image of a compact set, hence a compact subgroup of $(\mathbb{R}_{>0}, \times)$. But $(\mathbb{R}_{>0}, \times) \cong (\mathbb{R}, +)$ via the logarithm, and the only compact subgroup of $(\mathbb{R}, +)$ is $\{0\}$ (any other subgroup is either discrete and unbounded, e.g. $c\mathbb{Z}$ for $c \neq 0$, or dense in $\mathbb{R}$ and hence non-compact, e.g. if it contains two incommensurate numbers); so $\log \Delta(G) = \{0\}$, i.e. $\Delta \equiv 1$. Therefore $\mu_L(Sx_0) = \mu_L(S)$ for every x_0 : μ_L is *also* right-invariant, hence equal (up to the normalisation already fixed) to μ_R . $\square$

Remark. For a compact group one conventionally normalises the (now unique, bi-invariant) Haar measure to have total volume $\int_G d\mu = 1$, in direct analogy with $\frac{1}{|G|} \sum_{g \in G}$ for a finite group; with this normalisation, the substitution rule $\int_G dg F(g g_0) = \int_G dg F(g)$ used in Part B's compact-group unitarity proof is exactly the statement that μ is right-invariant (equivalently, by Theorem 26.2, left-invariant), i.e. the continuous analogue of the rearrangement theorem for finite groups.

Example 26.1. The Haar measure of $SU(2)$. We compute the (bi-invariant, by Theorem 26.2, since $SU(2)$ is compact) Haar measure explicitly. Write $u \in SU(2)$ as $u = \begin{pmatrix} \alpha & \beta \\ -\beta^* & \alpha^* \end{pmatrix}$ with $\alpha, \beta \in \mathbb{C}$, $|\alpha|^2 + |\beta|^2 = 1$: this exhibits $SU(2)$ as the unit sphere $S^3 \subset \mathbb{R}^4$ via $(\operatorname{Re} \alpha, \operatorname{Im} \alpha, \operatorname{Re} \beta, \operatorname{Im} \beta)$. Since left- and right-multiplication by a fixed $u_0 \in SU(2)$ act on $\mathbb{C}^2 = (\alpha, \beta)$ (viewed as $\mathbb{R}^4$) as an orthogonal (indeed unitary) linear transformation — it preserves $|\alpha|^2 + |\beta|^2$, being built from unitary 2×2 matrices acting on the columns of u — both translations preserve the ordinary round volume element of $S^3 \subset \mathbb{R}^4$. Hence *the Haar measure of $SU(2)$ is, up to normalisation, the round volume element of the unit 3-sphere*, which we now make explicit in coordinates.

Set $\alpha = \cos \frac{\theta}{2} e^{i\psi}$, $\beta = \sin \frac{\theta}{2} e^{i\phi}$, with $\theta \in [0, \pi]$, ϕ, ψ each ranging over a period of 2π (these are Hopf coordinates on S^3 ; one checks directly that $|\alpha|^2 + |\beta|^2 = \cos^2 \frac{\theta}{2} + \sin^2 \frac{\theta}{2} = 1$ automatically). A direct computation of the induced metric on S^3 (or, equivalently, of the Jacobian in Eq. (5)) gives the volume element

$$d\mu_{SU(2)} \propto \sin \theta \, d\theta \, d\phi \, d\psi \quad (\theta \in [0, \pi]),$$

exactly the familiar $\sin \theta \, d\theta \, d\phi$ of ordinary spherical coordinates, now with an extra flat angle ψ . (Concretely: writing $\alpha = x_1 + ix_2$, $\beta = x_3 + ix_4$ and computing $ds^2 = \sum_i dx_i^2$ restricted to the sphere in these coordinates gives $ds^2 = \frac{1}{4}d\theta^2 + \cos^2 \frac{\theta}{2} d\psi^2 + \sin^2 \frac{\theta}{2} d\phi^2$,

whose associated volume form is $\sqrt{\det g} d\theta d\phi d\psi = \frac{1}{2} \sin \frac{\theta}{2} \cos \frac{\theta}{2} d\theta d\phi d\psi = \frac{1}{4} \sin \theta d\theta d\phi d\psi$; as a check, $\int_0^\pi \int_0^{2\pi} \int_0^{2\pi} \frac{1}{4} \sin \theta d\theta d\phi d\psi = \frac{1}{4} \cdot 2 \cdot 2\pi \cdot 2\pi = 2\pi^2$, correctly reproducing the volume of the unit 3-sphere.) Note that at $\theta = 0$ the coordinate ϕ becomes degenerate (only $\alpha = e^{i\psi}$ survives), so the locus $\theta = 0$ is not a single point but the whole circle $\{(\cos 0 \cdot e^{i\psi}, 0) : \psi \in [0, 2\pi)\}$ traced out by ψ ; likewise $\theta = \pi$ degenerates ψ and is the circle traced out by ϕ . These two circles are exactly the degenerate fibres of the Hopf fibration $S^3 \rightarrow S^2$ over the poles of the base S^2 — unlike the actual poles $\theta = 0, \pi$ of the ordinary 2-sphere, they are one-dimensional, not single points — but since $\sin \theta = 0$ there, they still correctly carry zero measure in $d\mu_{SU(2)}$.

27 The Lie Group–Lie Algebra Correspondence

We now make precise the dictionary between structures on the group G and structures on its Lie algebra $\mathfrak{g}$, extending the discussion of subgroups and normal subgroups from Part A to the continuous setting.

27.1 Subalgebras and Ideals

Definition 27.1. Let $\mathfrak{X}$ be a Lie algebra. A subspace $\mathfrak{X}' \subseteq \mathfrak{X}$ is called a **Lie subalgebra** if it is closed under the bracket:

$$a, b \in \mathfrak{X}' \implies [a, b] \in \mathfrak{X}'.$$

A subalgebra $\mathfrak{z} \subseteq \mathfrak{X}$ is called an **invariant subalgebra**, or **ideal**, if it absorbs brackets with the *whole* of $\mathfrak{X}$, not merely with itself:

$$a \in \mathfrak{X}, b \in \mathfrak{z} \implies [a, b] \in \mathfrak{z}.$$

(Every ideal is in particular a subalgebra, taking $a \in \mathfrak{z} \subseteq \mathfrak{X}$ above.)

Theorem 27.1. Let G be a Lie group with Lie algebra $\mathfrak{g}$.

- (a) If $H \leq G$ is a (connected) Lie subgroup, its Lie algebra $\mathfrak{h}$ is a Lie subalgebra of $\mathfrak{g}$.
- (b) If, moreover, $H \trianglelefteq G$ is a *normal* (invariant) subgroup, then $\mathfrak{h}$ is an **ideal** of $\mathfrak{g}$.
- (c) Conversely, every Lie subalgebra $\mathfrak{h} \subseteq \mathfrak{g}$ is the Lie algebra of a unique connected Lie subgroup $H \leq G$; and $H \trianglelefteq G$ if $\mathfrak{h}$ is an ideal.

Proof. (a) A tangent vector to a curve lying entirely in H is, *a fortiori*, a tangent vector to a curve in G , so $\mathfrak{h} \subseteq \mathfrak{g}$ as a subspace; and $\mathfrak{h}$ is closed under the bracket by Theorem 25.1 applied within H itself (the curve $C(s)$ built there from two curves in H again lies in H , since H is a subgroup).

(b) Let $h \in \mathfrak{h}$ be tangent at $s = 0$ to a curve $H(s) \subset H$, and let $a \in \mathfrak{g}$ be tangent at $t = 0$ to a curve $A(t) \subset G$. Normality of H means $A(t)H(s)A(t)^{-1} \in H$ for *every* t and *every* s . Fix t : the curve $s \mapsto A(t)H(s)A(t)^{-1}$ lies entirely in H , so its tangent vector at $s = 0$,

$$\left. \frac{d}{ds} [A(t)H(s)A(t)^{-1}] \right|_{s=0} = A(t)hA(t)^{-1},$$

is an element of $\mathfrak{h}$ — *for every value of t* . Thus $t \mapsto A(t)hA(t)^{-1}$ is a curve lying entirely in the fixed subspace $\mathfrak{h} \subset \mathfrak{g}$, so its own derivative at $t = 0$ is again in $\mathfrak{h}$ (differentiating a curve valued

in a linear subspace stays in that subspace):

$$[a, h] = \left. \frac{d}{dt} [A(t)hA(t)^{-1}] \right|_{t=0} = ah - ha \in \mathfrak{h}.$$

Since $a \in \mathfrak{g}$ and $h \in \mathfrak{h}$ were arbitrary, $\mathfrak{h}$ is an ideal.

(c) This is the (harder) direction of Lie's second and third theorems: every subalgebra integrates to a unique connected subgroup (obtained, e.g., as the subgroup generated by $\exp(\mathfrak{h})$), and the normality statement follows by reversing the differentiation argument of (b) and re-exponentiating. We do not reprove Lie's theorems here; the content relevant to us is the dictionary itself, which we use freely below. $\square$

Theorem 27.2. If $\mathfrak{h}$ and $\mathfrak{k}$ are both ideals of a Lie algebra $\mathfrak{g}$, then so is their **bracket** $[\mathfrak{h}, \mathfrak{k}] := \text{span}\{[b, c] : b \in \mathfrak{h}, c \in \mathfrak{k}\}$.

Proof. Let $a \in \mathfrak{g}$, $b \in \mathfrak{h}$, $c \in \mathfrak{k}$. The Jacobi identity, rearranged, gives

$$[a, [b, c]] = [[a, b], c] + [b, [a, c]].$$

(This is simply $a*(b*c) = -b*(c*a) - c*(a*b) = -b*(c*a) + (a*b)*c$, using antisymmetry twice, which is $[[a, b], c] + [b, [a, c]]$ after renaming $(a*b)*c = [[a, b], c]$ and $-b*(c*a) = b*(a*c) = [b, [a, c]]$.) Now $[a, b] \in \mathfrak{h}$ because $\mathfrak{h}$ is an ideal, so $[[a, b], c] \in [\mathfrak{h}, \mathfrak{k}]$; and $[a, c] \in \mathfrak{k}$ because $\mathfrak{k}$ is an ideal, so $[b, [a, c]] \in [\mathfrak{h}, \mathfrak{k}]$ (indeed already in $\text{span}\{[b, c']\} \subseteq [\mathfrak{h}, \mathfrak{k}]$). Hence $[a, [b, c]] \in [\mathfrak{h}, \mathfrak{k}]$ for every generator $[b, c]$ of $[\mathfrak{h}, \mathfrak{k}]$ and every $a \in \mathfrak{g}$, so (by linearity in the second slot) $[a, z] \in [\mathfrak{h}, \mathfrak{k}]$ for every $z \in [\mathfrak{h}, \mathfrak{k}]$: it is an ideal. $\square$

27.2 Homomorphisms and Isomorphisms

Definition 27.2. A map $\Phi : G \rightarrow G'$ between Lie groups is a (Lie group) **homomorphism** if it is a group homomorphism, $\Phi(g_1g_2) = \Phi(g_1)\Phi(g_2)$, and is analytic. Every such Φ induces a map $\psi : \mathfrak{g} \rightarrow \mathfrak{g}'$ on the Lie algebras,

$$\psi(a) := \left. \frac{d}{dt} \Phi(\exp(ta)) \right|_{t=0} \quad (a \in \mathfrak{g}),$$

which is automatically *linear* and satisfies $\psi([a, b]) = [\psi(a), \psi(b)]$ (a **Lie algebra homomorphism**): linearity and the bracket property both follow by differentiating Φ along the constructions of Theorems 25.1 and 25.2 and using that Φ is a group homomorphism at every step.

Φ is a (Lie group) **isomorphism** if it is, in addition, one-to-one and onto (with analytic inverse); ψ is then a Lie algebra isomorphism, $\psi : \mathfrak{g} \xrightarrow{\sim} \mathfrak{g}'$.

Example 27.1. Let $\Phi : U(1) \rightarrow SO(2)$, $\Phi(e^{ix}) = \begin{pmatrix} \cos x & -\sin x \\ \sin x & \cos x \end{pmatrix}$, an isomorphism of Lie groups (both are the circle group, realised in two different ways). The Lie algebra of $U(1)$ is $\mathfrak{u}(1) = i\mathbb{R}$ (purely imaginary numbers, under commutation — trivially abelian, since $U(1) \subset \mathbb{C}$ is commutative), with generator $\Theta = i$; the induced map is

$$\psi(i) = \left. \frac{d}{dt} \Phi(e^{it}) \right|_{t=0} = \left. \frac{d}{dt} \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix} \right|_{t=0} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

which is indeed the generator of $\mathfrak{so}(2)$ found in Part B, confirming ψ is (as it must be, Φ being an isomorphism) an isomorphism $\mathfrak{u}(1) \xrightarrow{\sim} \mathfrak{so}(2)$ of the two (one-dimensional, automatically abelian) Lie algebras.

Theorem 27.3. Let $\Phi : G \rightarrow G'$ be a Lie group homomorphism with **kernel** $K = \{g \in G : \Phi(g) = e'\}$. Then:

- (a) K is a normal (invariant) subgroup of G .
- (b) If K is **discrete** (i.e. every point of K has a neighbourhood in G containing no other point of K), then the induced Lie algebra map $\psi : \mathfrak{g} \rightarrow \mathfrak{g}'$ is **injective**. If, moreover, Φ is **surjective** (equivalently, since Φ is then an open map onto a group of the same dimension as G/K , whenever G' is connected and $\dim G = \dim G'$), ψ is in fact an **isomorphism**.

Proof. (a) K is a subgroup since Φ is a homomorphism ($\Phi(k_1 k_2) = \Phi(k_1)\Phi(k_2) = e'e' = e'$, $\Phi(k^{-1}) = \Phi(k)^{-1} = e'$), and normal since $\Phi(gkg^{-1}) = \Phi(g)\Phi(k)\Phi(g)^{-1} = \Phi(g)e'\Phi(g)^{-1} = e'$ for any $g \in G, k \in K$.

(b) Since K is discrete, there is a neighbourhood E of $e \in G$ with $E \cap K = \{e\}$. Shrinking E if necessary, Φ is one-to-one on E : if $\Phi(g_1) = \Phi(g_2)$ for $g_1, g_2 \in E$, then $g_1 g_2^{-1} \in K \cap (EE^{-1})$, which, for E small enough, forces $g_1 g_2^{-1} = e$, i.e. $g_1 = g_2$. In particular Φ is injective in a neighbourhood of e , so its derivative there, $\psi : \mathfrak{g} \rightarrow \mathfrak{g}'$, is injective: if $\psi(a) = 0$ for some $a \neq 0$, the one-parameter subgroup $t \mapsto \exp(ta)$ would be mapped by Φ to a curve in G' tangent to 0 at $t = 0$, and standard facts about analytic maps then force $\Phi(\exp(ta)) = e'$ for all small t , contradicting injectivity of Φ on E (unless $a = 0$). This proves the first (injectivity) claim of (b), *without* any further hypothesis.

For the second (isomorphism) claim, injectivity of ψ already gives $\dim \mathfrak{g} \leq \dim \mathfrak{g}'$. If in addition Φ is surjective, then (assuming, as is the case in every application below, that G' is connected) Φ is also an *open* map near e — a standard consequence of the constant-rank theorem once one knows Φ is a local homeomorphism onto its image and that image, being a subgroup containing a neighbourhood of e' , is all of the connected group G' — so Φ maps a neighbourhood of e *onto* a neighbourhood of e' , forcing $\dim \mathfrak{g} = \dim \mathfrak{g}'$ and hence ψ to be onto as well. Being linear, injective and (now) onto, ψ is a vector-space isomorphism, and we already know it respects the bracket; hence ψ is a Lie algebra isomorphism. (Every use of this theorem in these notes — $\mathbb{R} \rightarrow SO(2)$, $SU(2) \rightarrow SO(3)$, $\tilde{G} \rightarrow \tilde{G}/K$ — has Φ manifestly surjective, so the isomorphism conclusion applies.) $\square$

Remark. Significance: the Lie algebra determines the Lie group only locally. Theorem 27.3 says that a homomorphism with discrete kernel gives an *isomorphism* of Lie algebras, even though Φ itself need not be one-to-one on all of G (only near the identity) — G and G' can be globally quite different (different topology, e.g. one compact and the other not, or one simply connected and the other not) while sharing exactly the same Lie algebra. Two Lie groups with the same Lie algebra are called **locally isomorphic**. Determining *which* globally distinct Lie groups share a given Lie algebra is precisely the subject of Section 30 below.

Example 27.2. $(\mathbb{R}, +)$ and $SO(2)$ are locally, but not globally, isomorphic. Consider $\Phi : \mathbb{R} \rightarrow SO(2)$, $\Phi(t) = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix}$ (rotation by t). This is a Lie group homomorphism,

$\Phi(t_1 + t_2) = \Phi(t_1)\Phi(t_2)$, with kernel $K = 2\pi\mathbb{Z}$: discrete, but countably *infinite*. By Theorem 27.3, the induced map $\psi : (\mathfrak{R}, +) \rightarrow \mathfrak{so}(2)$ is a Lie algebra isomorphism — both are one-dimensional abelian Lie algebras, so this is no surprise — yet Φ itself is very far from a group isomorphism: it is an ∞ -to-one covering map. Globally, $(\mathbb{R}, +)$ is non-compact and simply connected (no non-contractible loops, being an interval), whereas $SO(2) \cong U(1)$ is compact and *not* simply connected (a loop that winds around the circle cannot be contracted to a point without leaving the group). This is the simplest possible illustration of Section 30: $(\mathbb{R}, +)$ is the *universal covering group* of $SO(2)$, and every other connected Lie group with the same Lie algebra $\mathfrak{so}(2)$ is a quotient $\mathbb{R}/(c\mathbb{Z})$ for some $c > 0$, all of which are isomorphic to $SO(2)$ after rescaling t .

Example 27.3. The covering homomorphism $SU(2) \rightarrow SO(3)$. Every $u \in SU(2)$ acts on the real, three-dimensional space of traceless Hermitian 2×2 matrices, $X = \vec{x} \cdot \vec{\sigma}$ ($\vec{x} \in \mathbb{R}^3$), by $X \mapsto uXu^\dagger$; since $uXu^\dagger$ is again traceless Hermitian, this defines $R(u) \in GL(3, \mathbb{R})$ via $R(u)\vec{x} \cdot \vec{\sigma} := u(\vec{x} \cdot \vec{\sigma})u^\dagger$. Because $\det(uXu^\dagger) = \det(X) = -|\vec{x}|^2$ (using $\det(\vec{x} \cdot \vec{\sigma}) = -(x_1^2 + x_2^2 + x_3^2)$) and $\text{tr}((uXu^\dagger)(uX'u^\dagger)) = \text{tr}(XX')$, the map $R(u)$ preserves the length of $\vec{x}$, so $R(u) \in O(3)$; continuity in u and $R(1) = 1$ (with $SU(2)$ connected) force $R(u) \in SO(3)$. One checks directly $R(u_1u_2) = R(u_1)R(u_2)$, so $\Phi := R : SU(2) \rightarrow SO(3)$ is an analytic Lie group homomorphism.

Its kernel consists of those u with $uXu^\dagger = X$ for *every* traceless Hermitian X , i.e. u commutes with all of $\mathfrak{su}(2)$ (as Hermitian matrices via $X = i \cdot (\text{anti-Hermitian generator})$); by Schur's Lemma (Part B) applied to the (irreducible, as we verify in Section 31) defining representation of $SU(2)$ on $\mathbb{C}^2$, u must be a multiple of the identity, and unitarity plus $\det u = 1$ forces $u = \pm 1$. So $K = \ker \Phi = \{1, -1\} \cong \mathbb{Z}_2$: discrete (indeed finite). By Theorem 27.3, Φ induces an isomorphism $\mathfrak{su}(2) \xrightarrow{\sim} \mathfrak{so}(3)$ (consistent with both being three-dimensional with structure constants $i\epsilon_{abc}$, resp. ϵ_{abc} , in suitable bases), while globally Φ is a **two-to-one** covering map: $SO(3) \cong SU(2)/\mathbb{Z}_2$. This is the algebraic origin of the physicist's "spin- $\frac{1}{2}$ particles pick up a minus sign under a 2π rotation": a full (2π) rotation in $SO(3)$ lifts to *either* $u = \exp(i\pi \hat{n} \cdot \vec{\sigma}) = -1$ *or* back to $+1$ depending on which of the two sheets of the double cover one started on, and a 4π rotation is needed to guarantee a return to $+1$.

28 Direct Products, Direct Sums, and Tensor Products

Before turning to the adjoint representation and to the classification of representations of $\mathfrak{su}(2)$, we need some general constructions for building new groups, vector spaces, and representations out of old ones. This is also the right place to sort out a persistent and genuinely confusing notational habit in the physics literature: the symbol $\otimes$ is routinely used for *two completely different constructions* — combining two *groups* into a bigger group, and combining two *vector spaces* into a bigger vector space — which happen to behave very differently (additively versus multiplicatively in dimension). We first give careful, formal definitions of all three constructions in play, and then show how each is used to build representations.

28.1 Three Constructions

Definition 28.1. Let G, H be groups. Their **direct product** $G \times H$ is the group whose underlying set is the Cartesian product $\{(g, h) : g \in G, h \in H\}$, with componentwise multiplication

$$(g_1, h_1) \cdot (g_2, h_2) := (g_1g_2, h_1h_2),$$

identity (e_G, e_H) , and inverse $(g, h)^{-1} = (g^{-1}, h^{-1})$. If G, H are Lie groups, so is $G \times H$, with

$$\dim(G \times H) = \dim G + \dim H$$

(as a manifold, $G \times H$ is simply the product manifold, and near the identity a pair of independent parameters (ξ, η) for G and H separately parametrises $G \times H$). The direct product describes two symmetries acting *independently*, with no interaction between the two factors: G and H commute inside $G \times H$ in the sense that (g, e_H) and (e_G, h) always commute for any $g \in G, h \in H$.

Definition 28.2. Let V, W be (finite-dimensional, real or complex) vector spaces with bases $\{e_i^V\}_{i=1}^{\dim V}, \{e_j^W\}_{j=1}^{\dim W}$. Their **direct sum** $V \oplus W$ is the vector space of ordered pairs (v, w) , $v \in V, w \in W$, with componentwise addition and scalar multiplication, $\alpha(v_1, w_1) + \beta(v_2, w_2) := (\alpha v_1 + \beta v_2, \alpha w_1 + \beta w_2)$; a basis is given by the union $\{(e_i^V, 0)\} \cup \{(0, e_j^W)\}$, so that

$$\dim(V \oplus W) = \dim V + \dim W.$$

Theorem 28.1. There exist (unique) linear maps

$$\iota_V : V \rightarrow V \oplus W, \quad \iota_W : W \rightarrow V \oplus W, \quad \pi_V : V \oplus W \rightarrow V, \quad \pi_W : V \oplus W \rightarrow W$$

(the natural inclusions and projections) with $\pi_V \circ \iota_V = \mathbb{1}_V$, $\pi_W \circ \iota_W = \mathbb{1}_W$, $\pi_V \circ \iota_W = 0$, $\pi_W \circ \iota_V = 0$.

Proof. Take $\iota_V(v) = (v, 0)$, $\iota_W(w) = (0, w)$, $\pi_V(v, w) = v$, $\pi_W(v, w) = w$; all the stated identities are immediate from these definitions. $\square$

Definition 28.3. Let V, W be as above. Their **tensor product** $V \otimes W$ is the vector space with basis $\{e_i^V \otimes e_j^W\}$ (formal symbols, one for each *pair* (i, j) , so $\dim V \cdot \dim W$ of them in total), so that

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

For general vectors $v = \sum_i v^i e_i^V \in V$, $w = \sum_j w^j e_j^W \in W$, one defines

$$v \otimes w := \sum_{i,j} v^i w^j e_i^V \otimes e_j^W \in V \otimes W,$$

which is manifestly bilinear in (v, w) : $(\alpha v_1 + \beta v_2) \otimes w = \alpha(v_1 \otimes w) + \beta(v_2 \otimes w)$, and likewise in the second slot.

Theorem 28.2. (Universal property.) To every bilinear map $\Phi : V \times W \rightarrow U$ (for any vector space U) there corresponds a *unique* linear map $\hat{\Phi} : V \otimes W \rightarrow U$ such that $\hat{\Phi}(v \otimes w) = \Phi(v, w)$ for all $v \in V, w \in W$.

Proof. Define $\hat{\Phi}$ on the basis by $\hat{\Phi}(e_i^V \otimes e_j^W) := \Phi(e_i^V, e_j^W)$ and extend linearly; this is well defined since $\{e_i^V \otimes e_j^W\}$ is a *basis* (no relations to check), and one verifies $\hat{\Phi}(v \otimes w) = \sum_{i,j} v^i w^j \Phi(e_i^V, e_j^W) = \Phi(v, w)$ using bilinearity of Φ . Uniqueness holds because a linear map is determined by its values on a basis, and every basis vector $e_i^V \otimes e_j^W$ is of the required product form $v \otimes w$. $\square$

Remark. A notational warning. The table below summarises the three constructions just defined; note in particular that the direct product of *groups* behaves *additively* in dimension, exactly like the direct sum of vector spaces, and *not* like the tensor product.

Construction	Combines	Elements	Dimension
Direct product	two <i>groups</i>	pairs (g, h)	$\dim G + \dim H$
Direct sum	two <i>vector spaces</i>	pairs (v, w)	$\dim V + \dim W$
Tensor product	two <i>vector spaces</i>	sums of $v \otimes w$	$\dim V \cdot \dim W$

Despite this, physicists very commonly write $G \otimes H$ for the direct product of *groups* — the Standard Model gauge group, for instance, is almost universally written $SU(3) \otimes SU(2) \otimes U(1)$, even though, by Definition 28.1, this is the **direct product** $SU(3) \times SU(2) \times U(1)$: three mutually commuting, non-interacting symmetry groups, of total dimension $8 + 3 + 1 = 12$ (*not* $8 \times 3 \times 1 = 24$). The symbol $\otimes$ is, properly, reserved for the tensor product of *vector spaces* (or of representations, Section 28.2 below) — as in “the quark fields transform in the $(\mathbf{3}, \mathbf{2}, \frac{1}{6})$ representation”, which really is built as a tensor product of representation spaces, dimension $3 \times 2 = 6$ per generation. Throughout the rest of this chapter we use $\times$ exclusively for direct products of groups and $\otimes$ exclusively for tensor products of vector spaces/representations, and we flag explicitly, below, exactly how the two constructions interact.

28.2 Representations Built From These Constructions

The construction we will actually need throughout the rest of this chapter combines two representations of the *same* group G into a new representation of G , again on the tensor product vector space $V \otimes W$.¹

Definition 28.4. Let Γ^V, Γ^W be representations of a group G on V, W respectively. The **tensor product representation** $\Gamma^V \otimes \Gamma^W$ acts on $V \otimes W$ (Definition 28.3) by

$$(\Gamma^V \otimes \Gamma^W)(g)(v \otimes w) := (\Gamma^V(g)v) \otimes (\Gamma^W(g)w) \quad (g \in G),$$

extended linearly to all of $V \otimes W$ (well defined by Theorem 28.2, applied to the bilinear map $(v, w) \mapsto \Gamma^V(g)v \otimes \Gamma^W(g)w$). One checks immediately that this is again a representation: $(\Gamma^V \otimes \Gamma^W)(g_1 g_2) = (\Gamma^V \otimes \Gamma^W)(g_1)(\Gamma^V \otimes \Gamma^W)(g_2)$, inherited directly from $\Gamma^V(g_1 g_2) = \Gamma^V(g_1)\Gamma^V(g_2)$ and likewise for Γ^W .

Remark. Not every element of $V \otimes W$ is of the simple form $v \otimes w$. A general element $\sum_{ij} A^{ij} e_i^V \otimes e_j^W$ is of product form $v \otimes w$ only if the matrix $A^{ij} = v^i w^j$ has rank one; a generic matrix A does not. This is, quite literally, the mathematical statement of **quantum entanglement**: if V, W are the state spaces of two subsystems, a generic joint state cannot be written as a simple product $v \otimes w$ of a state of the first subsystem with a state of the second — it is *entangled*.

¹A more refined treatment would distinguish this from the **external** tensor product representation $\Gamma^V \boxtimes \Gamma^W$ of the *direct product group* $G \times H$ (possibly $G \neq H$), defined by $(\Gamma^V \boxtimes \Gamma^W)(g, h) := \Gamma^V(g) \otimes \Gamma^W(h)$ for $(g, h) \in G \times H$ — the construction physicists mean, for instance, when they say quark fields transform in the $(\mathbf{3}, \mathbf{2}, \frac{1}{6})$ of $SU(3) \times SU(2) \times U(1)$, with each factor acting on its own tensor slot. The construction defined here is recovered from it by restricting to the diagonal subgroup $\{(g, g) : g \in G\} \cong G \subset G \times G$. We will not need to distinguish the two further.

Example 28.1. A representation of a direct product group that is *not* a tensor product representation. Let $G = \{e, a\} \cong \mathbb{Z}_2$ and $H = \{e, b\} \cong \mathbb{Z}_2$, and let the direct product $G \times H$ (Definition 28.1) act on $\mathbb{C}^2$ by

$$\Gamma(e, e) = \mathbb{1}, \quad \Gamma(a, e) = \Gamma(e, b) = X, \quad \Gamma(a, b) = X^2 = \mathbb{1}, \quad X := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

One checks directly that this respects the group law of $G \times H$: e.g. $(a, e)(e, b) = (a, b)$ in $G \times H$, and indeed $\Gamma(a, e)\Gamma(e, b) = X \cdot X = \mathbb{1}$, while $\Gamma(a, b)$ was *defined* to be $\mathbb{1}$ — consistent. So Γ is a genuine two-dimensional representation of $G \times H \cong \mathbb{Z}_2 \times \mathbb{Z}_2$.

This is not a tensor product representation, i.e. there is no way to write $\Gamma = \Gamma^V \otimes \Gamma^W$ for representations Γ^V of G , Γ^W of H (necessarily one-dimensional, since $1 \cdot 2 = 2 \cdot 1 = 2$ are the only ways to factor $\dim = 2$). Suppose Γ^W is one-dimensional, i.e. a character $\chi : H \rightarrow \mathbb{C}^\times$; then $\Gamma(g, h)$ would have to equal $\Gamma^V(g) \cdot \chi(h)$, forcing in particular

$$\Gamma(a, e) = \Gamma^V(a) \cdot \chi(e) = \Gamma^V(a) \cdot 1, \quad \Gamma(e, b) = \Gamma^V(e) \cdot \chi(b) = 1 \cdot \chi(b) = \chi(b) \mathbb{1} :$$

the second identity would force $\Gamma(e, b)$ to be a *scalar multiple of the identity matrix*. But $\Gamma(e, b) = X$ has eigenvalues $+1, -1$ and so is manifestly not proportional to $\mathbb{1}$ — a contradiction. The symmetric argument (with Γ^V one-dimensional instead) rules out the other factorisation. So Γ genuinely mixes the two basis vectors in a way that cannot be attributed to G alone or to H alone acting “in its own slot”, which is exactly the sense in which it fails to be a tensor product representation.

(In fact Γ is still *reducible*: diagonalising X in the basis $\psi_\pm := \frac{1}{\sqrt{2}}(1, \pm 1)$ gives $\Gamma(a, e)\psi_\pm = \pm\psi_\pm$ and $\Gamma(e, b)\psi_\pm = \pm\psi_\pm$, so Γ decomposes as a *direct sum* $(\chi_+ \otimes \chi_+) \oplus (\chi_- \otimes \chi_-)$ of two one-dimensional tensor product representations, where $\chi_+ \equiv 1$ is trivial and $\chi_-(a) = \chi_-(b) = -1$. This illustrates the general pattern: a representation of $G \times H$ need not itself be a tensor product, but — for finite or compact groups — it always decomposes into a *direct sum* of them.)

Theorem 28.3. Let t^V, t^W denote the Lie-algebra representations induced (as in Section 27) by Γ^V, Γ^W . The Lie-algebra representation induced by the tensor product $\Gamma^V \otimes \Gamma^W$ is **not** $t^V \otimes t^W$, but rather the **Leibniz combination**

$$T(a) := t^V(a) \otimes \mathbb{1}_W + \mathbb{1}_V \otimes t^W(a) \quad (a \in \mathfrak{g}).$$

Proof. By definition, $T(a) = \frac{d}{dt} [(\Gamma^V \otimes \Gamma^W)(\exp(ta))] \Big|_{t=0} = \frac{d}{dt} [\Gamma^V(\exp(ta)) \otimes \Gamma^W(\exp(ta))] \Big|_{t=0}$. The tensor product of two matrices depending on t obeys the ordinary Leibniz (product) rule for differentiation — immediate componentwise, since $(M(t) \otimes N(t))_{(ij)(kl)} = M_{ik}(t)N_{jl}(t)$ is an ordinary product of scalar functions of t — so

$$T(a) = \frac{d\Gamma^V(\exp(ta))}{dt} \Big|_0 \otimes \Gamma^W(e) + \Gamma^V(e) \otimes \frac{d\Gamma^W(\exp(ta))}{dt} \Big|_0 = t^V(a) \otimes \mathbb{1}_W + \mathbb{1}_V \otimes t^W(a). \quad \square$$

Remark. This “Leibniz rule at the Lie-algebra level” is exactly the familiar statement, from angular momentum addition in quantum mechanics, that the total angular momentum of a two-particle system is $\vec{J} = \vec{J}^{(1)} \otimes \mathbb{1} + \mathbb{1} \otimes \vec{J}^{(2)}$ (usually abbreviated $\vec{J}^{(1)} + \vec{J}^{(2)}$), *not* $\vec{J}^{(1)} \otimes \vec{J}^{(2)}$: angular momentum operators act on a tensor product by the sum rule of

Theorem 28.3, precisely because they generate (infinitesimal) group transformations, which act on tensor products by Definition 28.4. This will be the starting point for the Clebsch–Gordan decomposition of Section 32.

Definition 28.5. Given representations Γ^V on V and Γ^W on W of the *same* group G , their **direct sum** $\Gamma^V \oplus \Gamma^W$ acts on $V \oplus W$ (Definition 28.2) by $(\Gamma^V \oplus \Gamma^W)(g)(v, w) := (\Gamma^V(g)v, \Gamma^W(g)w)$; in block-matrix form, using the inclusions and projections of Theorem 28.1, this is $\begin{pmatrix} \Gamma^V(g) & 0 \\ 0 & \Gamma^W(g) \end{pmatrix}$, exactly the completely reducible form encountered in Part B. Unlike the tensor product representations above, the direct sum of two representations of G is always of *this* G itself (never of a direct product $G \times H$): it simply lets G act on V and W “side by side”, without mixing them.

Remark. Symmetric and antisymmetric tensor powers. Given a single representation Γ of G on a d -dimensional space V , its square $\Gamma \otimes \Gamma$ acts on $V \otimes V$, but this action preserves the splitting of $V \otimes V$ into the **symmetric** part $\text{Sym}^2 V = \text{span}\{e_i \otimes e_j + e_j \otimes e_i\}$ (dimension $\frac{1}{2}d(d+1)$) and the **antisymmetric** part $\Lambda^2 V = \text{span}\{e_i \otimes e_j - e_j \otimes e_i\}$ (dimension $\frac{1}{2}d(d-1)$), since $(\Gamma \otimes \Gamma)(g)$ commutes with the swap $e_i \otimes e_j \leftrightarrow e_j \otimes e_i$ for every g :

$$V \otimes V = \text{Sym}^2 V \oplus \Lambda^2 V, \quad d^2 = \frac{1}{2}d(d+1) + \frac{1}{2}d(d-1).$$

For $G = SO(d)$ acting on $V = \mathbb{R}^d$ by its defining representation, the antisymmetric part $\Lambda^2 V$ has exactly the dimension $\frac{1}{2}d(d-1) = \dim \mathfrak{so}(d)$ of the adjoint representation defined below, and indeed *is* (isomorphic to) the adjoint representation of $SO(d)$: an antisymmetric matrix $a \in \mathfrak{so}(d)$ is itself naturally an element of $\Lambda^2 \mathbb{R}^d$ (via $a = \sum_{i < j} a_{ij}(e_i \otimes e_j - e_j \otimes e_i)$), and conjugation $a \mapsto RaR^{-1} = RaR^T$ for $R \in SO(d)$ matches exactly the action of $\Gamma \otimes \Gamma$ on $\Lambda^2 V$. This is why, for instance, the generators of $SO(3)$ (three of them, $\frac{1}{2} \cdot 3 \cdot 2 = 3$) can be labelled either by an antisymmetric pair of indices $[ij]$ or, using ϵ_{ijk} , by a single vector index k — the two labellings are related by the isomorphism $\Lambda^2 \mathbb{R}^3 \cong \mathbb{R}^3$ special to three dimensions.

29 The Adjoint Representation

A Lie algebra always possesses one representation that requires no external input to construct: its action on *itself*.

Definition 29.1. Let $\mathfrak{g}$ be a Lie algebra. For $a \in \mathfrak{g}$, define the linear map $\text{ad}(a) : \mathfrak{g} \rightarrow \mathfrak{g}$ by

$$\text{ad}(a)b := [a, b] \quad (b \in \mathfrak{g}).$$

The assignment $a \mapsto \text{ad}(a)$ is called the **adjoint representation** of $\mathfrak{g}$. In a basis $\{t_a\}$ with structure constants $[t_a, t_b] = f_{ab}^c t_c$, the matrix elements of $\text{ad}(t_a)$ (an $n \times n$ matrix, $n = \dim \mathfrak{g}$, acting on the n -dimensional carrier space $\mathfrak{g}$ itself) are given by the structure constants. Since the generators t_a carry *lower* indices, we contract them with *upper* indices and write the matrix of t_a with a lower row index b and an upper column index c :

$$(T_a^{\text{ad}})_b^c := f_{ba}^c, \quad \text{so that} \quad (T_a^{\text{ad}})_b^c t_c = f_{ba}^c t_c = [t_b, t_a] = -\text{ad}(t_a)t_b.$$

Matrix products contract an upper with a lower index, $(T_a T_d)_b^e = (T_a)_b^c (T_d)_c^e$. The index order f_{ba}^c (rather than f_{ab}^c) and the resulting minus sign are exactly what makes T_a^{ad} obey the

same commutation relations as the t_a , $[T_a^{\text{ad}}, T_d^{\text{ad}}] = f_{ad}^e T_e^{\text{ad}}$, for the ordinary matrix product (see Theorem 29.1). Equivalently, T_a^{ad} is minus the transpose of the matrix $(\text{ad}(t_a))^c_b = f_{ab}^c$ one obtains by writing the coefficients of $\text{ad}(t_a)t_b$ in a column.

Theorem 29.1. ad is indeed a representation of $\mathfrak{g}$: for all $a, b \in \mathfrak{g}$,

$$[\text{ad}(a), \text{ad}(b)] = \text{ad}([a, b]).$$

Proof. Apply both sides to an arbitrary $c \in \mathfrak{g}$. The left side gives, using the definition of ad twice,

$$[\text{ad}(a), \text{ad}(b)]c = \text{ad}(a)\text{ad}(b)c - \text{ad}(b)\text{ad}(a)c = [a, [b, c]] - [b, [a, c]].$$

The right side gives $\text{ad}([a, b])c = [[a, b], c]$. So the claimed identity is exactly

$$[a, [b, c]] - [b, [a, c]] - [[a, b], c] = 0,$$

which is precisely the Jacobi identity of Definition 24.1, rearranged (recall $[a, [b, c]] + [b, [c, a]] + [c, [a, b]] = 0$; using antisymmetry, $[b, [c, a]] = -[b, [a, c]]$ and $[c, [a, b]] = -[[a, b], c]$, giving exactly the identity above). $\square$

Remark. This is a satisfying converse to the observation, in Section 24, that the Jacobi identity descends from associativity of the group law: here we see that the Jacobi identity is, conversely, *exactly* the statement that “multiplication by elements of $\mathfrak{g}$ ” (i.e. ad) is itself a bona fide representation of $\mathfrak{g}$. Every Lie algebra therefore comes equipped, for free, with a representation on a space of the same dimension as the group itself — this is the representation physicists use, for example, to describe gauge bosons (which transform in the adjoint representation of the gauge group, e.g. the eight gluons for $SU(3)$ with $\dim \mathfrak{su}(3) = 3^2 - 1 = 8$).

Definition 29.2. The corresponding representation at the group level, the **adjoint representation** of G , is

$$\text{Ad}(A)a := AaA^{-1} \quad (A \in G, a \in \mathfrak{g}),$$

i.e. $\text{Ad}(A)$ is the linear map $\mathfrak{g} \rightarrow \mathfrak{g}$ given by conjugation by A (note that $AaA^{-1} \in \mathfrak{g}$ indeed, being the tangent vector at $t = 0$ to the curve $A \exp(ta)A^{-1} \in G$). One checks immediately that $\text{Ad}(A_1A_2) = \text{Ad}(A_1)\text{Ad}(A_2)$ (both sides send $a \mapsto A_1A_2aA_2^{-1}A_1^{-1}$), so Ad is indeed an n -dimensional representation of G .

Theorem 29.2. For every $a \in \mathfrak{g}$ and $t \in \mathbb{R}$,

$$\text{Ad}(\exp(ta)) = \exp(t \text{ad}(a)) \quad (\text{as linear operators on } \mathfrak{g}).$$

In particular, ad is exactly the Lie-algebra representation induced (in the sense of Section 27) by the group representation Ad .

Proof. Fix $a \in \mathfrak{g}$ and $b \in \mathfrak{g}$, and define $F(t) := \text{Ad}(\exp(ta))b = \exp(ta)b \exp(-ta) \in \mathfrak{g}$. Differentiating the matrix product using the ordinary product rule,

$$\dot{F}(t) = a \exp(ta)b \exp(-ta) - \exp(ta)b \exp(-ta)a = aF(t) - F(t)a = [a, F(t)] = \text{ad}(a)F(t).$$

So $F(t)$ solves the *linear*, constant-coefficient ODE $\dot{F} = \text{ad}(a)F$ with initial condition $F(0) = b$; the unique solution of such an ODE is $F(t) = \exp(t \text{ad}(a))b$ (exactly as for the ordinary scalar equation $\dot{f} = \lambda f$, $f(t) = e^{\lambda t}f(0)$, since the same power-series argument applies verbatim to the operator $\text{ad}(a)$ acting on the finite-dimensional space $\mathfrak{g}$). Comparing the two expressions for $F(t)$, and since $b \in \mathfrak{g}$ was arbitrary, $\text{Ad}(\exp(ta)) = \exp(t \text{ad}(a))$ as operators on $\mathfrak{g}$. Differentiating both sides at $t = 0$ recovers $\text{ad}(a)$ as the derivative of Ad along $\exp(ta)$, i.e. exactly the induced Lie-algebra representation of Ad , by definition. $\square$

Physical perspective

The identity $\text{Ad}(\exp(ta)) = \exp(t \text{ad}(a))$ is used constantly in practice: it is the statement that conjugating a group element by $\exp(ta)$, to all orders in t , resums into the Baker–Campbell–Hausdorff-type exponentiated commutator series $b \mapsto b + t[a, b] + \frac{t^2}{2}[a, [a, b]] + \cdots$ familiar from, e.g., the Heisenberg-picture time evolution of an operator under a similarity transformation, or from computing how a field transforms under a finite (rather than infinitesimal) gauge transformation once its infinitesimal transformation law (its representation of $\mathfrak{g}$) is known.

30 Simply Connected Groups and the Universal Covering Group

We saw in Examples 27.2 and 27.3 that a given Lie algebra can be the Lie algebra of several genuinely different (globally non-isomorphic) connected Lie groups, related to one another by homomorphisms with discrete kernel. We now classify *all* of them, for a fixed Lie algebra $\mathfrak{g}$, in terms of a single distinguished representative: the **universal covering group**. This requires a little topology.

30.1 Homotopy of Paths and Loops

Definition 30.1. Let G be a (connected) Lie group. A **path** in G is a continuous map $g : [0, 1] \rightarrow G$; it runs from $g(0)$ to $g(1)$. A path with $g(0) = g(1) = x_0$ (returning to its starting point) is called a **loop** based at x_0 .

Two paths $g, \tilde{g} : [0, 1] \rightarrow G$ with the *same* endpoints, $g(0) = \tilde{g}(0)$ and $g(1) = \tilde{g}(1)$, are called **homotopic** if there is a continuous family of paths $F(t, s)$, $t, s \in [0, 1]$, interpolating between them with the endpoints held fixed throughout the deformation:

$$F(t, 0) = g(t), \quad F(t, 1) = \tilde{g}(t), \quad F(0, s) = g(0), \quad F(1, s) = g(1) \quad \text{for all } s \in [0, 1].$$

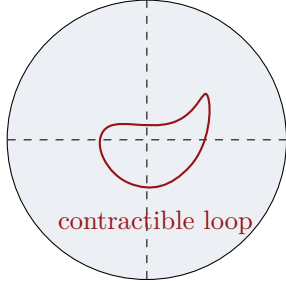
A loop is called **contractible** if it is homotopic to the constant loop $g(t) \equiv x_0$. G is called **simply connected** if G is connected and *every* loop in G is contractible.

Example 30.1. $SU(2)$ is simply connected. We saw in Example 26.1 that $SU(2)$ is, as a manifold, the unit sphere $S^3 \subset \mathbb{R}^4$. Topologically, *every* loop on S^n for $n \geq 2$ is contractible: intuitively, a loop is a one-dimensional object and so (generically) misses at least one point $p \in S^n$; but $S^n \setminus \{p\}$ is topologically an ordinary (contractible) n -dimensional space $\mathbb{R}^n$ (via stereographic projection), inside which the loop can be shrunk to a point without ever needing to pass through p . Hence $SU(2) \cong S^3$ is simply connected.

Example 30.2. $SO(3)$ is *not* simply connected. Every rotation is determined by an axis $\hat{n} \in S^2$ and an angle $\varphi \in [0, \pi]$ (angles beyond π are redundant, since rotating by $2\pi - \varphi$ about $-\hat{n}$ gives the same rotation as φ about $\hat{n}$); so $SO(3)$ can be pictured as the solid ball

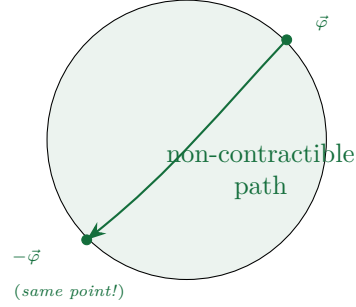
$B = \{\vec{\varphi} = \varphi \hat{n} : \varphi \in [0, \pi]\} \subset \mathbb{R}^3$ of radius π , **with each boundary point $\vec{\varphi}$ (at $\varphi = \pi$) identified with its antipode $-\vec{\varphi}$** — since a rotation by π about $\hat{n}$ equals the rotation by π about $-\hat{n}$. This identified-ball picture makes $SO(3)$ topologically the real projective space $\mathbb{RP}^3$, which is *not* simply connected, as Fig. 13 below illustrates.

$SO(3)$ as a ball of radius π



(stays in the interior)

identify $\vec{\varphi} \sim -\vec{\varphi}$ on ∂B



(endpoints are the same group element)

Figure 13: $SO(3)$ pictured as a solid ball of radius π with antipodal boundary points identified. A loop entirely in the interior (left) can always be shrunk to a point. A path running from one boundary point $\vec{\varphi}$ to its antipode $-\vec{\varphi}$ (right) is, because of the identification, already a *loop* (it starts and ends at the same group element), but it cannot be contracted: any attempt to shrink it must eventually break through the identified boundary sphere. Traversing such a path *twice* (i.e. a “ 4π rotation”), however, *is* contractible — consistent with $\pi_1(SO(3)) \cong \mathbb{Z}_2$.

A straight path from a boundary point $\vec{\varphi}$ to its identified antipode $-\vec{\varphi}$ (Fig. 13, right) is a genuine loop (it begins and ends at the very same rotation), yet it cannot be continuously shrunk to the constant loop without at some stage crossing the boundary sphere discontinuously. One can show that traversing such a loop *twice* is contractible, so that $\pi_1(SO(3)) \cong \mathbb{Z}_2$ (only two homotopy classes of loops exist). This factor of two is not a mathematical curiosity: it is the precise topological origin of the need for $SU(2)$, the double cover of $SO(3)$, in the quantum mechanics of spin.

30.2 The Universal Covering Group

Theorem 30.1. For every (finite-dimensional, real) Lie algebra $\mathfrak{g}$ there exists a **unique** (up to isomorphism) **simply connected** Lie group $\tilde{G}$ with Lie algebra $\mathfrak{g}$, called the **universal covering group** of any connected Lie group with Lie algebra $\mathfrak{g}$. Every other connected Lie group G with the same Lie algebra $\mathfrak{g}$ is isomorphic to a quotient

$$G \cong \tilde{G}/K$$

for some **discrete** subgroup K of the **centre** $Z(\tilde{G}) = \{z \in \tilde{G} : zg = gz \ \forall g \in \tilde{G}\}$ of $\tilde{G}$ (discreteness of K ensures, by Theorem 27.3, that the quotient map $\tilde{G} \rightarrow \tilde{G}/K$ induces an isomorphism on Lie algebras, as it must if G and $\tilde{G}$ share the Lie algebra $\mathfrak{g}$; and K must lie in the centre precisely so that $\tilde{G}/K$ is a *group* at all with K as the kernel of a homomorphism from $\tilde{G}$, which forces K to be normal, and a discrete normal subgroup of a connected group is automatically central).

Remark. This theorem “bails down” (reduces) the classification of *connected Lie groups* to two much more tractable problems: classifying (finite-dimensional) *Lie algebras* — a purely algebraic problem, solved completely for semisimple Lie algebras by the Cartan classification — and, for each such Lie algebra, classifying the (finite) discrete subgroups of the centre of its universal covering group, an easy finite-group-theory problem once $Z(\tilde{G})$ is known.

Example 30.3. $SU(2)$ is the universal covering group of $SO(3)$. We showed $\mathfrak{su}(2) \cong \mathfrak{so}(3)$ (Example 27.3) and $SU(2)$ is simply connected (Example 30.1), so $SU(2) = \widetilde{SO(3)}$. The centre of $SU(2)$ is $Z(SU(2)) = \{1, -1\} \cong \mathbb{Z}_2$ (a matrix commuting with every $SU(2)$ matrix must, by Schur’s Lemma applied to the irreducible defining representation, be a multiple of 1 , and unitarity with unit determinant forces it to be ± 1). The only discrete subgroups of $\mathbb{Z}_2$ are $\{1\}$ and all of $\mathbb{Z}_2$, giving exactly two connected Lie groups with Lie algebra $\mathfrak{su}(2) \cong \mathfrak{so}(3)$:

$$\tilde{G}/\{1\} = SU(2) \quad \text{and} \quad \tilde{G}/\mathbb{Z}_2 = SU(2)/\{\pm 1\} \cong SO(3),$$

matching the double cover found directly in Example 27.3. (For $SU(N)$, $N \geq 2$, the centre is the larger cyclic group $\mathbb{Z}_N = \{e^{2\pi i k/N} 1 : k = 0, \dots, N-1\}$, giving correspondingly more connected groups sharing $\mathfrak{su}(N)$, such as $PSU(N) := SU(N)/\mathbb{Z}_N$.)

Physical perspective

Projective representations and the spin–statistics of representation theory. Since $\tilde{G}$ is simply connected, Theorem 30.1 together with the group–algebra correspondence guarantees that *every* representation of $\mathfrak{g}$ exponentiates to an honest representation of $\tilde{G}$. It does **not**, however, always descend to an honest representation of a non-simply-connected $G = \tilde{G}/K$: a representation D of $\mathfrak{g}$ descends to G if and only if the corresponding representation of $\tilde{G}$ sends every element of K to the identity operator.

The paradigm example is the **spin- $\frac{1}{2}$ representation** of $\mathfrak{su}(2) \cong \mathfrak{so}(3)$ (constructed explicitly in Section 31): it exponentiates to an honest, two-dimensional representation of $SU(2)$, but the non-trivial central element $-1 \in Z(SU(2)) = K$ is represented by $-\mathbb{1}_{2 \times 2} \neq \mathbb{1}_{2 \times 2}$, *not* the identity. So the spin- $\frac{1}{2}$ representation does *not* descend to an honest representation of $SO(3) = SU(2)/\mathbb{Z}_2$: instead it defines only a **projective representation** of $SO(3)$ (a representation up to an overall sign/phase, $D(R_1)D(R_2) = \pm D(R_1 R_2)$), which is exactly the familiar statement that a spinor picks up a factor of -1 under a rotation by 2π (an element of $SO(3)$ that is *not* the identity element of $SU(2)$ ’s preferred lift, even though it *is* the identity of $SO(3)$). By contrast, integer-spin representations (spin $0, 1, 2, \dots$, built in Section 31) *do* represent $-1 \in SU(2)$ trivially, and so descend to honest representations of $SO(3)$ — consistent with the fact that only integer-spin fields (not half-integer spin) can be built directly out of tensors (rather than spinors) of $SO(3)$.

31 Classification of the Irreducible Representations of $\mathfrak{su}(2) \cong \mathfrak{so}(3)$

We now put all of the machinery developed so far — the Lie bracket, the adjoint/Casimir idea, Schur’s Lemma from Part B — to work on the single most important example in all of physics: the classification of every finite-dimensional representation of $\mathfrak{su}(2)$.

31.1 Ladder Operators and the Casimir Invariant

Let D be a finite-dimensional representation of $\mathfrak{su}(2)$ on a complex vector space V , and write $t_a := D(t_a)$ (a slight abuse of notation) for the images of the Hermitian generators satisfying $[t_a, t_b] = i\epsilon_{abc}t_c$ (Section 25.4). We may take complex linear combinations of the t_a (working in the complexification $\mathfrak{su}(2)_{\mathbb{C}} \cong \mathfrak{sl}(2, \mathbb{C})$ mentioned in Section 24); the most useful combinations are the **ladder operators**

$$t_{\pm} := t_1 \pm it_2.$$

Theorem 31.1. The ladder operators satisfy

$$[t_3, t_{\pm}] = \pm t_{\pm}, \quad [t_+, t_-] = 2t_3, \quad t_{\pm}^{\dagger} = t_{\mp} \quad (\text{if } D \text{ is unitary}).$$

Moreover, the **Casimir operator**

$$t^2 := t_1^2 + t_2^2 + t_3^2 = t_3^2 + \frac{1}{2}(t_+t_- + t_-t_+)$$

commutes with *every* element of $\mathfrak{su}(2)$: $[t^2, t_a] = 0$ for $a = 1, 2, 3$.

Proof. Using $[t_1, t_2] = it_3$, $[t_2, t_3] = it_1$, $[t_3, t_1] = it_2$ (the antisymmetrised structure constants $i\epsilon_{abc}$ from Section 25.4):

$$[t_3, t_{\pm}] = [t_3, t_1] \pm i[t_3, t_2] = it_2 \pm i(-it_1) = it_2 \pm t_1 = \pm(t_1 \pm it_2) = \pm t_{\pm},$$

$$[t_+, t_-] = [t_1 + it_2, t_1 - it_2] = -i[t_1, t_2] + i[t_2, t_1] = -i(it_3) + i(-it_3) = t_3 + t_3 = 2t_3.$$

For the Casimir, using $[t_a, t_a^2] = 0$ trivially is not enough since t^2 mixes different a ; instead compute directly, e.g.

$$\begin{aligned} [t^2, t_3] &= [t_1^2, t_3] + [t_2^2, t_3] = t_1[t_1, t_3] + [t_1, t_3]t_1 + t_2[t_2, t_3] + [t_2, t_3]t_2 \\ &= t_1(-it_2) + (-it_2)t_1 + t_2(it_1) + (it_1)t_2 = 0, \end{aligned}$$

using $[A, B^2] = B[A, B] + [A, B]B$ (Leibniz rule for the commutator) and $[t_1, t_3] = -it_2$, $[t_2, t_3] = it_1$; the terms cancel in pairs. The identical computation (or a cyclic relabelling $1 \rightarrow 2 \rightarrow 3 \rightarrow 1$) gives $[t^2, t_1] = [t^2, t_2] = 0$. Finally, $t^2 = t_1^2 + t_2^2 + t_3^2$ and $t_+t_- + t_-t_+ = (t_1 + it_2)(t_1 - it_2) + (t_1 - it_2)(t_1 + it_2) = 2(t_1^2 + t_2^2)$, giving the stated rewriting $t^2 = t_3^2 + \frac{1}{2}(t_+t_- + t_-t_+)$. $\square$

Remark. Because $[t^2, t_a] = 0$ for all a , Schur's Lemma (Part B) applies: if D is *irreducible*, t^2 must act as a scalar multiple $\lambda \mathbb{1}_V$ of the identity on the whole representation space. This single number λ (to be identified with $j(j+1)$ below) will label the irreducible representations, exactly as the quantum number ℓ (or j) labels angular-momentum multiplets in quantum mechanics — indeed t^2 *is*, physically, the total angular momentum (or spin) operator squared.

A direct computation, using $[t_+, t_-] = 2t_3$, gives two operator identities that will do all the remaining work:

$$t_+t_- = t^2 - t_3^2 + t_3, \quad t_-t_+ = t^2 - t_3^2 - t_3. \quad (6)$$

(From Theorem 31.1, $t_+t_- = \frac{1}{2}(t_+t_- + t_-t_+) + \frac{1}{2}[t_+, t_-] = (t^2 - t_3^2) + t_3$, and similarly with the sign of the commutator reversed for t_-t_+ .)

31.2 The Highest-Weight Construction

Fix an irreducible representation D with Casimir eigenvalue λ . Since t_3 is Hermitian (if D is unitary; we assume this, as we may by Part B's unitarity theorem) it can be diagonalised, with real eigenvalues m (its **weights**) and orthonormal eigenvectors ψ_m , $t_3\psi_m = m\psi_m$.

Theorem 31.2. t_+ raises the weight by 1 and t_- lowers it by 1: if $t_3\psi_m = m\psi_m$, then $t_3(t_\pm\psi_m) = (m \pm 1)(t_\pm\psi_m)$, i.e. $t_\pm\psi_m \propto \psi_{m\pm 1}$ (or is zero).

Proof. $t_3(t_\pm\psi_m) = ([t_3, t_\pm] + t_\pm t_3)\psi_m = (\pm t_\pm + t_\pm m)\psi_m = (m \pm 1)t_\pm\psi_m$, using Theorem 31.1. $\square$

Theorem 31.3. (Classification theorem.) The finite-dimensional irreducible representations of $\mathfrak{su}(2)$ are labelled by a single number

$$j \in \left\{0, \frac{1}{2}, 1, \frac{3}{2}, 2, \dots\right\} \quad (\text{i.e. } 2j \in \mathbb{Z}_{\geq 0}),$$

called the **spin** (or angular momentum). The representation D^j has dimension $2j + 1$, an orthonormal basis $\{\psi_m^j\}_{m=-j, -j+1, \dots, j}$ of t_3 -eigenstates, Casimir eigenvalue $\lambda = j(j + 1)$, and

$$\begin{aligned} t_3 \psi_m^j &= m \psi_m^j, \\ t_+ \psi_m^j &= \sqrt{(j - m)(j + m + 1)} \psi_{m+1}^j, \\ t_- \psi_m^j &= \sqrt{(j + m)(j - m + 1)} \psi_{m-1}^j, \end{aligned}$$

(with $t_+\psi_j^j = 0$ and $t_-\psi_{-j}^j = 0$ understood, the square roots vanishing there).

Proof. Since D is finite-dimensional, t_3 has only finitely many eigenvalues (weights); by Theorem 31.2 they are spaced by integers along each “ladder” generated by repeated application of t_+ or t_- to a given eigenvector, and since there are only finitely many, each such ladder must *terminate* at both ends. Let j denote the largest weight occurring (the **highest weight**) and $\psi_j := \psi_j^j$ a corresponding (normalised) eigenvector; termination at the top means

$$t_+ \psi_j = 0.$$

Apply $t_- t_+$ (Eq. (6)) to ψ_j : since $t^2 \psi_j = \lambda \psi_j$ (Schur, as D is irreducible) and $t_3 \psi_j = j \psi_j$,

$$0 = t_- t_+ \psi_j = (t^2 - t_3^2 - t_3) \psi_j = (\lambda - j^2 - j) \psi_j \implies \lambda = j(j + 1).$$

Now build $\psi_{j-1}, \psi_{j-2}, \dots$ by repeated lowering, $\psi_{m-1} \propto t_- \psi_m$; Theorem 31.2 guarantees each is a t_3 -eigenvector with the stated eigenvalue, and irreducibility guarantees that repeated lowering from ψ_j eventually reaches *every* vector in V (the subspace generated this way is invariant under $t_3, t_\pm$, hence under all of $\mathfrak{su}(2)_\mathbb{C}$, hence under $\mathfrak{su}(2)$ itself, so by irreducibility it must be all of V). By finite-dimensionality, this downward ladder must also terminate: there is a smallest weight $m_{\min} = j - N$ (for some non-negative integer N , since the weights are equally spaced by Theorem 31.2) with $t_- \psi_{m_{\min}} = 0$. Applying $t_+ t_-$ (Eq. (6)) to $\psi_{m_{\min}}$:

$$0 = t_+ t_- \psi_{m_{\min}} = (t^2 - t_3^2 + t_3) \psi_{m_{\min}} = (\lambda - m_{\min}^2 + m_{\min}) \psi_{m_{\min}} \implies \lambda = m_{\min}(m_{\min} + 1).$$

Combining this with $\lambda = j(j + 1)$ and $m_{\min} = j - N$, and expanding both sides of $j(j + 1) = (j - N)(j - N + 1)$ fully:

$$\begin{aligned} \underbrace{j^2 + j}_{\text{left}} &= \underbrace{(j - N)^2 - (j - N)}_{\text{right}} = j^2 - 2jN + N^2 - j + N \\ \implies 2j + 2jN &= N^2 + N \implies 2j(1 + N) = N(N + 1), \end{aligned}$$

and since $N + 1 \neq 0$, this gives $2j = N$, i.e. $j = N/2$. So j is a non-negative half-integer, $m_{\min} = j - N = j - 2j = -j$, and the weights run over $m = j, j - 1, \dots, -j$ in $2j + 1$ integer steps, giving $\dim D^j = 2j + 1$.

Finally, the explicit coefficients in the ladder action follow from computing the (positive, since D is unitary) norm $\|t_- \psi_m^j\|^2 = \langle \psi_m^j | t_- | \psi_m^j \rangle = \lambda - m^2 + m = j(j+1) - m(m-1) = (j-m+1)(j+m)$ (using Eq. (6) and $\lambda = j(j+1)$), and taking the positive square root as the (real, by convention) proportionality constant defining $\psi_{m-1}^j := t_- \psi_m^j / \sqrt{(j-m+1)(j+m)}$; the coefficient for t_+ follows identically from $\|t_+ \psi_m^j\|^2 = \langle \psi_m^j | t_+ | \psi_m^j \rangle = \lambda - m^2 - m = (j-m)(j+m+1)$. $\square$

Remark. This is exactly the familiar quantum-mechanical construction of angular-momentum multiplets — $j = 0$ (dimension 1, trivial representation), $j = \frac{1}{2}$ (dimension 2, the defining/spinor representation of $SU(2)$, with $t_a = \frac{1}{2}\sigma_a$ as in Section 25.4), $j = 1$ (dimension 3, the vector/adjoint representation, isomorphic to the defining representation of $SO(3)$), $j = \frac{3}{2}$ (dimension 4), and so on — now derived from nothing but the abstract commutation relations $[t_a, t_b] = i\epsilon_{abc}t_c$, with no reference to any particular physical system. As flagged in Section 30: the integer- j representations ($j = 0, 1, 2, \dots$) are honest representations of $SO(3)$, while the half-integer- j ones ($j = \frac{1}{2}, \frac{3}{2}, \dots$) are honest representations of $SU(2)$ but only *projective* representations of $SO(3)$.

32 The Clebsch–Gordan Decomposition

Given two irreducible representations D^{j_1}, D^{j_2} of $\mathfrak{su}(2)$, their tensor product $D^{j_1} \otimes D^{j_2}$ (Definition 28.4 and Theorem 28.3) is, in general, *reducible*: it decomposes into a direct sum of irreducibles D^j . Finding exactly which j ’s occur, and how the corresponding basis vectors are built out of $\psi_{m_1}^{j_1} \otimes \psi_{m_2}^{j_2}$, is the **Clebsch–Gordan problem** — ubiquitous in physics under the name “addition of angular momentum”.

32.1 Counting States: Which j Occur?

Write $T_3 := t_3^{(1)} \otimes \mathbb{1} + \mathbb{1} \otimes t_3^{(2)}$ for the total-spin generator on $V = V^{j_1} \otimes V^{j_2}$ (Theorem 28.3). The product basis vectors $\psi_{m_1}^{j_1} \otimes \psi_{m_2}^{j_2}$ are simultaneous eigenvectors of T_3 with eigenvalue $m = m_1 + m_2$, so the multiplicity of weight m in V ,

$$N(m) := \#\{(m_1, m_2) : m_1 \in \{-j_1, \dots, j_1\}, m_2 \in \{-j_2, \dots, j_2\}, m_1 + m_2 = m\},$$

is simply the number of lattice points on the diagonal line $m_1 + m_2 = m$ inside the rectangle $[-j_1, j_1] \times [-j_2, j_2]$.

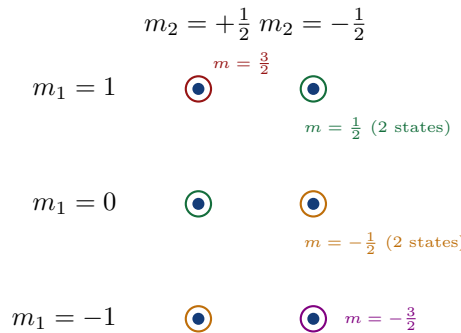


Figure 14: Weight-counting diagram for $D^1 \otimes D^{1/2}$ (dimension $3 \times 2 = 6$). Each dot is a product state $\psi_{m_1}^1 \otimes \psi_{m_2}^{1/2}$; states are grouped by total weight $m = m_1 + m_2$. The unique $m = \frac{3}{2}$ state is the highest weight of a $D^{3/2}$ multiplet (dimension 4, using up one state from each of the $m = \frac{3}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{3}{2}$ levels); the *remaining* state at $m = \frac{1}{2}$ is then the highest weight of a second multiplet, $D^{1/2}$ (dimension 2). Total: $4 + 2 = 6$, matching $D^1 \otimes D^{1/2} = D^{3/2} \oplus D^{1/2}$.

Theorem 32.1. For $j_1 \geq j_2$ (without loss of generality),

$$D^{j_1} \otimes D^{j_2} \cong D^{j_1+j_2} \oplus D^{j_1+j_2-1} \oplus \dots \oplus D^{j_1-j_2},$$

each irreducible D^j ($j = |j_1 - j_2|, \dots, j_1 + j_2$, in integer steps) occurring with multiplicity exactly one.

Proof. The top weight $m = j_1 + j_2$ occurs with multiplicity $N(j_1 + j_2) = 1$ (only $m_1 = j_1, m_2 = j_2$ achieves it): this state, $\psi_{j_1}^{j_1} \otimes \psi_{j_2}^{j_2}$, satisfies $T_+(\cdot) = 0$ (both factors are already annihilated by their own t_+ , and $T_+ = t_+^{(1)} \otimes \mathbb{1} + \mathbb{1} \otimes t_+^{(2)}$ by Theorem 28.3), so it is a highest-weight state and generates (by repeated lowering with T_- , exactly as in Theorem 31.3) a full copy of $D^{j_1+j_2}$, using up exactly one state at every weight $m = j_1 + j_2, j_1 + j_2 - 1, \dots, -(j_1 + j_2)$.

At the next weight down, $m = j_1 + j_2 - 1$, one has $N(j_1 + j_2 - 1) = 2$ (from $(m_1, m_2) = (j_1 - 1, j_2)$ and $(j_1, j_2 - 1)$, which are always distinct, since they differ in m_1 , and always both within range provided $j_1, j_2 \geq \frac{1}{2}$; if instead, say, $j_2 = 0$, then $N \equiv 1$ throughout and the construction below simply terminates immediately with the single multiplet D^{j_1}), of which one linear combination is already used by the $D^{j_1+j_2}$ multiplet just constructed (obtained by lowering its highest-weight state); the *orthogonal* combination within this two-dimensional weight space is annihilated by T_+ (being orthogonal to the image of T_- acting on the $m = j_1 + j_2$ state, and T_+, T_- are mutually adjoint for a unitary representation, so T_+ maps this orthogonal complement into the now-exhausted $m = j_1 + j_2$ level's orthogonal complement, i.e. to zero) — hence it is itself a new highest-weight state, generating a full copy of $D^{j_1+j_2-1}$.

Iterating this argument (at each successive weight m , one new highest-weight state is liberated, precisely accounting for the increase $N(m) - N(m + 1)$ in multiplicity, as long as N is still increasing) continues until m reaches $j_1 - j_2$: for $|m| \leq j_1 - j_2$, every one of the $2j_2 + 1$ values of $m_2 \in \{-j_2, \dots, j_2\}$ admits a valid $m_1 = m - m_2 \in \{-j_1, \dots, j_1\}$, so $N(m)$ has reached, and stays at, its maximum plateau value $2j_2 + 1$ throughout this range (it no longer *increases*, so no further new highest-weight states are liberated once $m < j_1 - j_2$); below $m = -(j_1 - j_2)$, $N(m)$ decreases again by exactly the same amount it increased, symmetric under $m \rightarrow -m$, and the construction is complete after exactly $2j_2 + 1$ multiplets have been extracted, at $j = j_1 + j_2, j_1 + j_2 - 1, \dots, j_1 - j_2$. $\square$

Remark. Dimension check. The claimed decomposition can be verified independently by summing dimensions: with $a := j_1 + j_2$, $b := j_1 - j_2 \geq 0$,

$$\sum_{j=b}^a (2j+1) = (a+1)^2 - b^2 = (a+b+1)(a-b+1) = (2j_1+1)(2j_2+1),$$

using $\sum_{j=0}^n (2j+1) = (n+1)^2$ (sum of the first $n+1$ odd numbers) applied to $\sum_{j=b}^a = \sum_{j=0}^a - \sum_{j=0}^{b-1}$, and factoring the difference of squares; this matches $\dim(D^{j_1} \otimes D^{j_2}) = (2j_1+1)(2j_2+1)$ exactly, confirming Theorem 32.1.

32.2 Clebsch–Gordan Coefficients

Definition 32.1. The (real, by convention — see below) expansion coefficients relating the two orthonormal bases of $V^{j_1} \otimes V^{j_2}$,

$$\psi_m^{(j)} = \sum_{m_1, m_2} \langle j_1 j_2; m_1 m_2 | j m \rangle \psi_{m_1}^{j_1} \otimes \psi_{m_2}^{j_2} \quad (m_1 + m_2 = m \text{ enforced}),$$

are called **Clebsch–Gordan coefficients**. They vanish unless $m = m_1 + m_2$ and unless $|j_1 - j_2| \leq j \leq j_1 + j_2$ (Theorem 32.1).

The proof of Theorem 32.1 is, in fact, a *recipe* for computing them: at each new top weight $m = j_1 + j_2 - k$ ($k = 0, 1, \dots, 2j_2$), find the state orthogonal to all previously constructed multiplets within the (finite-dimensional) weight- m subspace, fix its overall phase (**Condon–Shortley convention**: the coefficient of the term with the largest value of m_1 is chosen real and positive), and then generate the rest of that D^j multiplet by repeated application of $T_- = t_-^{(1)} \otimes \mathbb{1} + \mathbb{1} \otimes t_-^{(2)}$, using the known coefficients of Theorem 31.3 to keep track of normalisation. This procedure determines every Clebsch–Gordan coefficient uniquely (given the phase convention) and, remarkably, always produces *real* numbers.

Example 32.1. Two spin- $\frac{1}{2}$'s: $D^{1/2} \otimes D^{1/2} = D^1 \oplus D^0$. Write $|\uparrow\rangle := \psi_{1/2}^{1/2}$, $|\downarrow\rangle := \psi_{-1/2}^{1/2}$ for the two basis states of each spin- $\frac{1}{2}$ factor. By Theorem 32.1 ($j_1 = j_2 = \frac{1}{2}$), $D^{1/2} \otimes D^{1/2} = D^1 \oplus D^0$ (dimension check: $2 \times 2 = 4 = 3 + 1$). Following the recipe:

Top state ($m = 1$, unique): $\psi_1^{(1)} = |\uparrow\uparrow\rangle$ (shorthand for $|\uparrow\rangle \otimes |\uparrow\rangle$), a highest weight state for D^1 ($T_+|\uparrow\uparrow\rangle = 0$ since $t_+|\uparrow\rangle = 0$ in each factor).

Lower it using $T_- = t_-^{(1)} \otimes \mathbb{1} + \mathbb{1} \otimes t_-^{(2)}$ and Theorem 31.3 ($t_- \psi_{1/2}^{1/2} = \psi_{-1/2}^{1/2}$, coefficient $\sqrt{(\frac{1}{2} + \frac{1}{2})(\frac{1}{2} - \frac{1}{2} + 1)} = 1$):

$$T_- \psi_1^{(1)} = |\downarrow\uparrow\rangle + |\uparrow\downarrow\rangle, \quad \text{normalised: } \psi_0^{(1)} = \frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle).$$

(The normalisation follows the general coefficient $t_- \psi_m^j \rightarrow \sqrt{(j+m)(j-m+1)} \psi_{m-1}^j$ of Theorem 31.3 with $j = 1, m = 1$: $\sqrt{(1+1)(1-1+1)} = \sqrt{2}$.)

Lower once more ($j = 1, m = 0$: coefficient $\sqrt{(1+0)(1-0+1)} = \sqrt{2}$). Since $t_-|\downarrow\rangle = 0$ already (bottom of each individual spin- $\frac{1}{2}$ ladder), only one term survives in each product:

$$T_-|\uparrow\downarrow\rangle = (t_-|\uparrow\rangle) \otimes |\downarrow\rangle + |\uparrow\rangle \otimes (t_-|\downarrow\rangle) = |\downarrow\downarrow\rangle + 0, \quad T_-|\downarrow\uparrow\rangle = 0 + |\downarrow\downarrow\rangle,$$

so

$$T_- \psi_0^{(1)} = \frac{1}{\sqrt{2}}(T_-|\uparrow\downarrow\rangle + T_-|\downarrow\uparrow\rangle) = \frac{1}{\sqrt{2}}(2|\downarrow\downarrow\rangle) = \sqrt{2}|\downarrow\downarrow\rangle,$$

matching the predicted coefficient $\sqrt{2}$, and giving $\psi_{-1}^{(1)} = |\downarrow\downarrow\rangle$: this completes the **triplet** D^1 :

$$\psi_1^{(1)} = |\uparrow\uparrow\rangle, \quad \psi_0^{(1)} = \frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle), \quad \psi_{-1}^{(1)} = |\downarrow\downarrow\rangle.$$

The singlet D^0 : at $m = 0$ there are two product states, $|\uparrow\downarrow\rangle, |\downarrow\uparrow\rangle$, of which one combination, $\psi_0^{(1)}$ above, is used by the triplet; the orthogonal combination,

$$\psi_0^{(0)} = \frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle),$$

(sign and normalisation fixed by orthonormality to $\psi_0^{(1)}$, and the Condon–Shortley convention that the $m_1 = \frac{1}{2}$ term — i.e. $|\uparrow\downarrow\rangle$ — carries the + sign) must satisfy $T_+ \psi_0^{(0)} = 0$ and $T_- \psi_0^{(0)} = 0$ automatically, since it is alone at its weight and D^0 is one-dimensional; one checks this directly: $T_+ \psi_0^{(0)} = \frac{1}{\sqrt{2}}(t_+|\uparrow\rangle \otimes |\downarrow\rangle + |\uparrow\rangle \otimes t_+|\downarrow\rangle - t_+|\downarrow\rangle \otimes |\uparrow\rangle - |\downarrow\rangle \otimes t_+|\uparrow\rangle) = \frac{1}{\sqrt{2}}(0 + |\uparrow\uparrow\rangle - |\uparrow\uparrow\rangle - 0) = 0$, as required.

The non-zero Clebsch–Gordan coefficients $\langle \frac{1}{2} \frac{1}{2}; m_1 m_2 | j m \rangle$ are summarised below:

m_1	m_2	m	$j = 1$	$j = 0$
$\frac{1}{2}$	$\frac{1}{2}$	1	1	—
$\frac{1}{2}$	$-\frac{1}{2}$	0	$\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{2}}$
$-\frac{1}{2}$	$\frac{1}{2}$	0	$\frac{1}{\sqrt{2}}$	$-\frac{1}{\sqrt{2}}$
$-\frac{1}{2}$	$-\frac{1}{2}$	-1	1	—

Physical perspective

This is exactly the familiar quantum-mechanical rule for combining two spin- $\frac{1}{2}$ particles: the symmetric **triplet** $D^1 = \{|\uparrow\uparrow\rangle, \frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle), |\downarrow\downarrow\rangle\}$ and the antisymmetric **singlet** $D^0 = \frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle)$, ubiquitous from the hydrogen-atom hyperfine splitting to entangled two-qubit (Bell) states — the singlet state is, in fact, the maximally entangled state referenced in the remark following Definition 28.4.

33 Application: Spherical Harmonics as Representation Functions of $SO(3)$

As a worked application connecting the abstract classification of Section 31 back to very concrete objects, we identify the integer-spin representations D^ℓ ($\ell = 0, 1, 2, \dots$) inside the familiar space of functions on the sphere.

$SO(3)$ acts on functions $f : S^2 \rightarrow \mathbb{C}$ by $(\Gamma(R)f)(\hat{x}) := f(R^{-1}\hat{x})$ ($R \in SO(3)$, $\hat{x} \in S^2$); this is a (generally infinite-dimensional) representation, since $\Gamma(R_1 R_2)f(\hat{x}) = f((R_1 R_2)^{-1}\hat{x}) = f(R_2^{-1} R_1^{-1}\hat{x}) = \Gamma(R_2)\Gamma(R_1)f(\hat{x}) = (\Gamma(R_1)\Gamma(R_2)f)(\hat{x})$. Differentiating along $R = \exp(t\hat{e}_a)$ (rotation about axis a) at $t = 0$, exactly as in Section 27, gives the induced Lie-algebra representation — the familiar **orbital angular momentum operators**

$$L_a := -i\epsilon_{abc}x_b \frac{\partial}{\partial x_c} \quad (a = 1, 2, 3),$$

acting on functions of $\vec{x} \in \mathbb{R}^3$ (restricted to S^2 , i.e. $|\vec{x}| = 1$). Writing $L_a = \epsilon_{abc}x_b p_c$ with $p_c = -i\partial_c$ and using the canonical commutator $[x_a, p_b] = i\delta_{ab}$, a direct (if slightly lengthy) computation — entirely standard, and identical in structure to the Pauli-matrix computation of Section 25.4 — confirms that the L_a satisfy exactly the $\mathfrak{su}(2) \cong \mathfrak{so}(3)$ commutation relations of Section 25.4:

$$[L_a, L_b] = i\epsilon_{abc}L_c.$$

In spherical polar coordinates (θ, φ) on S^2 (related to $\vec{x}$ by $x_1 = \sin\theta \cos\varphi$, $x_2 = \sin\theta \sin\varphi$, $x_3 = \cos\theta$), the chain rule turns these into the familiar differential operators

$$\begin{aligned} L_3 &= -i \frac{\partial}{\partial \varphi}, \\ L_\pm &:= L_1 \pm iL_2 = e^{\pm i\varphi} \left(\pm \frac{\partial}{\partial \theta} + i \cot \theta \frac{\partial}{\partial \varphi} \right), \\ L^2 &= -\frac{1}{\sin\theta} \frac{\partial}{\partial \theta} \left(\sin\theta \frac{\partial}{\partial \theta} \right) - \frac{1}{\sin^2\theta} \frac{\partial^2}{\partial \varphi^2}, \end{aligned}$$

matching precisely the abstract $t_3, t_\pm, t^2$ of Section 31 (with $L_3 \leftrightarrow t_3$, $L_\pm \leftrightarrow t_\pm$, $L^2 \leftrightarrow t^2$). The **spherical harmonics** $Y_\ell^m(\theta, \varphi)$ are, by definition, the joint eigenfunctions

$$L^2 Y_\ell^m = \ell(\ell+1) Y_\ell^m, \quad L_3 Y_\ell^m = m Y_\ell^m,$$

so that, for each fixed ℓ , the $(2\ell + 1)$ functions $\{Y_\ell^m\}_{m=-\ell}^\ell$ furnish an explicit realisation of the abstract representation D^ℓ of Theorem 31.3, with $L_\pm Y_\ell^m = \sqrt{(\ell \mp m)(\ell \pm m + 1)} Y_\ell^{m\pm 1}$ exactly as in that theorem.

Remark. Why only integer ℓ occurs here. Single-valuedness of a genuine function on the sphere requires $Y_\ell^m(\theta, \varphi + 2\pi) = Y_\ell^m(\theta, \varphi)$; since $L_3 = -i\partial_\varphi$ has eigenfunctions $\propto e^{im\varphi}$, this forces $m \in \mathbb{Z}$, and hence (by Theorem 31.3) $\ell \in \mathbb{Z}_{\geq 0}$ as well. This is a concrete instance of the general fact, established in Section 30: functions on $S^2 \cong SO(3)/SO(2)$ (the sphere is literally the coset space of rotations modulo those fixing the north pole) can only realise *honest* representations of $SO(3)$ itself, and Section 30 showed these are precisely the integer-spin ones. The half-integer-spin representations of $\mathfrak{su}(2)$, by contrast, are realised not on functions of a point on the sphere, but on two-component *spinor* fields, which pick up the projective phase -1 discussed there.

The explicit closed-form solution of these eigenvalue equations (solving the eigenvalue problem for L^2 by separation of variables, the polar part reducing to the associated Legendre equation) gives the standard normalised formula

$$Y_\ell^m(\theta, \varphi) = (-1)^m \sqrt{\frac{2\ell + 1}{4\pi} \frac{(\ell - m)!}{(\ell + m)!}} P_\ell^m(\cos \theta) e^{im\varphi} \quad (-\ell \leq m \leq \ell),$$

where P_ℓ^m are the associated Legendre functions (in the convention already incorporating the Condon–Shortley phase into P_ℓ^m itself, the explicit $(-1)^m$ above is redundant with it; we follow the common convention, consistent with the ladder-operator phases already fixed above, in which $(-1)^m$ is displayed explicitly and P_ℓ^m is defined without it — consult whichever source’s P_ℓ^m you use and include the phase exactly once); we do not re-derive this classical formula here (it belongs properly to the analysis of the Legendre differential equation), but note that its *existence and structure* — one function for each of the $2\ell + 1$ values of m , related by the ladder operators $L_\pm$ exactly as prescribed by representation theory — is now fully explained, and indeed *predicted in advance*, by Theorem 31.3 alone, without solving any differential equation.

Physical perspective

This is the representation-theoretic underpinning of the multipole expansion in electrostatics, of the angular part of the hydrogen-atom wavefunction, and of the addition of orbital angular momenta via the very same Clebsch–Gordan coefficients of Section 32 (e.g. the product $Y_{\ell_1}^{m_1} Y_{\ell_2}^{m_2}$ of two spherical harmonics — itself a function on the sphere, with no radial dependence at all — decomposes into a sum of Y_L^M ’s whose coefficients are exactly products of Clebsch–Gordan coefficients (one for the m ’s, $\langle \ell_1 \ell_2; m_1 m_2 | LM \rangle$, and one fixing the overall normalisation via $\langle \ell_1 \ell_2; 00 | L0 \rangle$) as in Example 32.1; this purely angular integral $\int Y_L^M Y_{\ell_1}^{m_1} Y_{\ell_2}^{m_2} d\Omega$ is precisely what is known as the Gaunt coefficient). It is a striking illustration of the power of the present chapter that the entire qualitative structure of spherical harmonics — their number at each ℓ , their transformation under rotations, their recursion relations — follows from three lines of Lie algebra ($[t_a, t_b] = i\epsilon_{abc} t_c$), with no reference to the Legendre differential equation at all.

Part IV

Structure Theory and Classification of Semisimple Lie Algebras

34 Introduction: From a Single Algebra to a Classification

In Part C we learned how to pass from a Lie group to its Lie algebra, how representations of the algebra relate to representations of the group, and we solved, by hand, the representation theory of one particular algebra: $\mathfrak{su}(2) \cong \mathfrak{so}(3)$. That computation — ladder operators, a Casimir invariant, a highest weight terminating a finite tower of states — is the prototype for an enormously more general theory, and the goal of the present chapter is to set up the machinery that makes that generalisation possible.

The strategy, due to Killing and Cartan (1888–1894) and brought to its modern form by Weyl, Dynkin and others, is to show that *every* semisimple Lie algebra — no matter how large or exotic — can be reduced to a small, purely combinatorial object: a finite set of vectors in a Euclidean space (its **root system**), subject to a short list of rigid geometric constraints. Once the possible root systems are enumerated, one has, in effect, enumerated *every semisimple Lie algebra that can possibly exist*. This is one of the most complete classification theorems in all of mathematics, and it underlies the entire structure of Lie group theory as used in physics: every classical gauge group (the $SU(N)$ ’s, $SO(N)$ ’s, $Sp(N)$ ’s of the Standard Model and grand unified theories) and every classical global symmetry (isospin, flavour $SU(3)$, the Lorentz and Poincaré algebras) is an instance — or a close relative — of this classification.

We proceed in five stages. First (Section 35) we sharpen the notions of *simple* and *semisimple* Lie algebra, introduce the **Killing form**, and prove Cartan’s criteria characterising solvability and semisimplicity in terms of it. Second (Section 36) we introduce the *complexification* of a real Lie algebra, which lets us treat, e.g., $\mathfrak{su}(2)$ and $\mathfrak{sl}(2, \mathbb{C})$, or the rotation and boost generators of the Lorentz algebra, on the same algebraic footing — trading reality for enormous computational convenience, exactly as we did with the ladder operators $t_{\pm}$ in Part C. Third (Section 37) we generalise the ladder-operator construction of $\mathfrak{su}(2)$ to an arbitrary semisimple Lie algebra: choosing a **Cartan subalgebra** $\mathfrak{h}$ (the analogue of the single generator t_3), the rest of the algebra organises itself into **root spaces**, simultaneous eigenspaces of $\text{ad}(\mathfrak{h})$, labelled by vectors α (the **roots**) in the dual of $\mathfrak{h}$. Fourth (Section 38) we show that the roots obey rigid integrality and angle constraints — direct descendants of the $\mathfrak{su}(2)$ representation theory of Part C, applied to each root individually. Finally (Section 39–Section 40) we show that the entire root system is already determined by a small subset of **simple roots**, encode their pairwise relations in the **Cartan matrix** and its pictorial form, the **Dynkin diagram**, and state the classification theorem that completely enumerates all possible diagrams — and hence all simple complex Lie algebras.

Physical perspective

This chapter is the mathematical machine behind statements physicists take for granted, such as “the possible gauge groups are $SU(N)$, $SO(N)$, $Sp(N)$, and five exceptional cases,” or “the Standard Model is $SU(3) \times SU(2) \times U(1)$, and $SU(5)$ or $SO(10)$ grand-unify it.” Once you know the classification, choosing a gauge group is choosing a Dynkin diagram — a piece of combinatorics — and the representation content (particle multiplets) is then read off from the same root-and-weight machinery developed here and continued in the next chapter.

35 Simple and Semisimple Lie Algebras

35.1 Ideals, Simplicity, and Semisimplicity

Recall from Part C that a subspace $I \subseteq \mathfrak{g}$ of a Lie algebra $\mathfrak{g}$ is an **ideal** if $[\mathfrak{g}, I] \subseteq I$, i.e. bracketing with *anything* in $\mathfrak{g}$ never leaves I (this is the Lie-algebra analogue of a normal subgroup, and indeed is exactly what one obtains by linearising the defining property of a normal subgroup at the identity). Every Lie algebra has (at least) the two trivial ideals $\{0\}$ and $\mathfrak{g}$ itself.

Definition 35.1. Let $\mathfrak{g}$ be a (finite-dimensional) Lie algebra.

- (a) $\mathfrak{g}$ is **simple** if $\mathfrak{g}$ is non-abelian and has no ideals other than $\{0\}$ and $\mathfrak{g}$ itself.
- (b) $\mathfrak{g}$ is **semisimple** if $\mathfrak{g}$ contains no nonzero *abelian* ideal.

Remark. The exclusion of the abelian case in (a) is purely a matter of convention, needed to make the classification theorems clean: an abelian Lie algebra of dimension 1 trivially has no ideals other than $\{0\}$ and itself (it has no proper nonzero subspaces to speak of, when 1-dimensional), but we do not want to call it “simple,” since it carries no interesting bracket structure at all (its bracket is identically zero). Every simple Lie algebra is automatically semisimple: it has no proper nonzero ideals at all, hence in particular no nonzero *abelian* ideal.

We will shortly prove (Theorem 35.4) that semisimplicity is equivalent to the more intuitive statement that $\mathfrak{g}$ decomposes as a direct sum of simple ideals — “semisimple” literally means “built, without remainder, from simple pieces.”

35.2 The Killing Form

The tool that makes simplicity and semisimplicity *computable* — rather than a property one must verify by exhibiting or excluding ideals by hand — is a canonical symmetric bilinear form carried by every Lie algebra, built from the adjoint representation of Part C.

Definition 35.2. Let $\mathfrak{g}$ be a Lie algebra. Its **Killing form** is the symmetric bilinear form

$$K(a, b) := \text{Tr}(\text{ad}(a) \text{ad}(b)) \quad (a, b \in \mathfrak{g}),$$

where $\text{ad}(a) : \mathfrak{g} \rightarrow \mathfrak{g}$, $\text{ad}(a)b = [a, b]$, is the adjoint representation and the trace is that of a linear operator on the (finite-dimensional) vector space $\mathfrak{g}$.

Symmetry, $K(a, b) = K(b, a)$, is immediate from cyclicity of the trace, $\text{Tr}(AB) = \text{Tr}(BA)$. In a basis $\{t_a\}$ with structure constants $[t_a, t_b] = f_{ab}^c t_c$, the matrix of t_a in the adjoint representation is $(T_a^{\text{ad}})^b_c = f_{ba}^c$ (Definition 29.1), so that, with the trace taken over the contracted index pair,

$$K_{ab} := K(t_a, t_b) = (T_a^{\text{ad}})^d_c (T_b^{\text{ad}})^c_d = f_{ca}^d f_{db}^c = f_{ac}^d f_{bd}^c,$$

a totally explicit formula once the structure constants are known.

Theorem 35.1. The Killing form is **invariant**: for all $a, b, c \in \mathfrak{g}$,

$$K([a, b], c) = K(a, [b, c]), \quad \text{equivalently} \quad K([a, b], c) + K(b, [a, c]) = 0.$$

Proof. By definition, $K([a, b], c) = \text{Tr}(\text{ad}([a, b])\text{ad}(c))$. Since ad is a representation of $\mathfrak{g}$ (Part

C), $\text{ad}([a, b]) = [\text{ad}(a), \text{ad}(b)]$, so

$$\begin{aligned} K([a, b], c) &= \text{Tr}(\text{ad}(a) \text{ad}(b) \text{ad}(c)) - \text{Tr}(\text{ad}(b) \text{ad}(a) \text{ad}(c)) \\ &= \text{Tr}(\text{ad}(a) \text{ad}(b) \text{ad}(c)) - \text{Tr}(\text{ad}(a) \text{ad}(c) \text{ad}(b)) \quad (\text{cyclicity}) \\ &= \text{Tr}(\text{ad}(a) [\text{ad}(b) \text{ad}(c) - \text{ad}(c) \text{ad}(b)]) \\ &= \text{Tr}(\text{ad}(a) \text{ad}([b, c])) = K(a, [b, c]), \end{aligned}$$

(in the second line we used cyclicity of the trace, $\text{Tr}(\text{ad}(b) \text{ad}(a) \text{ad}(c)) = \text{Tr}(\text{ad}(a) \text{ad}(c) \text{ad}(b))$, shifting the last factor to the front), which is the first identity. For the second form, apply the first identity with $(a, b, c) \mapsto (a, c, b)$ to get $K([a, c], b) = K(a, [c, b]) = -K(a, [b, c])$, using antisymmetry of the bracket. By symmetry of K , the left-hand side is $K([a, c], b) = K(b, [a, c])$, so

$$K(b, [a, c]) = -K(a, [b, c]).$$

Adding $K([a, b], c) = K(a, [b, c])$ (the first identity) to this gives

$$K([a, b], c) + K(b, [a, c]) = K(a, [b, c]) - K(a, [b, c]) = 0. \quad \square$$

Remark. Invariance says that $\text{ad}(a)$ is “antisymmetric” with respect to K for every $a \in \mathfrak{g}$: $K(\text{ad}(a)b, c) = -K(b, \text{ad}(a)c)$. This is the algebraic shadow of the group-level fact that $K(\text{Ad}(g)b, \text{Ad}(g)c) = K(b, c)$ for all g in the corresponding Lie group (proved directly from $\text{Ad}(g) \text{ad}(b) \text{Ad}(g)^{-1} = \text{ad}(\text{Ad}(g)b)$ and cyclicity of the trace) — the Killing form is preserved by the adjoint action of the whole group, and infinitesimally by the whole algebra.

35.3 Cartan’s Criteria

Cartan’s two criteria turn the abstract, structural notions of solvability and semisimplicity into concrete linear-algebra computations on the Killing form. We first recall that a Lie algebra $\mathfrak{g}$ is **solvable** if its derived series $\mathfrak{g}^{(0)} = \mathfrak{g}$, $\mathfrak{g}^{(k+1)} = [\mathfrak{g}^{(k)}, \mathfrak{g}^{(k)}]$ terminates at $\{0\}$ after finitely many steps (this notion has not come up explicitly before this chapter).

Theorem 35.2. (Cartan’s first criterion.) A Lie algebra $\mathfrak{g}$ is solvable if and only if

$$K(a, b) = 0 \quad \text{for all } a \in \mathfrak{g}, b \in [\mathfrak{g}, \mathfrak{g}].$$

Theorem 35.3. (Cartan’s second criterion.) A Lie algebra $\mathfrak{g}$ is semisimple if and only if its Killing form is **non-degenerate**, i.e. $\det(K_{ab}) \neq 0$ (equivalently: $K(a, b) = 0$ for all $b \in \mathfrak{g}$ forces $a = 0$).

Remark. The proofs of both criteria require a fair amount of linear-algebraic machinery (Cartan’s original arguments, or the modern proof via Lie’s theorem on solvable Lie algebras of linear operators) that would take us rather far afield from our main goal — reaching the classification of root systems. We will therefore *use* both criteria as computational tools throughout the rest of this chapter (they are entirely standard and can be found with full proofs in, e.g., Humphreys’ *Introduction to Lie Algebras and Representation Theory* or Fulton–Harris’ *Representation Theory*), while proving everything that follows *from* them in full detail.

Theorem 35.4. Every semisimple Lie algebra $\mathfrak{g}$ is either simple, or decomposes as a direct sum of simple ideals,

$$\mathfrak{g} = \mathfrak{g}_1 \oplus \mathfrak{g}_2 \oplus \cdots \oplus \mathfrak{g}_n \quad (\text{each } \mathfrak{g}_i \text{ simple, } [\mathfrak{g}_i, \mathfrak{g}_j] = 0 \text{ for } i \neq j).$$

Sketch of proof. If $\mathfrak{g}$ is not simple, it has a proper nonzero ideal I . Using non-degeneracy of K (Cartan's second criterion, since $\mathfrak{g}$ is semisimple) one shows that the K -orthogonal complement $I^\perp := \{a \in \mathfrak{g} : K(a, b) = 0 \ \forall b \in I\}$ is *also* an ideal, and that $\mathfrak{g} = I \oplus I^\perp$ as vector spaces, with $[I, I^\perp] = 0$ (invariance of K , Theorem 35.1, shows $I^\perp$ is an ideal; non-degeneracy of K restricted to $\mathfrak{g}$ then forces $I \cap I^\perp = 0$ and a dimension count gives $\mathfrak{g} = I \oplus I^\perp$). One checks that both I and $I^\perp$, being ideals of a semisimple algebra, are themselves semisimple (any abelian ideal of I would also be an abelian ideal of $\mathfrak{g}$), so by induction on dimension both split further into simple ideals, and $\mathfrak{g}$ is their direct sum. $\square$

Theorem 35.5. If $\mathfrak{g}$ is semisimple, the adjoint representation $\text{ad} : \mathfrak{g} \rightarrow \mathfrak{gl}(\mathfrak{g})$ is **faithful** (injective): $\text{ad}(a) = \text{ad}(b)$ implies $a = b$.

Proof. Suppose $\text{ad}(a) = \text{ad}(b)$, i.e. $\text{ad}(a - b) = 0$ as a linear operator on $\mathfrak{g}$. Then for every $c \in \mathfrak{g}$,

$$K(a - b, c) = \text{Tr}(\text{ad}(a - b) \text{ad}(c)) = \text{Tr}(0) = 0.$$

Thus $a - b$ lies in the K -orthogonal complement of all of $\mathfrak{g}$. Since $\mathfrak{g}$ is semisimple, K is non-degenerate (Theorem 35.3), so $K(a - b, c) = 0$ for all $c \in \mathfrak{g}$ forces $a - b = 0$, i.e. $a = b$. $\square$

Theorem 35.6. If $\mathfrak{g}$ is **simple**, the adjoint representation of $\mathfrak{g}$ on the vector space $V = \mathfrak{g}$ is **irreducible**.

Proof. A subspace $W \subseteq \mathfrak{g}$ is invariant under $\text{ad}(\mathfrak{g})$ — i.e. $\text{ad}(a)w \in W$ for all $a \in \mathfrak{g}, w \in W$ — if and only if $[\mathfrak{g}, W] \subseteq W$, which is precisely the statement that W is an *ideal* of $\mathfrak{g}$. Since $\mathfrak{g}$ is simple, its only ideals are $\{0\}$ and $\mathfrak{g}$ itself; hence its only ad -invariant subspaces are $\{0\}$ and $\mathfrak{g}$, which is exactly the definition of irreducibility. $\square$

35.4 Examples

Example 35.1. $\mathfrak{su}(2)$ is **simple (hence semisimple)**. Using the anti-Hermitian basis $t_a = \frac{i}{2}\sigma_a$ of Part B ($[t_a, t_b] = -\varepsilon_{abc}t_c$, so $f_{ab}^c = -\varepsilon_{abc}$ — *not* to be confused with Part C's Hermitian $t_a = \frac{1}{2}\sigma_a$, which satisfies $[t_a, t_b] = +i\varepsilon_{abc}t_c$ instead; the two bases share the symbol t_a but are not the same generators), one computes that

$$K_{ab} = f_{ac}^d f_{bd}^c = -\varepsilon_{acd}(-\varepsilon_{bdc}) = \varepsilon_{acd}\varepsilon_{bdc} = -2\delta_{ab},$$

using $\varepsilon_{acd}\varepsilon_{bcd} = 2\delta_{ab}$ and one further sign from reordering the last two indices of the second ε . In particular $\det(K_{ab}) = (-2)^3 = -8 \neq 0$: the Killing form is non-degenerate, so $\mathfrak{su}(2)$ is semisimple by Cartan's second criterion. One further checks directly (or simply recalls from the classification of 2×2 traceless anti-Hermitian matrices) that $\mathfrak{su}(2)$, being 3-dimensional and non-abelian, has no 1- or 2-dimensional ideals, hence is in fact simple.

Example 35.2. $\mathfrak{su}(2) \oplus \mathfrak{su}(3)$ is **semisimple but not simple**. The direct sum of two Lie algebras $\mathfrak{g}_1 \oplus \mathfrak{g}_2$ (Part C, with $[(a_1, a_2), (b_1, b_2)] := ([a_1, b_1], [a_2, b_2])$) has $\mathfrak{g}_1 \oplus \{0\}$ as a proper,

nonzero, *non-abelian* ideal (it is an ideal since $[\mathfrak{g}_1 \oplus \mathfrak{g}_2, \mathfrak{g}_1 \oplus \{0\}] = [\mathfrak{g}_1, \mathfrak{g}_1] \oplus \{0\} \subseteq \mathfrak{g}_1 \oplus \{0\}$, using $[\mathfrak{g}_1, \{0\}] = \{0\}$). A non-abelian proper ideal already disqualifies $\mathfrak{su}(2) \oplus \mathfrak{su}(3)$ from being simple (Definition 35.1(a) requires *no* proper nonzero ideal, abelian or not). Yet $\mathfrak{su}(2) \oplus \mathfrak{su}(3)$ *is* semisimple: its only ideals are $\{0\}$, $\mathfrak{su}(2) \oplus \{0\}$, $\{0\} \oplus \mathfrak{su}(3)$, and the whole algebra (this uses that $\mathfrak{su}(2)$ and $\mathfrak{su}(3)$ are themselves simple, so a general ideal's projections onto each factor must, by the same argument as Theorem 35.6, be $\{0\}$ or the whole factor), and none of these is abelian except $\{0\}$ itself. This is the generic situation described by Theorem 35.4: a semisimple but non-simple algebra is precisely a direct sum of two or more simple ideals.

Example 35.3. $\mathfrak{u}(N)$ is *not* semisimple. Recall $\mathfrak{u}(N) = \mathfrak{su}(N) \oplus \mathfrak{u}(1)$ as vector spaces (Part B), where $\mathfrak{u}(1) \cong \mathbb{R}$ is spanned by $i\mathbb{1}_N$, and (since $i\mathbb{1}_N$ commutes with everything) $[\mathfrak{u}(N), \mathfrak{u}(1)] = 0$: the $\mathfrak{u}(1)$ factor is a nonzero *abelian* ideal of $\mathfrak{u}(N)$ (it is central, hence trivially an ideal, and 1-dimensional, hence abelian). By Definition 35.1(b), the existence of this nonzero abelian ideal means $\mathfrak{u}(N)$ fails to be semisimple. Concretely: since $\text{ad}(i\mathbb{1}_N) = 0$ identically, the row and column of the Killing form corresponding to the $\mathfrak{u}(1)$ generator vanish, so K is (maximally) degenerate — consistent with Cartan's second criterion.

Example 35.4. $\mathfrak{so}(4) \cong \mathfrak{so}(3) \oplus \mathfrak{so}(3)$ is semisimple but not simple. Writing M_{ij} ($1 \leq i < j \leq 4$) for the six generators of $\mathfrak{so}(4)$ and defining $A_i := \frac{1}{2}\varepsilon_{ijk}M_{jk} + M_{i4}$, $B_i := \frac{1}{2}\varepsilon_{ijk}M_{jk} - M_{i4}$ for $i, j, k \in \{1, 2, 3\}$, one finds by direct computation from $[M_{ij}, M_{kl}] = \delta_{jk}M_{il} - \delta_{ik}M_{jl} - \delta_{jl}M_{ik} + \delta_{il}M_{jk}$ that $[A_i, A_j] = -2\varepsilon_{ijk}A_k$, $[B_i, B_j] = -2\varepsilon_{ijk}B_k$, and $[A_i, B_j] = 0$ (the overall factor of -2 , rather than -1 , simply reflects the normalisation chosen for A_i, B_i above and does not affect the isomorphism type): the A 's and B 's each close on their own copy of $\mathfrak{so}(3) \cong \mathfrak{su}(2)$ (after rescaling by $-\frac{1}{2}$ to match the standard normalisation), and the two copies commute with each other. Hence $\mathfrak{so}(4) \cong \mathfrak{so}(3) \oplus \mathfrak{so}(3)$, exactly as in Example 35.2: semisimple (direct sum of two simple algebras) but not simple (each $\mathfrak{so}(3)$ summand is a proper non-abelian ideal). This decomposition is the algebraic root of the well-known fact that $SO(4)$ is locally isomorphic to $SU(2) \times SU(2)$, with important applications in the study of the hydrogen atom's "accidental" $O(4)$ degeneracy and in four-dimensional (anti-)self-dual gauge theory.

36 Complexification and the Lorentz Algebra

36.1 Complexification and Real Forms

We already introduced the complexification of a Lie algebra in passing in Part C; we now make the construction precise and prove the key fact that makes it useful: representations of a real Lie algebra and of its complexification are in natural bijection, so that nothing is lost — and, as we will see, a great deal of computational convenience is gained — by working with the complexified algebra.

Definition 36.1. Let $\mathfrak{g}$ be a real Lie algebra with basis $\{t_a\}$ and structure constants $[t_a, t_b] = f_{ab}^c t_c$ ($f_{ab}^c \in \mathbb{R}$). Its **complexification** $\tilde{\mathfrak{g}} := \mathfrak{g}_{\mathbb{C}}$ is the complex vector space of formal sums $a + ib$ ($a, b \in \mathfrak{g}$), i.e. $\tilde{\mathfrak{g}} = \mathfrak{g} \oplus i\mathfrak{g}$ as a real vector space but now equipped with multiplication by i , with bracket extended $\mathbb{C}$ -bilinearly:

$$[a_1 + ib_1, a_2 + ib_2] := ([a_1, a_2] - [b_1, b_2]) + i([a_1, b_2] + [b_1, a_2]).$$

Equivalently: $\{t_a\}$ remains a basis of $\tilde{\mathfrak{g}}$ as a *complex* vector space, with the *same* structure constants f_{ab}^c , now allowing complex coefficients. If $\mathfrak{g}_{\mathbb{C}} \cong \tilde{\mathfrak{g}}$ for a given complex Lie algebra $\tilde{\mathfrak{g}}$, the real Lie algebra $\mathfrak{g}$ is called a **real form** of $\tilde{\mathfrak{g}}$.

Two features of this definition deserve emphasis. First, complexification does *not* change the number of generators: $\dim_{\mathbb{C}} \tilde{\mathfrak{g}} = \dim_{\mathbb{R}} \mathfrak{g}$ (we merely allow complex linear combinations of the *same* basis, rather than only real ones). Second, a single complex Lie algebra $\tilde{\mathfrak{g}}$ can have *several*, non-isomorphic real forms $\mathfrak{g}$ — we will see this explicitly for $\mathfrak{sl}(2, \mathbb{C})$ in a moment, and it is precisely the freedom exploited when we pass, e.g., from the compact rotation group to the non-compact Lorentz group while keeping “the same” complexified algebra.

Theorem 36.1. Let $\tilde{\mathfrak{g}} = \mathfrak{g}_{\mathbb{C}}$ be the complexification of a real Lie algebra $\mathfrak{g}$. There is a natural bijection between (finite-dimensional) complex-linear representations $\Gamma_{\tilde{\mathfrak{g}}}$ of $\tilde{\mathfrak{g}}$ and (finite-dimensional, complex-vector-space-valued) representations $\Gamma_{\mathfrak{g}}$ of $\mathfrak{g}$, given by restriction one way and complex-linear extension the other:

$$\Gamma_{\mathfrak{g}}(a) := \Gamma_{\tilde{\mathfrak{g}}}(a + i0) \quad (a \in \mathfrak{g}), \quad \Gamma_{\tilde{\mathfrak{g}}}(a + ib) := \Gamma_{\mathfrak{g}}(a) + i\Gamma_{\mathfrak{g}}(b).$$

Proof. Given $\Gamma_{\tilde{\mathfrak{g}}}$, define $\Gamma_{\mathfrak{g}}$ by restriction to $\mathfrak{g} \hookrightarrow \tilde{\mathfrak{g}}$, $a \mapsto a + i0$; since $\tilde{\mathfrak{g}}$ ’s bracket restricted to $\mathfrak{g}$ agrees with $\mathfrak{g}$ ’s own bracket (by construction of $\tilde{\mathfrak{g}}$) and $\Gamma_{\tilde{\mathfrak{g}}}$ is a homomorphism, $\Gamma_{\mathfrak{g}}([a, b]) = \Gamma_{\tilde{\mathfrak{g}}}([a, b] + i0) = [\Gamma_{\tilde{\mathfrak{g}}}(a + i0), \Gamma_{\tilde{\mathfrak{g}}}(b + i0)] = [\Gamma_{\mathfrak{g}}(a), \Gamma_{\mathfrak{g}}(b)]$: $\Gamma_{\mathfrak{g}}$ is indeed a representation of $\mathfrak{g}$.

Conversely, given a representation $\Gamma_{\mathfrak{g}}$ of $\mathfrak{g}$ on a complex vector space V , define $\Gamma_{\tilde{\mathfrak{g}}}(a + ib) := \Gamma_{\mathfrak{g}}(a) + i\Gamma_{\mathfrak{g}}(b)$; this is manifestly $\mathbb{C}$ -linear in $a + ib$, and one checks directly, using bilinearity of both brackets and the fact that $\Gamma_{\mathfrak{g}}$ is a representation of $\mathfrak{g}$, that

$$\begin{aligned} [\Gamma_{\tilde{\mathfrak{g}}}(a_1 + ib_1), \Gamma_{\tilde{\mathfrak{g}}}(a_2 + ib_2)] &= [\Gamma_{\mathfrak{g}}(a_1) + i\Gamma_{\mathfrak{g}}(b_1), \Gamma_{\mathfrak{g}}(a_2) + i\Gamma_{\mathfrak{g}}(b_2)] \\ &= ([\Gamma_{\mathfrak{g}}(a_1), \Gamma_{\mathfrak{g}}(a_2)] - [\Gamma_{\mathfrak{g}}(b_1), \Gamma_{\mathfrak{g}}(b_2)]) \\ &\quad + i([\Gamma_{\mathfrak{g}}(a_1), \Gamma_{\mathfrak{g}}(b_2)] + [\Gamma_{\mathfrak{g}}(b_1), \Gamma_{\mathfrak{g}}(a_2)]) \\ &= \Gamma_{\mathfrak{g}}([a_1, a_2] - [b_1, b_2]) + i\Gamma_{\mathfrak{g}}([a_1, b_2] + [b_1, a_2]) \\ &= \Gamma_{\tilde{\mathfrak{g}}}([a_1 + ib_1, a_2 + ib_2]), \end{aligned}$$

using the definition of the bracket on $\tilde{\mathfrak{g}}$ in the last step: $\Gamma_{\tilde{\mathfrak{g}}}$ is a representation of $\tilde{\mathfrak{g}}$. The two constructions are manifestly mutually inverse. $\square$

Remark. This theorem is the rigorous justification for a trick we already used freely in Part C: to classify the representations of the real algebra $\mathfrak{su}(2)$, we introduced the *complex* combinations $t_{\pm} = t_1 \pm it_2$ and worked entirely inside the complexification $\mathfrak{su}(2)_{\mathbb{C}} \cong \mathfrak{sl}(2, \mathbb{C})$, where $t_{\pm}$ and t_3 form a convenient basis (raising/lowering/Cartan generators) even though $t_{\pm} \notin \mathfrak{su}(2)$ itself. Theorem 36.1 guarantees that every representation of $\mathfrak{sl}(2, \mathbb{C})$ found this way restricts to a genuine representation of $\mathfrak{su}(2)$, and every representation of $\mathfrak{su}(2)$ arises this way — so no information is gained or lost, only computational convenience.

36.2 Example: $\mathfrak{sl}(2, \mathbb{C})$ and Its Two Real Forms

Example 36.1. $\mathfrak{su}(2)$ is a real form of $\mathfrak{sl}(2, \mathbb{C})$. With the basis $t_a = \frac{i}{2}\sigma_a$ of $\mathfrak{su}(2)$ ($[t_a, t_b] = -\varepsilon_{abc}t_c$), the complexification $\mathfrak{su}(2)_{\mathbb{C}}$ is, by Definition 36.1, the 3-(complex-)dimensional complex Lie algebra spanned by t_1, t_2, t_3 with complex coefficients allowed. Since $\{t_a\}$ are (as 2×2 complex matrices) linearly independent over $\mathbb{C}$ and span exactly the traceless 2×2

complex matrices (any traceless complex 2×2 matrix is a complex linear combination of $i\sigma_1, i\sigma_2, i\sigma_3$), we get $\mathfrak{su}(2)_{\mathbb{C}} \cong \mathfrak{sl}(2, \mathbb{C})$: the complexification of $\mathfrak{su}(2)$ is exactly the complex Lie algebra of traceless 2×2 complex matrices under the commutator, the algebra used to classify $\mathfrak{su}(2)$ representations in Part C.

Example 36.2. $\mathfrak{sl}(2, \mathbb{R})$ is a *different* real form of the same $\mathfrak{sl}(2, \mathbb{C})$. Take the basis of traceless 2×2 real matrices

$$b_1 = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad b_2 = \frac{1}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad b_3 = \frac{1}{2} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

A direct computation of matrix commutators gives

$$[b_1, b_2] = b_3, \quad [b_2, b_3] = -b_1, \quad [b_3, b_1] = -b_2,$$

so $\mathfrak{sl}(2, \mathbb{R})$ is again 3-dimensional with real structure constants; complexifying (allowing complex coefficients of the *same* basis b_1, b_2, b_3) evidently reproduces the traceless 2×2 complex matrices, i.e. $\mathfrak{sl}(2, \mathbb{R})_{\mathbb{C}} \cong \mathfrak{sl}(2, \mathbb{C})$ as well. So $\mathfrak{su}(2)$ and $\mathfrak{sl}(2, \mathbb{R})$ are *both* real forms of the same complex Lie algebra $\mathfrak{sl}(2, \mathbb{C})$.

They are, however, **not isomorphic as real Lie algebras**, and the Killing form detects this immediately. Writing out $\text{ad}(b_1), \text{ad}(b_2), \text{ad}(b_3)$ as 3×3 matrices in the basis (b_1, b_2, b_3) from the brackets above and computing traces of products (exactly as in Example 35.1) gives the off-diagonal entries $K(b_i, b_j) = 0$ for $i \neq j$ and

$$K(b_1, b_1) = 2, \quad K(b_2, b_2) = 2, \quad K(b_3, b_3) = -2.$$

The Killing form of $\mathfrak{sl}(2, \mathbb{R})$ is therefore **indefinite** (signature $(2, 1)$): non-degenerate, so $\mathfrak{sl}(2, \mathbb{R})$ is semisimple (consistent with $\det K = 2 \cdot 2 \cdot (-2) = -8 \neq 0$, the same *numerical* determinant as for $\mathfrak{su}(2)$ in Example 35.1 — a coincidence of these two particular, similarly-normalised bases rather than a general fact: $\det K$ itself is *not* basis-independent, since under a change of basis $t_a \mapsto S_a^b t_b$ the Gram-type matrix transforms as $K \mapsto S^T K S$, rescaling $\det K$ by $(\det S)^2 > 0$; what *is* basis-independent is the *sign* of $\det K$ and whether it vanishes, which is all Cartan’s second criterion actually uses), but *not* of one definite sign, unlike $\mathfrak{su}(2)$ ’s $K_{ab} = -2\delta_{ab}$, which is negative-definite. Since an isomorphism of real Lie algebras would have to carry one Killing form onto the other and thus preserve its signature, $\mathfrak{su}(2) \not\cong \mathfrak{sl}(2, \mathbb{R})$.

Remark. The two real forms in Examples 36.1 and 36.2 play very different roles: $\mathfrak{su}(2)$, with its negative-definite Killing form, is the **compact real form** of $\mathfrak{sl}(2, \mathbb{C})$ (it is, up to isomorphism, the unique real form for which the corresponding Lie group is compact — recall from Part C that a negative-definite ad-invariant form is exactly what makes the Haar measure argument for compactness go through). $\mathfrak{sl}(2, \mathbb{R}) \cong \mathfrak{so}(2, 1)$, with its indefinite Killing form, is the **split (or “normal”) real form**, associated with the non-compact group $SL(2, \mathbb{R})$. In general, a given complex simple Lie algebra has exactly one compact real form but can have several non-compact ones; distinguishing them, and understanding which groups are compact, is exactly the kind of “global” information that is invisible to the complexified algebra alone, as anticipated in Section 34.

36.3 Physical Application: The Lorentz Algebra $\mathfrak{so}(3, 1)$

We now apply exactly this circle of ideas to the Lorentz algebra, with dramatic effect: the complexified Lorentz algebra splits as a *direct sum* of two commuting copies of $\mathfrak{su}(2)_{\mathbb{C}}$, which is the algebraic origin of the entire (j, j') classification of relativistic fields (scalars, spinors, vectors, ...).

Let θ_p ($p = 1, 2, 3$) denote the generators of spatial rotations and θ_{p+3} ($p = 1, 2, 3$) the generators of boosts, satisfying (in the conventions of Part B, matching the antisymmetric $\mathfrak{su}(2)$ convention $[t_a, t_b] = -\varepsilon_{abc}t_c$ used throughout)

$$[\theta_p, \theta_q] = -\varepsilon_{pqr}\theta_r, \quad [\theta_p, \theta_{q+3}] = -\varepsilon_{pqr}\theta_{r+3}, \quad [\theta_{p+3}, \theta_{q+3}] = \varepsilon_{pqr}\theta_r.$$

The first relation says the rotations θ_p close on their own $\mathfrak{so}(3) \cong \mathfrak{su}(2)$ subalgebra; the second says the boosts θ_{p+3} transform as a 3-vector under rotations; the third — the key, distinctively relativistic relation — says that the commutator of two boosts is *not* zero but produces a rotation (the algebraic origin of Thomas precession). Because of this last relation, $\{\theta_{p+3}\}$ does not span an ideal, and $\mathfrak{so}(3, 1)$ has no proper ideal at all as a *real* Lie algebra: it is simple.

Example 36.3. Complexifying the Lorentz algebra. Passing to the complexification $\mathfrak{so}(3, 1)_{\mathbb{C}}$, define the complex combinations

$$A_p := \frac{1}{2}(\theta_p + i\theta_{p+3}), \quad B_p := \frac{1}{2}(\theta_p - i\theta_{p+3}) \quad (p = 1, 2, 3).$$

Using the three bracket relations above, a direct computation gives

$$\begin{aligned} [A_p, A_q] &= \frac{1}{4}([\theta_p, \theta_q] + i[\theta_p, \theta_{q+3}] + i[\theta_{p+3}, \theta_q] - [\theta_{p+3}, \theta_{q+3}]) \\ &= \frac{1}{4}(-\varepsilon_{pqr}\theta_r - i\varepsilon_{pqr}\theta_{r+3} - i\varepsilon_{pqr}\theta_{r+3} - \varepsilon_{pqr}\theta_r) \\ &= \frac{1}{4}(-2\varepsilon_{pqr}\theta_r - 2i\varepsilon_{pqr}\theta_{r+3}) = -\frac{1}{2}\varepsilon_{pqr}(\theta_r + i\theta_{r+3}) = -\varepsilon_{pqr}A_r, \end{aligned}$$

where we used $[\theta_{p+3}, \theta_q] = -[\theta_q, \theta_{p+3}] = -\varepsilon_{pqr}\theta_{r+3}$ (antisymmetrising the second relation). An identical computation gives $[B_p, B_q] = -\varepsilon_{pqr}B_r$, while for the cross-bracket the θ_r and θ_{r+3} contributions cancel instead of adding:

$$\begin{aligned} [A_p, B_q] &= \frac{1}{4}([\theta_p, \theta_q] - i[\theta_p, \theta_{q+3}] + i[\theta_{p+3}, \theta_q] + [\theta_{p+3}, \theta_{q+3}]) \\ &= \frac{1}{4}(-\varepsilon_{pqr}\theta_r + i\varepsilon_{pqr}\theta_{r+3} - i\varepsilon_{pqr}\theta_{r+3} + \varepsilon_{pqr}\theta_r) = 0. \end{aligned}$$

So A_1, A_2, A_3 satisfy exactly the $\mathfrak{su}(2)$ bracket relations of Example 35.1, likewise B_1, B_2, B_3 , and the two triples commute with each other:

$$\mathfrak{so}(3, 1)_{\mathbb{C}} \cong \mathfrak{su}(2)_{\mathbb{C}} \oplus \mathfrak{su}(2)_{\mathbb{C}} \cong \mathfrak{sl}(2, \mathbb{C}) \oplus \mathfrak{sl}(2, \mathbb{C}).$$

Physical perspective

This is one of the most consequential computations in relativistic physics. Theorem 36.1 tells us that classifying representations of $\mathfrak{so}(3, 1)_{\mathbb{C}}$ is exactly as good as classifying representations of the real Lorentz algebra $\mathfrak{so}(3, 1)$; and Example 36.3 tells us that this reduces to classifying representations of *two independent copies of $\mathfrak{su}(2)$* , which we already solved completely in Part C. A finite-dimensional representation of the Lorentz algebra is therefore labelled by a pair of $\mathfrak{su}(2)$ spins (j, j') , of dimension $(2j+1)(2j'+1)$: $(0, 0)$ is the scalar, $(\frac{1}{2}, 0)$ and $(0, \frac{1}{2})$ are the left- and right-handed Weyl spinors, $(\frac{1}{2}, \frac{1}{2})$ is the vector, and so on. Note carefully,

though, that $\mathfrak{so}(3, 1)$ is *simple* as a real Lie algebra (there is no way to split rotations from boosts using only *real* linear combinations, as we observed above): the splitting into two commuting pieces is a genuinely complex phenomenon, an instance of the general fact that a simple *real* Lie algebra's complexification can fail to be simple, exactly the reverse-looking but entirely consistent partner statement to Example 35.2 (where a semisimple-but-not-simple *complex* algebra arose instead).

37 Cartan Subalgebras and the Root Space Decomposition

From this point on we work with a **complex semisimple Lie algebra** $\tilde{\mathfrak{g}}$ (the complexification of some real Lie algebra of interest, by Section 36), and generalise, systematically, the ladder-operator diagonalisation of $\mathfrak{su}(2)_{\mathbb{C}} \cong \mathfrak{sl}(2, \mathbb{C})$ used in Part C: there, the single generator t_3 was diagonalised, and the remaining two generators $t_{\pm}$ organised themselves into its ± 1 -eigenspaces (under $\text{ad}(t_3)$, eigenvalues ± 1 ; the eigenvalues ± 2 belong instead to $\text{ad}(H_{\alpha})$ with $H_{\alpha} = 2t_3$, introduced below). We now find the correct generalisation of “ t_3 ” for an arbitrary $\tilde{\mathfrak{g}}$.

37.1 Cartan Subalgebras

Definition 37.1. Let $\tilde{\mathfrak{g}}$ be a complex semisimple Lie algebra. A **Cartan subalgebra** $\mathfrak{h} \subset \tilde{\mathfrak{g}}$ is a maximal abelian subalgebra of $\tilde{\mathfrak{g}}$ all of whose elements are *ad-semisimple*, i.e. $\text{ad}(h)$ is diagonalisable (over $\mathbb{C}$) for every $h \in \mathfrak{h}$ (maximal abelian: no strictly larger abelian subalgebra of $\tilde{\mathfrak{g}}$ contains it).

One can show that all Cartan subalgebras of a given $\tilde{\mathfrak{g}}$ are conjugate under the group generated by $\exp(\text{ad}(x))$, $x \in \tilde{\mathfrak{g}}$, so $\dim_{\mathbb{C}} \mathfrak{h}$ is an invariant of $\tilde{\mathfrak{g}}$ itself.

Definition 37.2. The dimension $\ell := \dim_{\mathbb{C}} \mathfrak{h}$ of a (any) Cartan subalgebra of $\tilde{\mathfrak{g}}$ is called the **rank** of $\tilde{\mathfrak{g}}$.

Let $\{h_1, \dots, h_{\ell}\}$ be a basis of $\mathfrak{h}$. Since $\mathfrak{h}$ is abelian, $[h_i, h_j] = 0$ for all i, j , so the operators $\text{ad}(h_1), \dots, \text{ad}(h_{\ell})$ are mutually commuting; since each is (by definition of $\mathfrak{h}$) diagonalisable, standard linear algebra guarantees they are **simultaneously** diagonalisable. Hence $\tilde{\mathfrak{g}}$ decomposes into simultaneous eigenspaces

$$\tilde{\mathfrak{g}}_{\alpha} := \{x \in \tilde{\mathfrak{g}} : [h, x] = \alpha(h)x \text{ for all } h \in \mathfrak{h}\},$$

labelled by linear functionals $\alpha \in \mathfrak{h}^*$ (i.e. $\alpha : \mathfrak{h} \rightarrow \mathbb{C}$ linear); concretely, if $\{x_i\}$ is a simultaneous eigenbasis with $[h_i, x] = \lambda_i x$, then $\alpha(h) := \sum_i \xi^i \lambda_i$ for $h = \sum_i \xi^i h_i$. By maximality of $\mathfrak{h}$, the *zero*-eigenspace is exactly $\mathfrak{h}$ itself: $\tilde{\mathfrak{g}}_0 = \mathfrak{h}$ (any $x \notin \mathfrak{h}$ commuting with all of $\mathfrak{h}$ would let us enlarge $\mathfrak{h}$ to the strictly bigger abelian subalgebra $\mathfrak{h} + \mathbb{C}x$, contradicting maximality).

Definition 37.3. A nonzero linear functional $\alpha \in \mathfrak{h}^* \setminus \{0\}$ is called a **root** of $\tilde{\mathfrak{g}}$ (with respect to $\mathfrak{h}$) if $\tilde{\mathfrak{g}}_{\alpha} \neq \{0\}$. We write $\Delta \subset \mathfrak{h}^*$ for the (finite) set of all roots, and $\tilde{\mathfrak{g}}_{\alpha}$ for the corresponding **root space**. This gives the **root space decomposition**

$$\tilde{\mathfrak{g}} = \mathfrak{h} \oplus \bigoplus_{\alpha \in \Delta} \tilde{\mathfrak{g}}_{\alpha}.$$

Theorem 37.1. For roots (or zero) α, β and $x_\alpha \in \tilde{\mathfrak{g}}_\alpha$, $x_\beta \in \tilde{\mathfrak{g}}_\beta$:

$$[x_\alpha, x_\beta] \in \tilde{\mathfrak{g}}_{\alpha+\beta}.$$

In particular, if $\alpha + \beta \neq 0$ is *not* a root, then $[x_\alpha, x_\beta] = 0$; and if $\alpha + \beta = 0$, then $[x_\alpha, x_{-\alpha}] \in \mathfrak{h}$.

Proof. For any $h \in \mathfrak{h}$, the Jacobi identity gives

$$[h, [x_\alpha, x_\beta]] = [[h, x_\alpha], x_\beta] + [x_\alpha, [h, x_\beta]] = \alpha(h)[x_\alpha, x_\beta] + \beta(h)[x_\alpha, x_\beta] = (\alpha + \beta)(h)[x_\alpha, x_\beta].$$

So $[x_\alpha, x_\beta]$ is a simultaneous $\text{ad}(\mathfrak{h})$ -eigenvector with eigenvalue functional $\alpha + \beta$, i.e. $[x_\alpha, x_\beta] \in \tilde{\mathfrak{g}}_{\alpha+\beta}$ by definition (Definition 37.3 together with the general eigenspace decomposition; here $\tilde{\mathfrak{g}}_{\alpha+\beta} = \{0\}$ counts as a valid, if trivial, case, forcing $[x_\alpha, x_\beta] = 0$ whenever $\alpha + \beta$ is neither a root nor zero). If $\alpha + \beta = 0$, the relevant eigenspace is $\tilde{\mathfrak{g}}_0 = \mathfrak{h}$. $\square$

37.2 The Killing Form on $\mathfrak{h}$, and the Induced Form on Roots

Theorem 37.2. If $\alpha, \beta \in \mathfrak{h}^*$ (roots or zero) with $\alpha + \beta \neq 0$, then $K(x_\alpha, x_\beta) = 0$ for all $x_\alpha \in \tilde{\mathfrak{g}}_\alpha$, $x_\beta \in \tilde{\mathfrak{g}}_\beta$.

Proof. By invariance of the Killing form (Theorem 35.1), for any $h \in \mathfrak{h}$,

$$0 = K([h, x_\alpha], x_\beta) + K(x_\alpha, [h, x_\beta]) = \alpha(h)K(x_\alpha, x_\beta) + \beta(h)K(x_\alpha, x_\beta) = (\alpha + \beta)(h)K(x_\alpha, x_\beta).$$

Since $\alpha + \beta \neq 0$ as a functional on $\mathfrak{h}$, we may choose $h \in \mathfrak{h}$ with $(\alpha + \beta)(h) \neq 0$, forcing $K(x_\alpha, x_\beta) = 0$. $\square$

Theorem 37.3. The Killing form of a semisimple $\tilde{\mathfrak{g}}$, restricted to a Cartan subalgebra $\mathfrak{h}$, is **non-degenerate**.

Proof. Suppose not: there is a nonzero $h' \in \mathfrak{h}$ with $K(h', h) = 0$ for all $h \in \mathfrak{h}$. By Theorem 37.2 applied with $\beta = 0$ (so $x_\beta = h' \in \mathfrak{h} = \tilde{\mathfrak{g}}_0$) and any root $\alpha \neq 0$ (so $\alpha + \beta = \alpha \neq 0$), we also have $K(x_\alpha, h') = 0$ for every $x_\alpha \in \tilde{\mathfrak{g}}_\alpha$ and every root α . Combining both statements and using the root space decomposition $\tilde{\mathfrak{g}} = \mathfrak{h} \oplus \bigoplus_\alpha \tilde{\mathfrak{g}}_\alpha$, we conclude $K(h', y) = 0$ for *every* $y \in \tilde{\mathfrak{g}}$. But $\tilde{\mathfrak{g}}$ is semisimple, so K is non-degenerate on all of $\tilde{\mathfrak{g}}$ by Cartan's second criterion (Theorem 35.3) — contradiction. Hence no such h' exists, and $K|_{\mathfrak{h}}$ is non-degenerate. $\square$

Non-degeneracy of $K|_{\mathfrak{h}}$ lets us transport each root $\alpha \in \mathfrak{h}^*$ to an element $h_\alpha \in \mathfrak{h}$ via

$$K(h_\alpha, h) = \alpha(h) \quad \text{for all } h \in \mathfrak{h}$$

(h_α exists and is unique precisely because a non-degenerate bilinear form identifies a vector space with its dual). This, in turn, defines a symmetric bilinear form directly on the roots (or more generally on all of $\mathfrak{h}^*$):

$$\langle \alpha, \beta \rangle := K(h_\alpha, h_\beta) = \alpha(h_\beta) = \beta(h_\alpha).$$

One further shows — this is where semisimplicity is used most essentially, via a theorem of Cartan not reproduced here — that $\langle \cdot, \cdot \rangle$ is real and **positive-definite** when restricted to the real span $\mathfrak{h}_{\mathbb{R}}^*$ of the roots: the roots of any complex semisimple Lie algebra live in a genuine finite-dimensional *Euclidean* space. This is what makes the forthcoming classification a problem in ordinary Euclidean geometry.

37.3 Worked Example: $\mathfrak{su}(2)_{\mathbb{C}} \cong \mathfrak{sl}(2, \mathbb{C})$

Example 37.1. Take $\mathfrak{h} = \mathbb{C}h_3$ (rank $\ell = 1$), the complexification of $\mathbb{R}t_3 \subset \mathfrak{su}(2)$. Using $t_{\pm} = t_1 \pm it_2$ and $[t_3, t_{\pm}] = \pm t_{\pm}$ (Part C's ladder-operator relations, up to the convention t_3 here playing the role of h_3), the root space decomposition is simply

$$\mathfrak{sl}(2, \mathbb{C}) = \underbrace{\mathbb{C}t_3}_{\mathfrak{h}} \oplus \underbrace{\mathbb{C}t_+}_{\mathfrak{g}_{\alpha}} \oplus \underbrace{\mathbb{C}t_-}_{\mathfrak{g}_{-\alpha}}, \quad \alpha(t_3) = 1,$$

i.e. there is exactly one pair of roots, $\pm\alpha$, with $\alpha(t_3) = 1$: the “root system of $\mathfrak{su}(2)$ ” is the simplest possible one, two opposite points on a line. Computing $\text{ad}(t_3)$ explicitly in the basis (t_3, t_+, t_-) gives $\text{ad}(t_3) = \text{diag}(0, 1, -1)$, so

$$K(t_3, t_3) = \text{Tr}(\text{ad}(t_3)^2) = 0^2 + 1^2 + (-1)^2 = 2,$$

and hence $h_{\alpha} = \alpha(t_3)/K(t_3, t_3) \cdot t_3 = t_3/2$ (solving $K(h_{\alpha}, t_3) = \alpha(t_3) = 1$ for $h_{\alpha} = ct_3$: $c \cdot 2 = 1 \Rightarrow c = \frac{1}{2}$), and $\langle \alpha, \alpha \rangle = K(h_{\alpha}, h_{\alpha}) = \frac{1}{4} \cdot 2 = \frac{1}{2}$.

37.4 Worked Example: $\mathfrak{su}(3)_{\mathbb{C}}$ and the Gell-Mann Matrices

Example 37.2. For the physically central case of $\mathfrak{su}(3)$ (flavour/colour symmetry), a standard Hermitian basis is given by the eight traceless Hermitian 3×3 **Gell-Mann matrices** $\lambda_1, \dots, \lambda_8$ (generators are $t_a = \frac{i}{2}\lambda_a$, exactly as $t_a = \frac{i}{2}\sigma_a$ for $\mathfrak{su}(2)$):

$$\begin{aligned} \lambda_1 &= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \\ \lambda_5 &= \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \quad \lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad \lambda_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}. \end{aligned}$$

Only λ_3 and λ_8 are diagonal, so $\mathfrak{h} := \mathbb{C}t_3 \oplus \mathbb{C}t_8$ (with $t_3 = \frac{i}{2}\lambda_3$, $t_8 = \frac{i}{2}\lambda_8$) is a Cartan subalgebra: $\mathfrak{su}(3)$ has **rank** $\ell = 2$. The remaining six generators pair up, exactly as t_1, t_2 did for $\mathfrak{su}(2)$, into three ladder pairs

$$E_{\pm(1,0)} \propto t_1 \pm it_2, \quad E_{\pm(\frac{1}{2}, \frac{\sqrt{3}}{2})} \propto t_4 \pm it_5, \quad E_{\pm(-\frac{1}{2}, \frac{\sqrt{3}}{2})} \propto t_6 \pm it_7,$$

each an eigenvector of $\text{ad}(t_3)$ and $\text{ad}(t_8)$ simultaneously. Explicitly, writing $[t_3, E_{\alpha}] = \alpha_1 E_{\alpha}$, $[t_8, E_{\alpha}] = \alpha_2 E_{\alpha}$ and reading off the eigenvalues from the commutators of the Gell-Mann matrices (a standard, if lengthy, direct computation) gives *six* nonzero roots, forming a regular hexagon in the Euclidean plane $\mathfrak{h}_{\mathbb{R}}^* \cong \mathbb{R}^2$:

$$\Delta = \left\{ \pm(1, 0), \pm\left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right), \pm\left(-\frac{1}{2}, \frac{\sqrt{3}}{2}\right) \right\}.$$

We will re-derive this root system systematically, from the Cartan matrix alone, in Example 41.1 below, and draw it explicitly in Fig. 16.

38 Root Strings and the Master Formula

We now generalise the $\mathfrak{su}(2)$ representation theory of Part C — ladder operators terminating a finite tower of states — to the whole of $\tilde{\mathfrak{g}}$ at once, by showing that *every single root* α comes with its own copy of $\mathfrak{sl}(2, \mathbb{C})$, under which the entire algebra organises itself into finite-dimensional representations. This is the source of essentially every integrality and rigidity statement in the classification.

38.1 The Chevalley Basis Attached to a Root

Definition 38.1. For a root $\alpha \in \Delta$, define

$$H_\alpha := \frac{2h_\alpha}{\langle \alpha, \alpha \rangle} \in \mathfrak{h},$$

where $h_\alpha \in \mathfrak{h}$ is as in Section 37 ($K(h_\alpha, h) = \alpha(h)$). Choosing (it is a standard fact, which we do not reprove, that this is always possible) $E_\alpha \in \tilde{\mathfrak{g}}_\alpha$, $E_{-\alpha} \in \tilde{\mathfrak{g}}_{-\alpha}$ with $[E_\alpha, E_{-\alpha}] = H_\alpha$, the triple $\{H_\alpha, E_\alpha, E_{-\alpha}\}$ is called the **Chevalley basis** attached to α .

Theorem 38.1. The Chevalley basis attached to any root α satisfies

$$[H_\alpha, E_{\pm\alpha}] = \pm 2 E_{\pm\alpha}, \quad [E_\alpha, E_{-\alpha}] = H_\alpha,$$

i.e. $\{H_\alpha, E_\alpha, E_{-\alpha}\}$ spans a subalgebra of $\tilde{\mathfrak{g}}$ isomorphic to $\mathfrak{sl}(2, \mathbb{C})$, with $H_\alpha, E_\alpha, E_{-\alpha}$ playing the roles of (twice) t_3, t_+, t_- respectively.

Proof. Since $E_{\pm\alpha} \in \tilde{\mathfrak{g}}_{\pm\alpha}$, by Definition 37.3 we have $[h, E_{\pm\alpha}] = (\pm\alpha)(h)E_{\pm\alpha}$ for all $h \in \mathfrak{h}$; taking $h = H_\alpha = 2h_\alpha/\langle \alpha, \alpha \rangle \in \mathfrak{h}$ gives

$$[H_\alpha, E_{\pm\alpha}] = \pm \alpha(H_\alpha) E_{\pm\alpha} = \pm \frac{2\alpha(h_\alpha)}{\langle \alpha, \alpha \rangle} E_{\pm\alpha} = \pm \frac{2\langle \alpha, \alpha \rangle}{\langle \alpha, \alpha \rangle} E_{\pm\alpha} = \pm 2E_{\pm\alpha},$$

using $\alpha(h_\alpha) = K(h_\alpha, h_\alpha) = \langle \alpha, \alpha \rangle$. The relation $[E_\alpha, E_{-\alpha}] = H_\alpha$ is exactly the normalisation chosen in Definition 38.1 (consistent by Theorem 37.1, since $[E_\alpha, E_{-\alpha}] \in \tilde{\mathfrak{g}}_0 = \mathfrak{h}$ automatically; only the overall scale of $E_{\pm\alpha}$ needs fixing to hit H_α exactly, which is always possible using non-degeneracy of K on $\tilde{\mathfrak{g}}_\alpha \times \tilde{\mathfrak{g}}_{-\alpha}$, itself a consequence of Theorems 35.3 and 37.2). $\square$

38.2 Root Spaces Are One-Dimensional

Lemma 38.2. Let α be a root of the complex semisimple Lie algebra $\tilde{\mathfrak{g}}$ with Cartan subalgebra $\mathfrak{h}$ ($\ell = \dim \mathfrak{h}$). Then

- (a) the root space $\tilde{\mathfrak{g}}_\alpha$ is one-dimensional, $\tilde{\mathfrak{g}}_\alpha = \mathbb{C}E_\alpha$;
- (b) $k\alpha$ is not a root for any integer $k \geq 2$; in particular 2α is not a root;
- (c) $\alpha/2$ is not a root.

Consequently the only multiples of α that are roots are $\pm\alpha$.

Proof. Let $\{H_\alpha, E_\alpha, E_{-\alpha}\}$ be the Chevalley triple of Theorem 38.1 and consider

$$V := \mathfrak{h} \oplus \bigoplus_{c \neq 0} \tilde{\mathfrak{g}}_{c\alpha},$$

the Cartan subalgebra together with all root spaces along the line through α (the sum runs over those $c \in \mathbb{C} \setminus \{0\}$ for which $c\alpha$ is a root). By Theorem 37.1, $\text{ad}(E_{\pm\alpha})$ maps $\tilde{\mathfrak{g}}_{c\alpha}$ into $\tilde{\mathfrak{g}}_{(c\pm 1)\alpha}$,

and for $h \in \mathfrak{h}$ we have $[E_{\pm\alpha}, h] = \mp\alpha(h)E_{\pm\alpha}$. Hence V is invariant under the triple, i.e. it is a finite-dimensional representation of $\mathfrak{sl}(2, \mathbb{C}) \cong \mathfrak{su}(2)_{\mathbb{C}}$, which by Theorem 36.1 comes from a representation of $\mathfrak{su}(2)$ and is therefore *completely reducible* (Corollary 20.4). Moreover $\text{ad}(H_{\alpha})$ acts on $\tilde{\mathfrak{g}}_{\alpha}$ as multiplication by $c\alpha(H_{\alpha}) = 2c$, and on $\mathfrak{h}$ as 0.

Step 1: counting multiplets of integer spin. In an irreducible representation of spin j the eigenvalues of $\text{ad}(H_{\alpha}) = 2t_3$ are $2m$, $m = -j, \dots, j$. The eigenvalue 0 occurs (exactly once) if and only if j is an integer. So the number of integer-spin irreducible summands of V equals the dimension of the zero-weight space of V , which is $\dim \mathfrak{h} = \ell$.

Step 2: these are all accounted for. Write $\mathfrak{h} = \ker \alpha \oplus \mathbb{C}H_{\alpha}$, where $\ker \alpha = \{h \in \mathfrak{h} : \alpha(h) = 0\}$ has dimension $\ell - 1$ (note $\alpha(H_{\alpha}) = 2 \neq 0$). Every $h \in \ker \alpha$ satisfies $[H_{\alpha}, h] = 0$ and $[E_{\pm\alpha}, h] = 0$, so it spans a spin-0 multiplet: this gives $\ell - 1$ integer-spin multiplets. In addition, $\text{span}\{E_{\alpha}, H_{\alpha}, E_{-\alpha}\}$ is a spin-1 multiplet (the adjoint representation of $\mathfrak{sl}(2)$). That makes ℓ integer-spin multiplets, so by Step 1 there is no further one: the remaining summand V' of V (a complement, which exists by complete reducibility) contains only half-integer spins.

Step 3: conclusions. (a) The weight-2 space of V is $\tilde{\mathfrak{g}}_{\alpha}$, and the only multiplet that contains weight 2 is the spin-1 triple, where it occurs once. So $\dim \tilde{\mathfrak{g}}_{\alpha} = 1$. (b) A root space $\tilde{\mathfrak{g}}_{k\alpha}$, $k \geq 2$ an integer, has even weight $2k \geq 4$, and would require an integer-spin multiplet of spin ≥ 2 ; there is none. So $\tilde{\mathfrak{g}}_{k\alpha} = 0$. (c) Suppose $V' \neq 0$. Every half-integer-spin multiplet contains the state $m = \frac{1}{2}$, of weight 1, which lies in $\tilde{\mathfrak{g}}_{\alpha/2}$; so $\alpha/2$ would be a root. But then $\beta := \alpha/2$ is a root, and (b) applied to β says that $2\beta = \alpha$ is not a root — a contradiction. Hence $V' = 0$ and $\alpha/2$ is not a root. (The argument for β in (b) involves only the triple of β , with $H_{\beta} = 2H_{\alpha}$, so it is not circular.) $\square$

38.3 The α -String Through β , and the Master Formula

Definition 38.2. Fix roots $\alpha, \beta \in \Delta$ with $\beta \neq \pm\alpha$. The α -string through β is the set of all $k \in \mathbb{Z}$ for which $\beta + k\alpha$ is a root (together with $k = 0$, since β itself is a root):

$$\{\beta + k\alpha : -q \leq k \leq p\}, \quad p, q \in \mathbb{Z}_{\geq 0},$$

where we anticipate (Theorem 38.3 below) that this set of roots is automatically an unbroken, consecutive string.

Theorem 38.3. (Master formula.) The α -string through β is unbroken (no “gaps”), and its two endpoints $p, q \in \mathbb{Z}_{\geq 0}$ satisfy

$$p - q = -\frac{2\langle\alpha, \beta\rangle}{\langle\alpha, \alpha\rangle}.$$

In particular, $\frac{2\langle\alpha, \beta\rangle}{\langle\alpha, \alpha\rangle} \in \mathbb{Z}$ for **every** pair of roots α, β .

Proof. Consider the subspace $V := \bigoplus_{k=-q'}^{p'} \tilde{\mathfrak{g}}_{\beta+k\alpha}$, where the sum runs over *all* integers k for which $\beta + k\alpha$ is a root or 0 is not possible here since $\beta \neq \pm\alpha$ rules out $\beta + k\alpha = 0$ for integer k unless β is a multiple of α , excluded by hypothesis; so every $\tilde{\mathfrak{g}}_{\beta+k\alpha}$ appearing is a genuine root space, 1-dimensional by Lemma 38.2, and let $[-q', p']$ (for now just the extremal k 's for which the root space is nonempty — a bounded range, since $\tilde{\mathfrak{g}}$ is finite-dimensional) denote its extent. By Theorem 37.1, $\text{ad}(E_{\alpha})$ maps $\tilde{\mathfrak{g}}_{\beta+k\alpha} \rightarrow \tilde{\mathfrak{g}}_{\beta+(k+1)\alpha}$ and $\text{ad}(E_{-\alpha})$ maps it to $\tilde{\mathfrak{g}}_{\beta+(k-1)\alpha}$, so V is invariant under the Chevalley $\mathfrak{sl}(2, \mathbb{C})$ -triple $\{H_{\alpha}, E_{\alpha}, E_{-\alpha}\}$ of Theorem 38.1: it is a

representation of $\mathfrak{sl}(2, \mathbb{C}) \cong \mathfrak{su}(2)_{\mathbb{C}}$, and by Theorem 36.1 corresponds to an honest representation of $\mathfrak{su}(2)$.

On $\tilde{\mathfrak{g}}_{\beta+k\alpha}$, $\text{ad}(H_\alpha)$ acts by the scalar $(\beta + k\alpha)(H_\alpha) = \beta(H_\alpha) + 2k$ (using $\alpha(H_\alpha) = 2$ from Theorem 38.1, and linearity). Since each individual root space is 1-dimensional, this representation of $\mathfrak{su}(2)_{\mathbb{C}}$ on V has every weight (eigenvalue of $\text{ad}(H_\alpha)$) occurring with multiplicity (at most) one, and it is generated from any single nonzero vector by repeated raising/lowering with $E_{\pm\alpha}$ — both of which are exactly the structural features that force V to be a *single irreducible* $\mathfrak{su}(2)$ representation (a reducible representation would have to repeat a weight, or have raising/lowering fail to connect two given weight spaces, contradicting either the multiplicity-one property or the definition of p', q' as the full extent of nonzero root spaces reached this way). By Part C's classification of finite-dimensional irreducible $\mathfrak{su}(2)$ representations, the $\text{ad}(H_\alpha)$ -eigenvalues on an irreducible representation run, in steps of 2, from a minimum $-2j$ to a maximum $+2j$ (j the spin); matching this to the string's two ends,

$$\beta(H_\alpha) + 2p' = 2j, \quad \beta(H_\alpha) - 2q' = -2j.$$

Adding these two equations gives $2\beta(H_\alpha) + 2(p' - q') = 0$, i.e.

$$p' - q' = -\beta(H_\alpha) = -\frac{2\beta(h_\alpha)}{\langle \alpha, \alpha \rangle} = -\frac{2\langle \alpha, \beta \rangle}{\langle \alpha, \alpha \rangle},$$

using $\beta(h_\alpha) = \langle \beta, \alpha \rangle = \langle \alpha, \beta \rangle$. Moreover, since the eigenvalues of an irreducible $\mathfrak{su}(2)_{\mathbb{C}}$ -representation form a single unbroken run from $-2j$ to $2j$ with no gaps, every intermediate root space $\tilde{\mathfrak{g}}_{\beta+k\alpha}$ ($-q' \leq k \leq p'$) is genuinely nonzero: the string has no gaps, so $p = p'$, $q = q'$ are exactly the string's endpoints as defined in Definition 38.2, completing the proof. $\square$

38.4 The Angle and Length Constraints

Theorem 38.4. (Root angle quantisation.) Let α, β be linearly independent roots, and let θ be the angle between them in the Euclidean space $(\mathfrak{h}_{\mathbb{R}}^*, \langle \cdot, \cdot \rangle)$. Then

$$n_{\beta\alpha} n_{\alpha\beta} = 4 \cos^2 \theta \in \{0, 1, 2, 3\}, \quad \text{where} \quad n_{\beta\alpha} := \frac{2\langle \alpha, \beta \rangle}{\langle \alpha, \alpha \rangle} \in \mathbb{Z}, \quad n_{\alpha\beta} := \frac{2\langle \alpha, \beta \rangle}{\langle \beta, \beta \rangle} \in \mathbb{Z},$$

and only the pairs $(\theta; |\alpha|^2/|\beta|^2)$ in Table 1 below are possible (up to exchanging $\alpha \leftrightarrow \beta$ and $\theta \leftrightarrow \pi - \theta$).

Proof. Both $n_{\beta\alpha}$ and $n_{\alpha\beta}$ are integers by Theorem 38.3. Their product is

$$n_{\beta\alpha} n_{\alpha\beta} = \frac{4\langle \alpha, \beta \rangle^2}{\langle \alpha, \alpha \rangle \langle \beta, \beta \rangle} = 4 \cos^2 \theta,$$

using $\langle \alpha, \beta \rangle = |\alpha||\beta| \cos \theta$ in the Euclidean space $\mathfrak{h}_{\mathbb{R}}^*$. Since $\theta \neq 0, \pi$ (α, β linearly independent), the Cauchy–Schwarz inequality is strict, $\cos^2 \theta < 1$, so $n_{\beta\alpha} n_{\alpha\beta} < 4$; being a non-negative integer (as $4 \cos^2 \theta \geq 0$), it can only be 0, 1, 2, 3. $\square$

Remark. Every entry in Table 1 is realised in nature: two roots at 90° occur whenever a semisimple algebra is (locally) a direct sum, such as $\mathfrak{so}(4) \cong \mathfrak{so}(3) \oplus \mathfrak{so}(3)$ (Example 35.4); $60^\circ/120^\circ$ with equal lengths is exactly the hexagonal $\mathfrak{su}(3)$ root system of Example 37.2; the 45° and 30° cases, with their length ratios $\sqrt{2}$ and $\sqrt{3}$, are exactly what will distinguish the “long” and “short” roots of $B_n/C_n/F_4$ and of G_2 respectively in Section 40.

$n_{\beta\alpha}n_{\alpha\beta}$	θ	$ \alpha ^2/ \beta ^2$	example
0	90°	undetermined	any two orthogonal roots
1	60° (or 120°)	1	any two roots of $\mathfrak{su}(3)$'s hexagon
2	45° (or 135°)	2	long/short roots of B_2, C_2
3	30° (or 150°)	3	long/short roots of G_2

Table 1: The only possible angles between two linearly independent roots, and the corresponding ratio of squared lengths (the longer root's length-squared over the shorter's). This finite list is the geometric heart of the entire classification: only a handful of “shapes” of root system are consistent with it.

39 Simple Roots

The root system Δ typically contains many roots (six for $\mathfrak{su}(3)$, far more for larger algebras) but, remarkably, is entirely determined by a small subset of them: the **simple roots**. This section identifies that subset and proves its key rigidity property, which is what ultimately makes a complete classification possible.

39.1 Positive and Simple Roots

Definition 39.1. Fix a basis $\{h_1, \dots, h_\ell\}$ of $\mathfrak{h}_\mathbb{R}$ chosen so that no root vanishes identically on the “positivity direction” (any generic choice works). Expand each root as $\beta = \sum_i x_i \varpi^i$ for some fixed dual basis $\{\varpi^i\}$ of $\mathfrak{h}_\mathbb{R}^*$, and call β **positive**, written $\beta > 0$, if its first nonvanishing coefficient $x_1, x_2, \dots$ (read in a fixed order) is positive; otherwise $\beta < 0$. We write Δ_+ for the set of positive roots.

Since $-\beta$ has all coefficients negated, exactly one of $\beta, -\beta$ is positive for every root $\beta \neq 0$: $\Delta = \Delta_+ \sqcup (-\Delta_+)$, a disjoint union. One checks directly from the definition that if $\alpha, \beta > 0$ and $\alpha + \beta$ happens to be a root, then $\alpha + \beta > 0$ as well (the first nonvanishing coefficient of a sum of two vectors agreeing in sign on their first nonvanishing coordinates stays of that sign).

Definition 39.2. A positive root $\alpha \in \Delta_+$ is called a **simple root** if it *cannot* be written as $\alpha = \beta + \gamma$ with $\beta, \gamma \in \Delta_+$ both positive roots. We write $\Pi = \{\alpha_1, \dots, \alpha_\ell\} \subset \Delta_+$ for the set of simple roots.

39.2 The Key Rigidity Theorem

Theorem 39.1. Let $\alpha, \beta \in \Pi$ be two **distinct** simple roots. Then:

- (a) $\beta - \alpha$ is *not* a root;
- (b) $\langle \alpha, \beta \rangle \leq 0$.

Proof. (a) Suppose, for contradiction, that $\beta - \alpha$ is a root. By the dichotomy above, either $\beta - \alpha > 0$ or $\alpha - \beta > 0$.

- If $\beta - \alpha > 0$: then $\beta = \alpha + (\beta - \alpha)$ is a sum of two positive roots ($\alpha \in \Pi \subset \Delta_+$, and $\beta - \alpha > 0$ by this case), contradicting that β is simple.
- If $\alpha - \beta > 0$: then $\alpha = \beta + (\alpha - \beta)$ is a sum of two positive roots, contradicting that α is simple.

Either way we reach a contradiction, so $\beta - \alpha$ is not a root.

(b) Consider the α -string through β , $\{\beta - q\alpha, \dots, \beta + p\alpha\}$ (Definition 38.2; applicable since α, β are distinct simple roots, hence linearly independent, so $\beta \neq \pm\alpha$). By part (a), $\beta - \alpha$ (the case $k = -1$) is not a root, so the string cannot extend below $k = 0$: $q = 0$. The master formula (Theorem 38.3) then gives

$$p = p - q = -\frac{2\langle\alpha, \beta\rangle}{\langle\alpha, \alpha\rangle}.$$

Since $p \geq 0$ (it is, by definition, a non-negative integer) and $\langle\alpha, \alpha\rangle > 0$ (positive-definiteness of the induced Euclidean form), this forces $\langle\alpha, \beta\rangle \leq 0$. $\square$

Remark. Combining Theorem 39.1(b) with the angle quantisation of Theorem 38.4: since $\langle\alpha, \beta\rangle \leq 0$ means $\cos\theta \leq 0$, the angle between any two *distinct* simple roots is restricted to

$$\theta \in \{90^\circ, 120^\circ, 135^\circ, 150^\circ\},$$

i.e. simple roots can never make an *acute* angle with one another. This single fact — that the “building blocks” of a root system are mutually non-acute — is the geometric engine behind the entire classification of Section 40: it is essentially the same rigidity that makes the walls of a convex polyhedral cone well defined.

Theorem 39.2. (a) Every positive root $\beta \in \Delta_+$ can be written as a sum of simple roots with **non-negative integer** coefficients, $\beta = \sum_{i=1}^{\ell} n_i \alpha_i$, $n_i \in \mathbb{Z}_{\geq 0}$.

(b) The simple roots $\Pi = \{\alpha_1, \dots, \alpha_\ell\}$ are linearly independent, hence form a **basis** of $\mathfrak{h}_{\mathbb{R}}^*$ (in particular there are exactly $\ell = \text{rank}(\mathfrak{g})$ of them).

Sketch of proof. (a) By strong induction on the **height** $\text{ht}(\beta) := \sum_i x_i$ (sum of coefficients in the expansion in the positivity basis, a positive integer combination that increases whenever a positive root is decomposed). If β is simple, we are done trivially. Otherwise, by Definition 39.2, $\beta = \gamma + \delta$ for some $\gamma, \delta \in \Delta_+$, each of strictly smaller height than β ; by the inductive hypothesis, both γ and δ are non-negative integer combinations of simple roots, hence so is their sum β .

(b) Suppose $\sum_i c_i \alpha_i = 0$ for real c_i , not all zero; split the index set into $I_+ = \{i : c_i > 0\}$ and $I_- = \{i : c_i < 0\}$ (at least one of these is nonempty), and let $v := \sum_{i \in I_+} c_i \alpha_i = -\sum_{i \in I_-} c_i \alpha_i =: w$ (both sides equal, since their difference is $\sum_i c_i \alpha_i = 0$). Then $\langle v, v \rangle = \langle v, w \rangle = \sum_{i \in I_+, j \in I_-} c_i (-c_j) \langle \alpha_i, \alpha_j \rangle$; every term has $c_i > 0$, $-c_j > 0$, and $\langle \alpha_i, \alpha_j \rangle \leq 0$ for $i \neq j$ (Theorem 39.1(b), noting $i \in I_+, j \in I_-$ forces $i \neq j$), so every term is ≤ 0 , giving $\langle v, v \rangle \leq 0$. Since $\langle \cdot, \cdot \rangle$ is positive-definite, this forces $v = w = 0$.

It remains to show $v = \sum_{i \in I_+} c_i \alpha_i = 0$ with all $c_i > 0$ forces $I_+ = \emptyset$ (symmetrically for I_- , using w). The previous trick cannot be repeated verbatim, since Theorem 39.1(b) alone does not rule out a nontrivial *positive* combination of simple roots vanishing — pairwise non-acute angles are consistent in principle with such a relation, and one instead needs the actual linear ordering used to define positivity. Recall from Definition 39.1 that positivity is defined via a fixed “positivity direction”, and every simple root is, by definition, positive; a linear functional realising the same sign choice on the finitely many nonzero vectors $\alpha_1, \dots, \alpha_\ell$ can therefore always be found: take $\lambda \in \mathfrak{h}_{\mathbb{R}}^*$ a sufficiently small, generic perturbation of the vector used to define > 0 in Definition 39.1, close enough that it reproduces the same sign on this finite set of nonzero vectors, so that $\lambda(\alpha_i) > 0$ for *every* $i = 1, \dots, \ell$ simultaneously. Then

$$0 = \lambda(v) = \sum_{i \in I_+} c_i \lambda(\alpha_i),$$

a sum of strictly positive terms (each $c_i > 0$ and $\lambda(\alpha_i) > 0$) if $I_+ \neq \emptyset$ — which is impossible unless $I_+ = \emptyset$. Symmetrically $I_- = \emptyset$. So all $c_i = 0$: the simple roots are linearly independent. Since (by part (a), together with $\Delta = \Delta_+ \sqcup (-\Delta_+)$) every root is an *integer* combination of the α_i , and the roots span $\mathfrak{h}_{\mathbb{R}}^*$ (this is part of the general theory of Cartan subalgebras, guaranteeing $\mathfrak{h}_{\mathbb{R}}^*$ has no “room to spare” beyond what the roots already see), the linearly independent set Π must in fact span all of $\mathfrak{h}_{\mathbb{R}}^*$, hence is a basis. $\square$

40 The Cartan Matrix, Dynkin Diagrams, and the Classification

We have now reduced an entire semisimple Lie algebra — an object with potentially hundreds of generators — to a finite, purely combinatorial gadget: the pairwise relations among $\ell = \text{rank}(\tilde{\mathfrak{g}})$ simple roots. This section packages that data into the **Cartan matrix** and its picture, the **Dynkin diagram**, and states the classification theorem that completely enumerates the possibilities.

40.1 The Cartan Matrix

Definition 40.1. Let $\Pi = \{\alpha_1, \dots, \alpha_\ell\}$ be the simple roots of $\tilde{\mathfrak{g}}$ (Definition 39.2). The **Cartan matrix** of $\tilde{\mathfrak{g}}$ is the $\ell \times \ell$ matrix with entries

$$A_{jk} := \frac{2\langle \alpha_j, \alpha_k \rangle}{\langle \alpha_k, \alpha_k \rangle} \quad (j, k = 1, \dots, \ell).$$

(Convention note: the literature is split on whether to normalise by the row or the column root’s length; we normalise by the *column*, $\langle \alpha_k, \alpha_k \rangle$, matching the convention used for the classical Lie algebras in Part E.)

By Theorem 38.3, $A_{jk} = -p_{jk}$, minus the length of the α_k -string through α_j counted *upward* from α_j (i.e. the largest p such that $\alpha_j + p\alpha_k$ is still a root) — since α_j, α_k are both simple, the string’s lower end vanishes, $q = 0$, by Theorem 39.1(a). The Cartan matrix therefore collects exactly the data needed to reconstruct the whole root system by repeatedly adding simple roots, as we will do explicitly in Example 40.2 below.

Theorem 40.1. The Cartan matrix has the following properties:

- (a) $A_{jj} = 2$ for every j ;
- (b) $A_{jk} \in \{0, -1, -2, -3\}$ for $j \neq k$, and $A_{jk} = 0 \iff A_{kj} = 0$;
- (c) $A_{jk}A_{kj} = 4\cos^2\theta_{jk} \in \{0, 1, 2, 3\}$ for $j \neq k$, where θ_{jk} is the angle between α_j, α_k ;
- (d) $\det A \neq 0$ (in fact $\det A > 0$).

Proof. (a) is immediate from the definition. (b): $A_{jk} = n_{\alpha_j, \alpha_k}$ is an integer by Theorem 38.3 (with the roles $\alpha \rightarrow \alpha_k, \beta \rightarrow \alpha_j$), and $A_{jk} \leq 0$ by Theorem 39.1(b); together with (c), $|A_{jk}| \leq 3$, giving $A_{jk} \in \{0, -1, -2, -3\}$; and $A_{jk} = 0$ iff $\langle \alpha_j, \alpha_k \rangle = 0$ iff $A_{kj} = 0$, since both are proportional to the same inner product. (c) is Theorem 38.4 applied to α_j, α_k (note $n_{\beta\alpha}$ there is exactly A_{jk} up to matching indices). (d) Write $A_{jk} = S_{jk}D_{kk}$ where $S_{jk} := \langle \alpha_j, \alpha_k \rangle$ (the positive-definite Gram matrix of the linearly independent vectors α_j , Theorem 39.2(b), so $\det S > 0$) and $D := \text{diag}(2/\langle \alpha_1, \alpha_1 \rangle, \dots)$ (diagonal, strictly positive entries, so $\det D > 0$), i.e. $A = SD$ (each *column* k of S rescaled by D_{kk}); then $\det A = \det S \cdot \det D > 0$. $\square$

40.2 Dynkin Diagrams

Definition 40.2. The **Dynkin diagram** of $\tilde{\mathfrak{g}}$ is the graph with one node for each simple root α_j , in which nodes $j \neq k$ are joined by $A_{jk}A_{kj} \in \{0, 1, 2, 3\}$ edges; whenever they are joined by 2 or 3 edges (unequal root lengths, $A_{jk} \neq A_{kj}$), an arrow is drawn on the edges pointing from the **long** root to the **short** root (the short root is the node k for which $|A_{jk}|$, not $|A_{kj}|$, is the *larger* of the two off-diagonal entries: since $A_{jk} = 2\langle\alpha_j, \alpha_k\rangle/\langle\alpha_k, \alpha_k\rangle$ by Definition 40.1, a *smaller* $\langle\alpha_k, \alpha_k\rangle$ — i.e. a shorter root k — makes $|A_{jk}|$ *larger*).

The Dynkin diagram, unlike the (in general asymmetric) Cartan matrix, is manifestly the same object seen from either end of an edge, and it encodes exactly the same information: the number of edges is the symmetric quantity $A_{jk}A_{kj}$, and the arrow (needed only when lengths differ) records which of the two asymmetric entries is -1 .

Theorem 40.2. (The classification theorem.) The Dynkin diagram of a simple complex Lie algebra is connected, and is necessarily one of the following: the four infinite families

$$A_n \ (n \geq 1), \quad B_n \ (n \geq 2), \quad C_n \ (n \geq 3), \quad D_n \ (n \geq 4),$$

or one of the five exceptional diagrams

$$E_6, \quad E_7, \quad E_8, \quad F_4, \quad G_2.$$

(The subscript is the rank ℓ , i.e. the number of nodes.) Every one of these diagrams is realised by exactly one simple complex Lie algebra, and conversely every simple complex Lie algebra has one of these diagrams: the list is a complete and non-redundant classification.

Remark. We do not prove Theorem 40.2 here — doing so rigorously (showing that no *other* configuration of unit and $\sqrt{2}$ -/ $\sqrt{3}$ -scaled vectors, pairwise at the angles of Table 1, can close up into a consistent, linearly independent, connected diagram) is a beautiful but somewhat lengthy exercise in Euclidean geometry, due in essence to Killing and made fully rigorous by Cartan; see Humphreys or Fulton–Harris for the complete argument. What we *have* proved — the integrality of the Cartan integers (Theorem 38.3), the restriction to the four angles of Table 1, and the non-acuteness of simple roots (Theorem 39.1) — already constitutes the hard core of the theorem’s proof; what remains is a finite (if intricate) case-by-case elimination of every diagram *not* on the list above.

Physical perspective

The four infinite families correspond to the classical matrix groups familiar from Part B: $A_n = \mathfrak{sl}(n+1, \mathbb{C})$ (i.e. $\mathfrak{su}(n+1)$ is its compact real form), $B_n = \mathfrak{so}(2n+1, \mathbb{C})$, $C_n = \mathfrak{sp}(2n, \mathbb{C})$, $D_n = \mathfrak{so}(2n, \mathbb{C})$. The five exceptional algebras have no classical realisation as matrices of a simple pattern; G_2 , the smallest, has appeared in physics as a candidate grand-unified gauge group and in the study of certain string-theory compactifications (G_2 -holonomy manifolds), while E_6, E_7, E_8 appear in grand unification and in heterotic string theory.

40.3 Small Examples

Example 40.1. Cartan matrices of low rank.

$$A_1 : \bullet \quad (2), \quad A_2 : \bullet \text{---} \bullet \quad \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix},$$

$$A_3 : \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}, \quad B_2 : \begin{pmatrix} 2 & -2 \\ -1 & 2 \end{pmatrix}, \quad C_2 : \begin{pmatrix} 2 & -1 \\ -2 & 2 \end{pmatrix}, \quad G_2 : \begin{pmatrix} 2 & -3 \\ -1 & 2 \end{pmatrix}.$$

Note B_2 and C_2 are transposes of each other, i.e. describe the *same* Dynkin diagram with the arrow reversed — indeed $B_2 \cong C_2$ as abstract Lie algebras (both $\cong \mathfrak{so}(5, \mathbb{C}) \cong \mathfrak{sp}(4, \mathbb{C})$), the first “accidental” isomorphism in the classification, exactly parallel to $\mathfrak{so}(3) \cong \mathfrak{su}(2)$ except that here it identifies two different classical families at rank 2. (These match the explicit root computations of Theorem 45.3 for B_n and its analogue for C_n in Part E, specialised to $n = 2$.)

Example 40.2. Reconstructing all positive roots of A_3 from its Cartan matrix. Take three simple roots $\alpha_1, \alpha_2, \alpha_3$ with $A_{12} = A_{21} = A_{23} = A_{32} = -1$ and $A_{13} = A_{31} = 0$ (equal lengths throughout, so we may normalise $\langle \alpha_i, \alpha_i \rangle = 2$ for all i , giving simply $\langle \alpha_i, \alpha_j \rangle = A_{ij}$). We build up the positive roots level by level, where the **level** (height) of $\sum n_i \alpha_i$ is $\sum n_i$ (Theorem 39.2(a)).

- *Level 1:* $\alpha_1, \alpha_2, \alpha_3$ (the simple roots themselves).
- *Level 2:* is $\alpha_i + \alpha_j$ ($i < j$) a root? Using the α_1 -string through α_2 : $q = 0$ (Theorem 39.1(a)), and $p = -A_{12} = 1$, so $\alpha_2 + \alpha_1$ is a root (and $\alpha_2 + 2\alpha_1$ is not, since the string has length $p + q + 1 = 2$). Likewise $\alpha_2 + \alpha_3$ is a root. But $\alpha_1 + \alpha_3$: the α_1 -string through α_3 has $p = -A_{13} = 0$, so $\alpha_1 + \alpha_3$ is *not* a root. Level 2 positive roots: $\alpha_1 + \alpha_2$, $\alpha_2 + \alpha_3$.
- *Level 3:* Consider $\alpha_1 + \alpha_2 + \alpha_3$. Is it obtainable? Adding α_3 to $\alpha_1 + \alpha_2$: the α_3 -string through $\alpha_1 + \alpha_2$ needs $\langle \alpha_3, \alpha_1 + \alpha_2 \rangle = \langle \alpha_3, \alpha_1 \rangle + \langle \alpha_3, \alpha_2 \rangle = 0 + (-1) = -1 \neq 0$; since the pairing is nonzero (and, working through the string logic as above) $\alpha_1 + \alpha_2 + \alpha_3$ is a root, obtained equally well as $\alpha_1 + (\alpha_2 + \alpha_3)$. Level 3 positive roots: $\alpha_1 + \alpha_2 + \alpha_3$ (only).
- *Level 4:* $\alpha_1 + 2\alpha_2 + \alpha_3$ would be the next candidate, obtained from the level-3 root $\alpha_1 + \alpha_2 + \alpha_3$ by adding another α_2 . Consider the α_2 -string through $\alpha_1 + \alpha_2 + \alpha_3$: its lower end already fails, $q = 0$, since $(\alpha_1 + \alpha_2 + \alpha_3) - \alpha_2 = \alpha_1 + \alpha_3$ is *not* a root (Level 2, above); and $\langle \alpha_2, \alpha_1 + \alpha_2 + \alpha_3 \rangle = A_{21} + A_{22} + A_{23} = -1 + 2 - 1 = 0$, so the master formula gives $p - q = -2(0)/2 = 0$, i.e. $p = 0$ too: the string does not extend upward either. Hence $\alpha_1 + 2\alpha_2 + \alpha_3$ is *not* a root, and the construction terminates.

In total, A_3 has $3 + 2 + 1 = 6$ positive roots (and 12 roots altogether), matching $\dim \mathfrak{sl}(4, \mathbb{C}) = 4^2 - 1 = 15 = \underbrace{3}_{\text{rank}} + \underbrace{12}_{\text{roots}}$, exactly the dimension count $\dim \tilde{\mathfrak{g}} = \ell + |\Delta|$ implied by the root space decomposition (Definition 37.3) together with each root space being 1-dimensional.

41 Worked Example: The Root System of $SU(3)$

We close by assembling every piece of machinery developed in this chapter on the single example that will recur constantly in the applications to follow: $\mathfrak{su}(3)_{\mathbb{C}}$, of type A_2 .

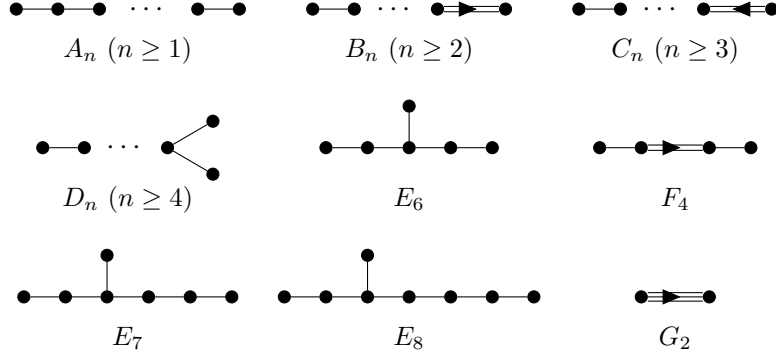


Figure 15: The Dynkin diagrams of the four infinite families and five exceptional simple complex Lie algebras (Theorem 40.2). Each node is a simple root; the number of lines joining two nodes is $A_{jk}A_{kj} \in \{1, 2, 3\}$ (no line at all when $A_{jk}A_{kj} = 0$, i.e. orthogonal roots); an arrow on a double or triple bond points from the long root to the short root.

Example 41.1. From the Cartan matrix to the root diagram. A_2 has rank $\ell = 2$ and Cartan matrix $\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$ (Example 40.1): both off-diagonal entries equal -1 , so by Theorem 40.1(c) the two simple roots α_1, α_2 have *equal length* and $A_{12}A_{21} = 1$ places them at 120° (Table 1). Level by level (as in Example 40.2):

- *Level 1:* α_1, α_2 .
- *Level 2:* the α_1 -string through α_2 has $q = 0$ (Theorem 39.1) and $p = -A_{12} = 1$, so $\alpha_1 + \alpha_2$ is a root; the string terminates there ($p + q + 1 = 2$ states), so $2\alpha_1 + \alpha_2$ is *not* a root, and the construction stops.

This gives exactly $2 + 1 = 3$ positive roots, hence 6 roots in total, plus rank 2: $\dim \mathfrak{su}(3)_\mathbb{C} = 2 + 6 = 8$, correctly reproducing $N^2 - 1 = 8$ for $\mathfrak{su}(N)$ at $N = 3$. Placing α_1 along the horizontal axis and α_2 at 120° to it, with $|\alpha_1| = |\alpha_2|$, the six roots $\{\pm\alpha_1, \pm\alpha_2, \pm(\alpha_1 + \alpha_2)\}$ automatically fall on the vertices of a regular hexagon — exactly the root system found by brute-force computation with the Gell-Mann matrices in Example 37.2, now derived from nothing but the Cartan matrix. Fig. 16 shows the result, together with the ladder operators $E_{\pm\alpha}$ attached to each root by Definition 38.1.

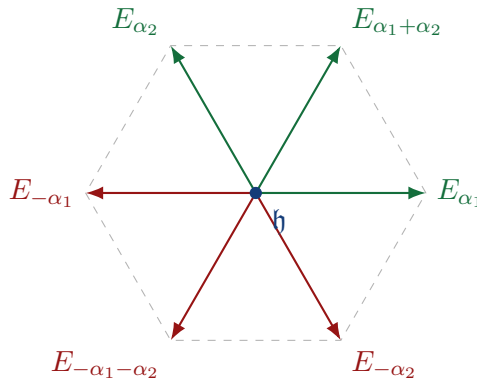


Figure 16: The root system of $\mathfrak{su}(3)_\mathbb{C} \cong A_2$: six roots at the vertices of a regular hexagon in $\mathfrak{h}_\mathbb{R}^* \cong \mathbb{R}^2$ (Examples 37.2 and 41.1). The two simple roots α_1, α_2 and the third positive root $\alpha_1 + \alpha_2$ are shown in green with their associated raising operators E_α ; the three negative roots, in red, carry the corresponding lowering operators $E_{-\alpha}$.

Part V

Weights, the Classical Lie Algebras, and the Lorentz Group

42 Introduction: Completing the Programme

Part D reduced the classification of semisimple Lie algebras to a piece of pure combinatorics: choose a Cartan subalgebra, extract the root system, and read off a Dynkin diagram from a short list of rigidity constraints. We ended that chapter with a promise: the same root-space machinery that classifies the *algebras* also classifies their *representations*. This chapter makes good on that promise, and then puts the machinery to work.

We proceed in four stages. First (Section 43) we extend the notion of a root — an eigenvalue of $\text{ad}(\mathfrak{h})$ on the algebra itself — to the notion of a **weight**, an eigenvalue of $\mathfrak{h}$ on an arbitrary representation. The set of weights of a representation turns out to be built from a single distinguished weight, the **highest weight**, exactly as the finite ladder of $\mathfrak{su}(2)$ states in Part C was built by lowering from a single top state. Along the way we meet the **Weyl group**, the symmetry group of the root system generated by reflections, and we explain why every finite-dimensional representation of a compact Lie algebra can be chosen unitary. Second (Section 44) we carry this machinery through in complete detail for $\mathfrak{su}(3)$, recovering the classification of representations by a pair of non-negative integers (Dynkin labels) that physicists know as the “ $SU(3)$ multiplets” of the eightfold way. Third (Section 45) we return to the classification theorem of Part D and close the loop in the other direction: rather than abstractly enumerating Dynkin diagrams, we construct the four infinite families of *classical* Lie algebras — $\mathfrak{su}(N)$, $\mathfrak{so}(2n+1)$, $\mathfrak{usp}(2n)$, $\mathfrak{so}(2n)$ — explicitly as matrix algebras, exhibit a Cartan subalgebra and root system for each by hand, and verify that they reproduce exactly the diagrams A_{N-1} , B_n , C_n , D_n of the classification theorem; we also record the natural invariant (Casimir) tensors these algebras carry. Finally (Section 47), as a physical capstone to the entire course, we apply everything we have built to the Lie algebra that matters most in relativistic physics: the Lorentz algebra $\mathfrak{so}(3,1)$. Its complexification splits as $\mathfrak{su}(2)_{\mathbb{C}} \oplus \mathfrak{su}(2)_{\mathbb{C}}$ — a fact we proved abstractly in Part D — and we now use that splitting to classify *all* finite-dimensional representations of the Lorentz group by a pair of spins (j, j') , recovering scalars, the left- and right-handed Weyl spinors, vectors, and so on, as the simplest members of an infinite family.

Physical perspective

This chapter is where the abstract classification of Part D turns into the representation theory physicists use daily: the weight diagrams of $\mathfrak{su}(3)$ are the particle multiplets of the eightfold way, the classical algebras $\mathfrak{su}(N)$, $\mathfrak{so}(N)$, $Sp(2n)$ are exactly the gauge groups available for building theories, and the representation theory of the Lorentz algebra is the origin of the classification of relativistic fields into scalars, spinors, vectors, and tensors.

43 Weights, the Weyl Group, and the Highest-Weight Classification

43.1 Weights of a Representation

Let $\mathfrak{g}$ be a semisimple Lie algebra with Cartan subalgebra $\mathfrak{h} = \text{span}\{H_1, \dots, H_\ell\}$ ($\ell = \text{rank } \mathfrak{g}$) and root system $\Delta \subset \mathfrak{h}^*$, as constructed in Part D. Let R be a finite-dimensional representation

of $\mathfrak{g}$ on a vector space V . Since the elements of $\mathfrak{h}$ commute among themselves, $\{R(H_i)\}_{i=1}^\ell$ is a family of commuting operators on V and can therefore be diagonalised simultaneously.

Definition 43.1. A **weight** of the representation R is a vector $\lambda = (\lambda_1, \dots, \lambda_\ell) \in \mathfrak{h}^*$ such that there exists a nonzero $\psi_\lambda \in V$ (a **weight vector**) with

$$R(H_i) \psi_\lambda = \lambda_i \psi_\lambda \quad (i = 1, \dots, \ell).$$

The subspace of all such ψ_λ for fixed λ is the **weight space** V_λ ; its dimension is the **multiplicity** of λ .

This is the direct generalisation of the eigenvalue m of t_3 labelling the states $|j, m\rangle$ of an $\mathfrak{su}(2)$ -multiplet in Part C, and of the roots themselves, which are simply the weights of the *adjoint* representation ($V = \mathfrak{g}$, $R = \text{ad}$), with the zero weight (multiplicity ℓ , spanned by $\mathfrak{h}$ itself) adjoined.

Exactly as the raising and lowering operators E_α shift $\mathfrak{su}(2)$ -weights by roots, they shift weights of any representation by roots:

Proposition 43.1. If $\psi_\lambda \in V_\lambda$ and $\alpha \in \Delta$, then $R(E_\alpha)\psi_\lambda$ is either zero or a weight vector of weight $\lambda + \alpha$.

Proof. For $H \in \mathfrak{h}$, using $[H, E_\alpha] = \alpha(H)E_\alpha$ from the root-space decomposition (Part D),

$$R(H) R(E_\alpha)\psi_\lambda = R(E_\alpha)R(H)\psi_\lambda + R([H, E_\alpha])\psi_\lambda = (\lambda(H) + \alpha(H)) R(E_\alpha)\psi_\lambda,$$

so $R(E_\alpha)\psi_\lambda$ is either 0 or an eigenvector of every $R(H)$ with eigenvalue $\lambda + \alpha$. $\square$

Restricting attention to a single root α , the subalgebra $\{E_\alpha, E_{-\alpha}, H_\alpha := [E_\alpha, E_{-\alpha}]\}$ is (after rescaling) an $\mathfrak{su}(2)$ subalgebra of $\mathfrak{g}$, exactly as in the root-string analysis of Part D. Every representation V of $\mathfrak{g}$ restricts to a representation of this $\mathfrak{su}(2)$, and the $\mathfrak{su}(2)$ representation theory of Part C applies verbatim: the eigenvalues of H_α on V are integers, and the string of weights $\lambda, \lambda - \alpha, \dots, \lambda - p\alpha$ obtained by repeated lowering (with $\lambda + \alpha, \dots$ not being weights, i.e. p maximal) satisfies the same root-string relation as for roots themselves. This immediately gives the analogue of the Master Formula of Part D:

Theorem 43.2 (Weight quantisation). For every weight λ of a finite-dimensional representation and every root α ,

$$\frac{2\langle\lambda, \alpha\rangle}{\langle\alpha, \alpha\rangle} \in \mathbb{Z}$$

where $\langle\cdot, \cdot\rangle$ is the Euclidean form on $\mathfrak{h}^*$ induced by the Killing form, as in Part D.

Proof. The quantity $2\langle\lambda, \alpha\rangle/\langle\alpha, \alpha\rangle$ is precisely the eigenvalue of H_α (the Cartan generator of the $\mathfrak{su}(2)$ triple associated to α) on the weight vector ψ_λ , normalised as in Part D so that this eigenvalue equals $2j_3$ for an $\mathfrak{su}(2)$ -multiplet of definite total spin content along α . Since eigenvalues of $2 \times$ (the third generator) in any finite-dimensional representation of $\mathfrak{su}(2)$ are integers (Part C), the claim follows. $\square$

In particular, taking $\lambda = \beta$ a simple root recovers the integrality statement of Part D as a special case: roots are weights of the adjoint representation.

43.2 The Weyl Group

The quantisation condition of Theorem 43.2 has a beautiful geometric consequence: it says that reflecting a weight in the hyperplane perpendicular to a root sends the weight set of a representation to itself.

Definition 43.2. For a root $\alpha \in \Delta$, the **Weyl reflection** $S_\alpha : \mathfrak{h}^* \rightarrow \mathfrak{h}^*$ is

$$S_\alpha(\beta) := \beta - \frac{2\langle \beta, \alpha \rangle}{\langle \alpha, \alpha \rangle} \alpha.$$

The **Weyl group** W of $\mathfrak{g}$ is the (finite) group of orthogonal transformations of $\mathfrak{h}^*$ generated by all the reflections S_α , $\alpha \in \Delta$.

Geometrically, S_α is the reflection in the hyperplane orthogonal to α : it fixes every vector perpendicular to α and sends $\alpha \mapsto -\alpha$, which is easily checked by substituting $\beta = \alpha$: $S_\alpha(\alpha) = \alpha - 2\alpha = -\alpha$. Because the integer in Theorem 43.2 is exactly the coefficient appearing in Definition 43.2, $S_\alpha(\lambda)$ is again a weight whenever λ is: the weight diagram of *any* representation is invariant under the full Weyl group, and in particular under all reflections in the simple roots, which already generate W (Part D showed every root is a W -image of a simple root, by the analogous argument applied to Δ itself).

Remark. This is an extremely efficient way to reconstruct or check a full weight diagram: start from one known weight (e.g. the highest weight, Definition 43.5 below), lower repeatedly by simple roots to generate new weights, and use the Weyl reflections as a consistency check — the resulting diagram must look identical from the perspective of every simple root, since S_{α_i} must map the diagram to itself.

43.3 Fundamental Weights and Highest Weight

Just as the simple roots $\{\alpha_1, \dots, \alpha_\ell\}$ form a basis of $\mathfrak{h}^*$ (Part D), there is a second, dual basis adapted to the integrality condition of Theorem 43.2.

Definition 43.3. The **fundamental weights** $\tilde{\lambda}_1, \dots, \tilde{\lambda}_\ell \in \mathfrak{h}^*$ are defined by

$$\frac{2\langle \tilde{\lambda}_i, \alpha_j \rangle}{\langle \alpha_j, \alpha_j \rangle} = \delta_{ij} \quad (i, j = 1, \dots, \ell).$$

By construction, $\tilde{\lambda}_i$ satisfies the quantisation condition of Theorem 43.2 with respect to every simple root (and hence, one can show, with respect to every root), so it is automatically an admissible weight. Any weight λ can be expanded uniquely in this basis,

$$\lambda = \sum_{i=1}^{\ell} n_i \tilde{\lambda}_i,$$

The coefficients are $n_j = 2\langle \lambda, \alpha_j \rangle / \langle \alpha_j, \alpha_j \rangle$, and Theorem 43.2, applied with $\alpha = \alpha_j$, shows immediately that each n_j is an integer.

Definition 43.4. The integers $(n_1, \dots, n_\ell)$ are the **Dynkin labels** of the weight $\lambda = \sum_i n_i \tilde{\lambda}_i$.

Definition 43.5. A weight λ of a representation R is a **highest weight** if $R(E_{\alpha_i})\psi_\lambda = 0$ for every simple root α_i , i.e. λ cannot be raised by any simple raising operator.

Since every positive root is a non-negative integer combination of simple roots (Part D), a weight annihilated by all the simple raising operators is automatically annihilated by *every* raising operator E_α , $\alpha > 0$: it is the top of the entire representation, exactly as $|j, j\rangle$ was the top of an $\mathfrak{su}(2)$ -multiplet in Part C. This is the content of the fundamental theorem of representation theory for semisimple Lie algebras, which we quote without proof (it is the natural but technical generalisation of the construction of $\mathfrak{su}(2)$ -multiplets by repeated lowering from the top state):

Theorem 43.3. (Highest weight theorem.) The finite-dimensional irreducible representations of a semisimple Lie algebra $\mathfrak{g}$ are in bijection with the highest weights $\lambda = \sum_i n_i \tilde{\lambda}_i$ having *non-negative integer* Dynkin labels $n_i \geq 0$. Every weight of the representation is obtained from λ by subtracting a non-negative integer combination of simple roots, and the full weight diagram is generated from λ by repeated lowering together with Weyl-group symmetry.

This is precisely the generalisation, promised at the end of Part D, of the $\mathfrak{su}(2)$ classification of Part C by a single non-negative integer (or half-integer spin j , i.e. Dynkin label $n = 2j$): for a general semisimple $\mathfrak{g}$ of rank ℓ , representations are labelled by ℓ non-negative integers, one Dynkin label per node of the Dynkin diagram.

43.4 Reality, Compactness, and Weyl's Unitarity Theorem

Everything above is purely algebraic (it uses only the complexified algebra $\tilde{\mathfrak{g}} = \mathfrak{g}_\mathbb{C}$ of Part D). To connect with physics — where representations must be *unitary* on a Hilbert space of states — we need to reintroduce reality.

Definition 43.6. A real Lie algebra $\mathfrak{g}_0$ is a **compact real form** of a complex semisimple Lie algebra $\tilde{\mathfrak{g}}$ if $(\mathfrak{g}_0)_\mathbb{C} = \tilde{\mathfrak{g}}$ and the Killing form of $\mathfrak{g}_0$ is negative definite.

Recall from Part D that $\mathfrak{su}(2)$ is the compact real form of $\mathfrak{sl}(2, \mathbb{C})$, while $\mathfrak{sl}(2, \mathbb{R})$ is a second, non-compact real form of the same complexification; more generally $\mathfrak{su}(N)$, $\mathfrak{so}(N)$, $\mathfrak{usp}(2n)$ (all introduced below) are the compact real forms of their complexifications. On a compact real form one may choose a basis $\{t_a\}$ in which the structure constants f_{ab}^c are completely antisymmetric and real, and the generators are represented by *anti-Hermitian* matrices, $t_a^\dagger = -t_a$; equivalently, writing $T_a = i t_a$ as physicists do, the T_a are Hermitian, $T_a^\dagger = T_a$.

Theorem 43.4. (Weyl's unitarity theorem.) Every finite-dimensional representation of a compact real form $\mathfrak{g}_0$ can be chosen unitary: there is an inner product on the representation space for which $R(t_a)$ is anti-Hermitian (equivalently, $R(T_a) = R(it_a)$ is Hermitian) for every generator.

We will not reprove this here (it follows from the existence of an invariant, positive-definite averaging over the corresponding compact group, generalising the finite-group averaging argument); the point that matters for what follows is its immediate corollary for the Cartan subalgebra.

Remark. Since the Cartan generators H_i can be taken Hermitian (they lie in $it_a \mapsto H_i$, real linear combinations of the anti-Hermitian t_a 's), their eigenvalues — the weights λ_i —

are automatically *real* numbers, not merely elements of an abstract dual space. This is why weight diagrams can be drawn as honest configurations of points in a real Euclidean space $\mathbb{R}^\ell$, and it is the representation-theoretic reason (rather than a coincidence) that all the root and weight diagrams drawn in Part D and below have real coordinates.

43.5 Complex Conjugate Representations

If R is a representation with Hermitian generators $R(T_a)$, its **complex conjugate representation** $\bar{R}$ is defined by $\bar{R}(T_a) := -R(T_a)^*$ (the minus sign is required so that $\bar{R}$ again satisfies the same, real-structure-constant commutation relations). On the Cartan generators, which are diagonal and real in a weight basis, complex conjugation simply flips the sign of every eigenvalue:

Proposition 43.5. If λ is a weight of R , then $-\lambda$ is a weight of $\bar{R}$. In particular the highest weight of $\bar{R}$ is $-w_0(\lambda)$, where $w_0 \in W$ is the Weyl group element sending the negative-weight chamber to the positive one.

A representation is called **self-conjugate** if $\bar{R} \cong R$, which happens exactly when the weight diagram of R is symmetric under $\lambda \mapsto -\lambda$. Self-conjugacy alone does not distinguish two inequivalent situations, told apart by the Frobenius–Schur indicator: R is **real** if it admits an invariant *symmetric* bilinear form (equivalently, can be realised by real matrices in a suitable basis), and **pseudoreal** (or quaternionic) if instead it admits only an invariant *antisymmetric* one (equivalently, is self-conjugate but genuinely complex, with no real form) — the fundamental doublet of $\mathfrak{su}(2)$ is the standard example of the latter: self-conjugate (its weights $\pm \frac{1}{2}$ are symmetric), yet not real, since the antisymmetric $\varepsilon_{\alpha\beta}$ of Section 47.4 realises the equivalence $\bar{R} \cong R$, not a symmetric form.

Example 43.1. The adjoint representation is always self-conjugate: its weights are the roots together with ℓ copies of the zero weight, and the root system Δ is symmetric under $\alpha \mapsto -\alpha$ by construction (Part D). This matches Section 43.4: the adjoint representation of a compact real form is literally the (real) vector space $\mathfrak{g}_0$ acted on by itself, so it must be a real representation in the representation-theoretic sense too.

44 Worked Example: The Representations of $\mathfrak{su}(3)$

We now carry the machinery of Section 43 through explicitly for $\mathfrak{su}(3)$, using the Cartan–Weyl basis and hexagonal root system A_2 constructed in Part D.

44.1 Setup: Simple Roots and Fundamental Weights

Recall from Part D that $\mathfrak{su}(3)_{\mathbb{C}}$ has rank $\ell = 2$, Cartan matrix $\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$, and simple roots α_1, α_2 of equal length at 120° ; the full root system is the hexagon $\{\pm\alpha_1, \pm\alpha_2, \pm(\alpha_1 + \alpha_2)\}$. Normalising $\langle \alpha_i, \alpha_i \rangle = 1$ and placing

$$\alpha_1 = (1, 0), \quad \alpha_2 = \left(-\frac{1}{2}, \frac{\sqrt{3}}{2}\right), \quad \alpha_1 + \alpha_2 = \left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right)$$

in $\mathfrak{h}_{\mathbb{R}}^* \cong \mathbb{R}^2$, Definition 43.3 gives the two fundamental weights (solve $2\langle \tilde{\lambda}_i, \alpha_j \rangle = \delta_{ij}$):

$$\tilde{\lambda}_1 = \left(\frac{1}{2}, \frac{1}{2\sqrt{3}}\right), \quad \tilde{\lambda}_2 = \left(0, \frac{1}{\sqrt{3}}\right),$$

with $|\tilde{\lambda}_1| = |\tilde{\lambda}_2| = 1/\sqrt{3}$ and the angle between them 60° (not a symmetry of the hexagon by a Weyl-group rotation, which for A_2 moves roots in multiples of 60° about the origin but does not fix the length while permuting $\tilde{\lambda}_1, \tilde{\lambda}_2$): rather, $\tilde{\lambda}_1$ and $\tilde{\lambda}_2$ are exchanged by $-w_0$, the composition of the longest Weyl group element w_0 with the order-2 automorphism of the A_2 Dynkin diagram that swaps its two nodes (the group-theoretic operation underlying charge conjugation $\mathbf{3} \leftrightarrow \bar{\mathbf{3}}$ below).

Remark. In physics language, $\mathfrak{h}^* \cong \mathbb{R}^2$ is spanned by (a rescaling of) the third component of isospin I_3 and the hypercharge Y ; the two axes of Fig. 17 below are exactly the (I_3, Y) plane in which every $SU(3)$ -flavour multiplet is conventionally drawn.

44.2 The Fundamental and Antifundamental Representations

The representation with highest weight $\tilde{\lambda}_1$ (Dynkin label $(1, 0)$) is the **fundamental representation $\mathbf{3}$** . Starting from $\tilde{\lambda}_1$ and lowering by simple roots according to Theorem 43.3: lowering once by α_1 gives $\tilde{\lambda}_1 - \alpha_1 = (-\frac{1}{2}, \frac{1}{2\sqrt{3}})$, which one checks has Dynkin label $(-1, 1)$, so it can be lowered once more, now by α_2 , giving $\tilde{\lambda}_1 - \alpha_1 - \alpha_2 = (0, -\frac{1}{\sqrt{3}}) = -\tilde{\lambda}_2$. No further lowering is possible (both remaining Dynkin labels are already ≤ 0), so $\mathbf{3}$ has exactly three weights, each of multiplicity one:

$$\tilde{\lambda}_1, \quad \tilde{\lambda}_1 - \alpha_1, \quad -\tilde{\lambda}_2 = \tilde{\lambda}_1 - \alpha_1 - \alpha_2,$$

forming an equilateral triangle, consistent with $\dim \mathbf{3} = 3$. By Proposition 43.5, the conjugate representation $\bar{\mathbf{3}}$ (highest weight $\tilde{\lambda}_2$, Dynkin label $(0, 1)$) has the negatives of these as its weights, an equilateral triangle pointing the opposite way.

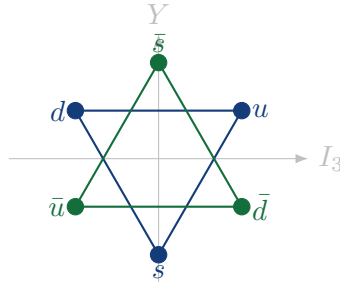


Figure 17: Weight diagrams of the fundamental $\mathbf{3}$ (blue, highest weight $\tilde{\lambda}_1$, labelled by the three light-quark flavours u, d, s) and antifundamental $\bar{\mathbf{3}}$ (green, highest weight $\tilde{\lambda}_2$) representations of $\mathfrak{su}(3)$, drawn in the (I_3, Y) plane.

44.3 The Adjoint Representation

The adjoint representation has highest weight $\alpha_1 + \alpha_2$ (the highest root), Dynkin label $(1, 1)$, and dimension 8. Its eight weights are the six nonzero roots of the hexagonal root system (Part D) together with the zero weight, which occurs with multiplicity $2 = \text{rank } \mathfrak{su}(3)$ (spanned by $\mathfrak{h}$ itself, Definition 43.1). Since $(1, 1)$ is of the self-conjugate form (n, n) , the adjoint representation is self-conjugate, and by Example 43.1 (the compact-real-form argument, not the weight symmetry alone, which only ever gives self-conjugacy) it is in fact real rather than merely pseudoreal.

Remark. More generally, for $\mathfrak{su}(3)$ every representation with Dynkin label of the form (n, n) is self-conjugate (real): complex conjugation acts on Dynkin labels by $(n_1, n_2) \mapsto (n_2, n_1)$, which is the statement that $\bar{\mathbf{3}}$ has the labels of $\mathbf{3}$ exchanged, and more generally that conjugating a representation swaps the roles of the two nodes of the A_2 Dynkin

diagram; a label fixed by this exchange is therefore self-conjugate. (For $\mathfrak{su}(3)$ self-conjugate representations are always real rather than pseudoreal, since $\mathfrak{su}(3)$, unlike $\mathfrak{su}(2)$, admits no invariant antisymmetric — quaternionic-type — structure on any of its representations.)

44.4 The Six-Dimensional Representation

The representation with Dynkin label $(2, 0)$, highest weight $2\tilde{\lambda}_1$, describes the symmetric product of two fundamentals, $(\mathbf{3} \otimes \mathbf{3})_{\text{sym}} = \mathbf{6}$: writing the weights of $\mathbf{3}$ as e_u, e_d, e_s (the three points of Fig. 17), the six weights of $\mathbf{6}$ are

$$2e_u, 2e_d, 2e_s, \quad e_u + e_d, e_u + e_s, e_d + e_s,$$

each of multiplicity one, forming a triangle similar to that of $\mathbf{3}$ but scaled by a factor of two and with the three edge midpoints filled in (Fig. 18). The dimension check follows from a general closed-form dimension formula for $\mathfrak{su}(3)$ representations (a special case of the Weyl dimension formula, which we do not derive here):

$$\dim(n_1, n_2) = \frac{(n_1 + 1)(n_2 + 1)(n_1 + n_2 + 2)}{2}, \quad (7)$$

which correctly reproduces $\dim(1, 0) = 3$, $\dim(1, 1) = 8$, and $\dim(2, 0) = 6$.

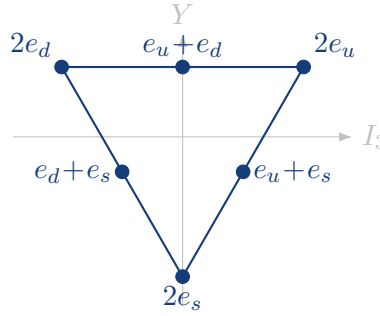


Figure 18: Weight diagram of the six-dimensional representation $\mathbf{6} = (2, 0)$ of $\mathfrak{su}(3)$: the symmetric square of the fundamental. The conjugate $\bar{\mathbf{6}} = (0, 2)$ has the opposite orientation, as for $\mathbf{3}$ vs. $\bar{\mathbf{3}}$.

44.5 The Decuplet and a Physical Application

The representation with Dynkin label $(3, 0)$, highest weight $3\tilde{\lambda}_1$, is the totally symmetric cube of the fundamental, $(\mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3})_{\text{sym}}$. Its weights are all points $n_u e_u + n_d e_d + n_s e_s$ with $n_u + n_d + n_s = 3$, $n_i \geq 0$ integers, giving $\binom{5}{2} = 10$ weights of multiplicity one, arranged in four rows of 4, 3, 2, 1 points — the familiar **decuplet** pattern, consistent with Eq. (7): $\dim(3, 0) = \frac{4 \cdot 1 \cdot 5}{2} = 10$.

Physical perspective

This is one of the most celebrated applications of Lie group theory to particle physics. Gell-Mann and Ne’eman (1961–62) organised the known spin- $\frac{3}{2}$ baryon resonances — the Δ (isospin quartet), Σ^* (isospin triplet), Ξ^* (isospin doublet) — into the top three rows of the decuplet in Fig. 19, and noted that the pattern of equally spaced masses (the Gell-Mann–Okubo mass formula, a consequence of $SU(3)$ being only approximately, not exactly, a symmetry) demanded a tenth particle: a strangeness $S = -3$, isospin-singlet baryon Ω^- , with a predicted mass around 1680 MeV. The Ω^- was discovered at Brookhaven in 1964 with essentially the predicted mass (measured at 1672 MeV, close to the ≈ 1680 MeV estimate),

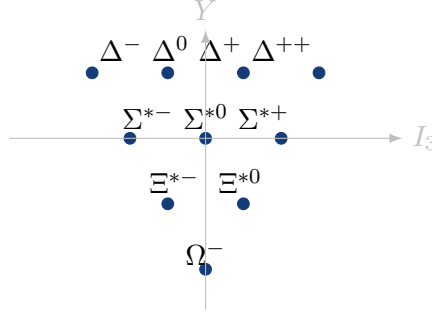


Figure 19: The baryon decuplet, the $(3, 0)$ representation of flavour $\mathfrak{su}(3)$: four rows of isospin multiplets ($I = 3/2, 1, 1/2, 0$) at descending hypercharge Y .

and with the predicted charge and decay properties, providing dramatic confirmation that the approximate flavour symmetry of the strong interactions is indeed $SU(3)$, generated by the algebra $\mathfrak{su}(3)$ studied in this section. (In the modern quark model, the ten states of Fig. 19 are the totally symmetric combinations of three quarks $uuu, uud, \dots, sss$ drawn from $\{u, d, s\}$, exactly the weight $n_u e_u + n_d e_d + n_s e_s$ construction above.)

45 Explicit Construction and Classification of the Classical Lie Algebras

Part D's classification theorem enumerated the possible Dynkin diagrams from purely combinatorial rigidity constraints. We now close the loop from the other end: we construct, completely explicitly, the four infinite matrix families that realise the diagrams A_{N-1}, B_n, C_n, D_n , verifying by direct computation that their Cartan subalgebras, root systems, and Cartan matrices are exactly what the classification predicts.

45.1 $\mathfrak{su}(N) \leftrightarrow A_{N-1}$

Let $G = SU(N) = \{g \in U(N) : \det g = 1\}$. Its Lie algebra $\mathfrak{su}(N)$ consists of the traceless anti-Hermitian $N \times N$ matrices, of real dimension $N^2 - 1$; physicists prefer the Hermitian generators $T = i a$ ($a \in \mathfrak{su}(N)$), so that $\mathfrak{su}(N) = \{iT : T^\dagger = T, \text{tr } T = 0\}$.

A Cartan subalgebra is given by the traceless diagonal Hermitian matrices, $\mathfrak{h} = \{\text{diag}(h_1, \dots, h_N) : \sum_i h_i = 0\}$, of dimension $\ell = N - 1 = \text{rank } \mathfrak{su}(N)$. An orthonormal basis (with respect to $\text{tr}(H_i H_j) = \delta_{ij}$, the trace form used throughout) is

$$H_k = \frac{1}{\sqrt{k(k+1)}} \text{diag}(\underbrace{1, \dots, 1}_k, -k, 0, \dots, 0), \quad k = 1, \dots, N-1, \quad (8)$$

which is exactly the Gell-Mann pattern of Part D's λ_3, λ_8 generalised to arbitrary N (at $N = 3$, H_1, H_2 reproduce $\lambda_3/\sqrt{2}, \lambda_8/\sqrt{2}$, the normalisation fixed by $\text{tr}(H_i H_j) = \delta_{ij}$ versus $\text{tr}(\lambda_a \lambda_b) = 2\delta_{ab}$).

The remaining $N(N-1)$ off-diagonal generators E_{ij} ($i \neq j$, the matrix with a single 1 in position (i, j)) satisfy $[H, E_{ij}] = (h_i - h_j)E_{ij}$ for $H = \text{diag}(h_1, \dots, h_N) \in \mathfrak{h}$: they are root vectors. To read off the corresponding root as a vector in $\mathfrak{h}_{\mathbb{R}}^* \cong \mathbb{R}^{N-1}$, introduce the **weight vectors** $w^{(1)}, \dots, w^{(N)} \in \mathbb{R}^{N-1}$ of the defining (N -dimensional) representation, whose k -th component is

the i -th diagonal entry of H_k :

$$w_k^{(i)} = \begin{cases} \frac{1}{\sqrt{k(k+1)}}, & i \leq k, \\ -\frac{k}{\sqrt{k(k+1)}}, & i = k+1, \\ 0, & i > k+1. \end{cases} \quad (9)$$

Proposition 45.1. The weight vectors Eq. (9) satisfy $|w^{(i)}|^2 = \frac{N-1}{N}$ for every i , $w^{(i)} \cdot w^{(j)} = -\frac{1}{N}$ for $i \neq j$, and $\sum_{i=1}^N w^{(i)} = 0$.

Proof. The last identity is immediate: $\sum_i w_k^{(i)} = k \cdot \frac{1}{\sqrt{k(k+1)}} - k \cdot \frac{1}{\sqrt{k(k+1)}} = 0$ for every k (it is nothing but $\text{tr } H_k = 0$, read off the diagonal). For the norms, a direct computation using Eq. (9) gives, for fixed i ,

$$|w^{(i)}|^2 = \sum_{k=1}^{N-1} (w_k^{(i)})^2 = \sum_{k=i}^{N-1} \frac{1}{k(k+1)} + \frac{(i-1)^2}{i(i-1)} \cdot [i \geq 2] \cdots$$

— rather than track this sum term by term, it is quicker to use $\sum_i (w^{(i)})^2 = \sum_i \sum_j w_j^{(i)} w_j^{(i)} = \sum_j \sum_i w_j^{(i)} w_j^{(i)} = \sum_j \text{tr}(H_j^2) = \sum_j 1 = N-1$ (each H_j has $\text{tr } H_j^2 = 1$ by orthonormality), together with the fact — manifest from Eq. (9), since H_k is invariant under any permutation of the first k indices and of the last $N-k$ indices, and the whole construction is covariant under the residual permutation symmetry of the N weight labels combined with an orthogonal change of Cartan basis — that all N vectors $w^{(i)}$ have the *same* length $c := |w^{(i)}|^2$. Hence $Nc = N-1$, i.e. $c = (N-1)/N$. For $i \neq j$, squaring $\sum_i w^{(i)} = 0$ gives $0 = |\sum_i w^{(i)}|^2 = \sum_i |w^{(i)}|^2 + \sum_{i \neq j} w^{(i)} \cdot w^{(j)} = N \cdot \frac{N-1}{N} + N(N-1)(w^{(i)} \cdot w^{(j)})$ (using, again by the permutation symmetry, that $w^{(i)} \cdot w^{(j)}$ is the same for every pair $i \neq j$), which solves to $w^{(i)} \cdot w^{(j)} = -1/N$. $\square$

Theorem 45.2. The roots of $\mathfrak{su}(N)_{\mathbb{C}}$ are $\alpha^{(ij)} := w^{(i)} - w^{(j)}$ for $i \neq j$ (there are $N(N-1)$ of them, matching $\dim \mathfrak{su}(N) - \text{rank } \mathfrak{su}(N) = (N^2 - 1) - (N-1) = N(N-1) = N^2 - N$). The simple roots are $\alpha^{(i)} := w^{(i)} - w^{(i+1)}$, $i = 1, \dots, N-1$, and the Cartan matrix is

$$A_{ij} = \begin{cases} 2, & i = j \\ -1, & |i - j| = 1, \\ 0, & |i - j| \geq 2 \end{cases}$$

exactly the A_{N-1} Cartan matrix of the classification theorem (Part D).

Proof. That $\alpha^{(ij)}$ is the root attached to E_{ij} was observed above. Using Proposition 45.1, $\langle \alpha^{(i)}, \alpha^{(i)} \rangle = |w^{(i)}|^2 - 2w^{(i)} \cdot w^{(i+1)} + |w^{(i+1)}|^2 = 2 \cdot \frac{N-1}{N} + \frac{2}{N} = 2$, and

$$\begin{aligned} \langle \alpha^{(i)}, \alpha^{(i+1)} \rangle &= (w^{(i)} - w^{(i+1)}) \cdot (w^{(i+1)} - w^{(i+2)}) \\ &= w^{(i)} \cdot w^{(i+1)} - w^{(i)} \cdot w^{(i+2)} - |w^{(i+1)}|^2 + w^{(i+1)} \cdot w^{(i+2)} \\ &= -\frac{1}{N} + \frac{1}{N} - \frac{N-1}{N} - \frac{1}{N} = -1, \end{aligned}$$

while for $|i - j| \geq 2$ every one of the four terms in the analogous expansion involves distinct labels, giving $\langle \alpha^{(i)}, \alpha^{(j)} \rangle = -\frac{1}{N} + \frac{1}{N} + \frac{1}{N} - \frac{1}{N} = 0$. Dividing by $\langle \alpha^{(i)}, \alpha^{(i)} \rangle = 2$ gives the stated Cartan matrix. $\square$

45.2 $\mathfrak{so}(2n+1) \leftrightarrow B_n$

The Lie algebra $\mathfrak{so}(2n+1)$ (real antisymmetric $(2n+1) \times (2n+1)$ matrices, $\dim = n(2n+1)$, rank $\ell = n$) admits a Cartan subalgebra of n independent, mutually commuting infinitesimal rotations, one in each of n orthogonal 2-planes inside $\mathbb{R}^{2n+1}$, leaving one axis fixed (the unpaired, “odd” direction):

$$H = \text{diag}(\theta_1 J, \theta_2 J, \dots, \theta_n J, 0), \quad J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Complexifying and diagonalising each block (J has eigenvectors $(1, \pm i)/\sqrt{2}$ with eigenvalues $\pm i$), the $2n$ paired coordinates acquire weights $\pm e^\ell$ ($\ell = 1, \dots, n$, the standard unit vectors of $\mathbb{R}^n \cong \mathfrak{h}_{\mathbb{R}}^*$) under the defining representation, while the one unpaired coordinate has weight 0.

Theorem 45.3. The roots of $\mathfrak{so}(2n+1)_{\mathbb{C}}$ are $\pm e^i \pm e^j$ ($i < j$, the “long” roots, $2n(n-1)$ of them) and $\pm e^i$ ($i = 1, \dots, n$, the “short” roots, $2n$ of them), for a total of $2n^2$ roots, matching $\dim - \text{rank} = n(2n+1) - n = 2n^2$. The simple roots are $\alpha_i = e^i - e^{i+1}$ ($i = 1, \dots, n-1$, long) and $\alpha_n = e^n$ (short), giving the Cartan matrix with $A_{n-1,n} = -2$, $A_{n,n-1} = -1$ (all other off-diagonal entries as in A_{n-1}) — exactly the B_n Dynkin diagram: a chain of $n-1$ simple bonds ending in a double bond, with the arrow pointing towards the short root α_n .

Sketch. The long roots arise, exactly as for $\mathfrak{su}(N)$, from generators mixing two paired 2-planes; the short roots arise from generators mixing a paired 2-plane with the single unpaired direction (there are 2 real generators per pair (ℓ , fixed axis), complexifying to the two root vectors $\pm e^\ell$). Since $|e^i \pm e^j|^2 = 2$ and $|e^i|^2 = 1$, one finds $\langle \alpha_{n-1}, \alpha_n \rangle = (e^{n-1} - e^n) \cdot e^n = -1$, so $A_{n-1,n} = 2(-1)/1 = -2$ while $A_{n,n-1} = 2(-1)/2 = -1$, as claimed (all other entries reproduce the A_{n-1} pattern of Theorem 45.2, since the long roots among themselves behave exactly as the $\mathfrak{su}(n)$ weights $e^1, \dots, e^n$ did there). $\square$

45.3 $\mathfrak{usp}(2n) \leftrightarrow C_n$

The **compact symplectic group** $USp(2n) = Sp(2n, \mathbb{C}) \cap U(2n)$ consists of unitary $2n \times 2n$ matrices g preserving an antisymmetric form, $g^T J g = J$, with $J = \begin{pmatrix} 0 & \mathbb{1}_n \\ -\mathbb{1}_n & 0 \end{pmatrix}$. Its Lie algebra, $\mathfrak{usp}(2n) = \{a : a^T j + j a = 0, a^\dagger = -a\}$, is most conveniently parametrised in $n \times n$ blocks as (writing Hermitian generators $A = ia$)

$$A = \begin{pmatrix} W & \bar{P} \\ P & -W^T \end{pmatrix}, \quad W^\dagger = W, \quad P^T = P,$$

i.e. an arbitrary Hermitian $n \times n$ block W together with a *complex symmetric* $n \times n$ block P ; counting real parameters, n^2 (Hermitian W) + $2 \cdot \frac{n(n+1)}{2}$ (complex symmetric P) = $n^2 + n(n+1) = n(2n+1) = \dim \mathfrak{usp}(2n)$, of rank $\ell = n$ (note this equals $\dim \mathfrak{so}(2n+1) = n(2n+1)$ for *every* n — a structural fact, since B_n and C_n are dual root systems with the same number of roots, $2n^2$, not a coincidence; what *is* an accident special to rank 2 is the actual isomorphism $\mathfrak{usp}(4) \cong \mathfrak{so}(5)$, since the root systems, though equal in size, are genuinely different shapes — long and short roots exchanged — for $n \geq 3$, as we now see).

A Cartan subalgebra is given by taking $W = \text{diag}(u_1, \dots, u_n)$ (real diagonal — genuinely Hermitian, as required, unlike a purely imaginary diagonal matrix, which would be anti-Hermitian and violate $W^\dagger = W$ above; this is the $\mathfrak{usp}(2n)$ analogue of the traceless real diagonal Cartan subalgebra of $\mathfrak{su}(n)$, but *without* the tracelessness constraint) and $P = 0$:

$$H = \text{diag}(u_1, \dots, u_n, -u_1, \dots, -u_n),$$

an n -dimensional (full rank) abelian subalgebra. The defining $2n$ -dimensional representation has weights $\pm e^i$ ($i = 1, \dots, n$), exactly as for $\mathfrak{so}(2n+1)$'s paired directions but now with *no* tracelessness constraint linking them.

Theorem 45.4. The roots of $\mathfrak{usp}(2n)_{\mathbb{C}}$ are $\pm e^i \pm e^j$ ($i < j$, from the off-diagonal entries of W and P mixing two indices, $2n(n-1)$ of them) and $\pm 2e^i$ ($i = 1, \dots, n$, from the diagonal entries of P , $2n$ of them): $2n^2$ roots in total, matching $\dim - \text{rank} = n(2n+1) - n = 2n^2$. The simple roots are $\alpha_i = e^i - e^{i+1}$ ($i = 1, \dots, n-1$, short, $|\alpha_i|^2 = 2$) and $\alpha_n = 2e^n$ (long, $|\alpha_n|^2 = 4$), giving $A_{n-1,n} = -1$, $A_{n,n-1} = -2$ — the C_n Dynkin diagram: the same chain as B_n , but with the double-bond arrow reversed, now pointing *away* from the long root α_n .

Sketch. The off-diagonal entries of the Hermitian block W contribute the generators of weight $e^i - e^j$ ($i \neq j$), exactly as the off-diagonal E_{ij} did for $\mathfrak{su}(n)$; the off-diagonal entries of the complex symmetric block P (and its conjugate $\bar{P}$) contribute weight $\pm(e^i + e^j)$ ($i < j$); and the *diagonal* entries of P (which are allowed to be nonzero precisely because P need only be symmetric, not traceless or off-diagonal) contribute the extra long roots $\pm 2e^i$ that have no analogue in $\mathfrak{su}(n)$ or $\mathfrak{so}(2n+1)$. The Cartan matrix entries follow exactly as in Theorem 45.3, with the roles of long and short exchanged: $\langle \alpha_{n-1}, \alpha_n \rangle = (e^{n-1} - e^n) \cdot 2e^n = -2$, so $A_{n-1,n} = 2(-2)/4 = -1$ and $A_{n,n-1} = 2(-2)/2 = -2$. $\square$

45.4 $\mathfrak{so}(2n) \leftrightarrow D_n$

Finally, $\mathfrak{so}(2n)$ (real antisymmetric $2n \times 2n$ matrices, $\dim = n(2n-1)$, $\text{rank } \ell = n$) is constructed exactly as $\mathfrak{so}(2n+1)$ was (Theorem 45.3), but now *without* an unpaired axis: all $2n$ coordinates pair up into n rotation planes, so the defining representation has weights $\pm e^i$ ($i = 1, \dots, n$) with *no* zero weight, and there is consequently no analogue of the short roots $\pm e^i$ of B_n .

Theorem 45.5. The roots of $\mathfrak{so}(2n)_{\mathbb{C}}$ are exactly $\pm e^i \pm e^j$ ($i < j$), $2n(n-1)$ of them, matching $\dim - \text{rank} = n(2n-1) - n = 2n(n-1)$. The simple roots are $\alpha_i = e^i - e^{i+1}$ ($i = 1, \dots, n-1$) and $\alpha_n = e^{n-1} + e^n$, all of equal length ($|\alpha_i|^2 = 2$), giving the *simply-laced* Cartan matrix of D_n : the A_{n-1} chain $\alpha_1, \dots, \alpha_{n-1}$, together with α_n attached to α_{n-2} (not α_{n-1}) by a single bond, producing the characteristic “fork” at the end of the diagram.

Sketch. $\langle \alpha_{n-2}, \alpha_n \rangle = (e^{n-2} - e^{n-1}) \cdot (e^{n-1} + e^n) = -1$, so α_{n-2} and α_n are connected by a single bond; while $\langle \alpha_{n-1}, \alpha_n \rangle = (e^{n-1} - e^n) \cdot (e^{n-1} + e^n) = |e^{n-1}|^2 - |e^n|^2 = 0$, so α_{n-1} and α_n are *not* connected — exactly the fork of the D_n diagram, with α_{n-1} and α_n as the two prongs. $\square$

45.5 Classical Groups as Rotation Groups over $\mathbb{R}, \mathbb{C}, \mathbb{H}$

The four families just constructed admit a single unifying description, which is one of the most elegant facts in the subject: each is the automorphism group of a vector space equipped with a natural inner product, where the vector space is built from one of the three real division algebras $\mathbb{R}, \mathbb{C}, \mathbb{H}$ (the quaternions).

Remark. $SO(n)$ preserves the standard real symmetric bilinear form $x^T y$ on $\mathbb{R}^n$: it is the group of “rotations” of real vectors. $SU(N)$ preserves the standard Hermitian form $w^\dagger w$ on $\mathbb{C}^N$, together with a choice of complex volume form ($\det g = 1$): it is the group of “rotations” of complex vectors. And $USp(2n)$ preserves the natural Hermitian norm on a *quaternionic* vector space $\mathbb{H}^n$: pairing up the $2n$ complex components of $\mathbb{C}^{2n}$ into n quaternions $q_a = z_a^{(1)} + z_a^{(2)}j$ (using $\mathbf{1}, i, j, k$ for the quaternion units, $j^2 = -1$, $ij = k$), the

condition $g^\dagger g = \mathbb{1}$ together with $g^T J g = J$ is exactly the statement that g commutes with right multiplication by j on each quaternionic slot, i.e. g is $\mathbb{H}$ -linear and norm-preserving: $USp(2n)$ is the “rotation group” of quaternionic vectors. This is why the three families are traditionally called the **classical groups**: they exhaust the possible “orthogonal-type” symmetry groups over the three associative real division algebras, with $\mathfrak{so}(2n)$ and $\mathfrak{so}(2n+1)$ being the even- and odd-dimensional cases of the same real story.

46 Invariant Tensors and Casimir Operators

Beyond the quadratic Casimir built from the Killing form (Part D), every classical Lie algebra carries further invariant tensors, of direct importance in physics (e.g. in identifying which field-theoretic couplings are gauge invariant, and in computing chiral anomalies).

Definition 46.1. For $\mathfrak{su}(N)$ with Hermitian generators T_a normalised as $\text{tr}(T_a T_b) = \frac{1}{2} \delta_{ab}$, the rank- r **invariant tensor** is the totally symmetrised trace

$$d_{a_1 \dots a_r} := \text{tr}(T_{(a_1} \dots T_{a_r)}),$$

where $(\dots)$ denotes symmetrisation over all $r!$ orderings (divided by $r!$).

Proposition 46.1. $d_{a_1 \dots a_r}$ is invariant under $T_a \rightarrow U^\dagger T_a U$ for any $U \in SU(N)$, and $C := d_{a_1 \dots a_r} T^{a_1} \dots T^{a_r}$ commutes with every generator (it is a higher **Casimir operator**).

Proof. Invariance under conjugation follows from cyclicity of the trace applied to each term of the symmetrisation. Infinitesimally, invariance under $T_a \rightarrow T_a + i\epsilon[T_c, T_a]$ means the sum of r terms obtained by replacing one factor at a time in $\text{tr}(T_{a_1} \dots T_{a_r})$ by $[T_c, T_{a_k}]$ must vanish after symmetrisation; this is again a direct consequence of cyclicity, e.g. for $r = 2$: $\text{tr}([T_c, T_{a_1}] T_{a_2}) + \text{tr}(T_{a_1} [T_c, T_{a_2}]) = \text{tr}(T_c T_{a_1} T_{a_2}) - \text{tr}(T_{a_1} T_c T_{a_2}) + \text{tr}(T_{a_1} T_c T_{a_2}) - \text{tr}(T_{a_1} T_{a_2} T_c) = \text{tr}(T_c T_{a_1} T_{a_2}) - \text{tr}(T_{a_1} T_{a_2} T_c) = 0$ by cyclicity, and the general r follows the same telescoping pattern. Since C is built from an invariant tensor contracted with all generators, $[T_b, C] = 0$ follows from the same infinitesimal invariance statement. $\square$

Remark. For $\mathfrak{su}(N)$, the first nontrivial higher invariant is the rank-3, totally symmetric **d -symbol**. The standard physics normalisation is $d_{abc} := 2 \text{tr}(T_a \{T_b, T_c\})$ — a fixed overall rescaling of the $r = 3$ case of Definition 46.1 ($d_{abc} = 2 \text{tr}(T_a T_b T_c)$) there, so the two differ by the conventional factor of 4; symmetrised traces of both even and odd rank can be nonzero for $\mathfrak{su}(N)$ when $N \geq 3$, since $\text{tr}(T_a) = 0$ already kills $r = 1$, and $d_{ab} \propto \delta_{ab}$ reduces to the quadratic Casimir). In physics, d_{abc} is precisely the **anomaly coefficient**: a chiral gauge theory is free of the triangle (Adler–Bell–Jackiw) anomaly only if d_{abc} vanishes when summed with appropriate signs over all Weyl fermions in the theory — a nontrivial group-theoretic constraint that, e.g., is what fixes the number of colours $N_c = 3$ in the Standard Model, via the requirement that the anomaly cancel *within each generation* of quarks and leptons separately (it constrains N_c , not the number of generations, which anomaly cancellation says nothing about).

Remark. For $\mathfrak{so}(2n)$, the generators are represented by antisymmetric matrices $A_{\mu\nu} = -A_{\nu\mu}$ in the vector representation (this is specific to $\mathfrak{so}(2n)$: the defining property of $\mathfrak{usp}(2n)$ ’s

generators, by contrast, is that they are never antisymmetric in any basis, only Hermitian and symplectic-traceless, Section 45.3 above, so no analogous Pfaffian invariant exists for $\mathfrak{usp}(2n)$, and an antisymmetric $2n \times 2n$ matrix carries, beyond $\det A = (\text{Pf } A)^2$, its own square root of the determinant, the **Pfaffian**

$$\text{Pf}(A) = \frac{1}{2^n n!} \varepsilon^{i_1 i_2 \dots i_{2n-1} i_{2n}} A_{i_1 i_2} \dots A_{i_{2n-1} i_{2n}},$$

an invariant of the vector representation not expressible as a trace, and one which changes sign under $O(2n) \setminus SO(2n)$ (an improper orthogonal transformation) while $\det A$ does not. The two inequivalent “chiral” Weyl-type spinor representations of $SO(2n)$ (unlike the single spinor representation of $SO(2n+1)$) are a related but logically separate fact: they are distinguished by the eigenvalue ± 1 of the chirality operator $\Gamma_{2n+1} := i^n \Gamma_1 \Gamma_2 \dots \Gamma_{2n}$ (built from the Clifford algebra generators, not from Pf directly) — the even-dimensional-orthogonal analogue of the parity distinction we meet again for the Lorentz group below.

47 Physical Application: The Lorentz Group and Its Representations

We close by using everything developed in this chapter, together with Part D’s abstract computation that $\mathfrak{so}(3,1)_{\mathbb{C}} \cong \mathfrak{su}(2)_{\mathbb{C}} \oplus \mathfrak{su}(2)_{\mathbb{C}}$, to classify the finite-dimensional representations of the Lorentz group explicitly, down to writing out the Weyl spinor representations by hand.

47.1 The Lorentz Group

Equip $\mathbb{R}^4$ with coordinates $x^\mu = (t, x, y, z)$ and the Minkowski metric $\eta = \text{diag}(1, -1, -1, -1)$, so that $\|x\|^2 = x^T \eta x = t^2 - x^2 - y^2 - z^2$ is the invariant interval.

Definition 47.1. The **Lorentz group** is $O(1,3) := \{\Lambda \in GL(4, \mathbb{R}) : \Lambda^T \eta \Lambda = \eta\}$, the linear transformations $x'^\mu = \Lambda^\mu_\nu x^\nu$ preserving $\|x\|^2$.

Taking determinants of $\Lambda^T \eta \Lambda = \eta$ gives $(\det \Lambda)^2 = 1$, so $\det \Lambda = \pm 1$; and the $(0,0)$ component of the same equation reads $(\Lambda^0_0)^2 - \sum_i (\Lambda^i_0)^2 = 1$, forcing $(\Lambda^0_0)^2 \geq 1$, i.e. $\Lambda^0_0 \geq 1$ (**orthochronous**, preserves the direction of time) or $\Lambda^0_0 \leq -1$. Combining the two independent signs gives four connected components, represented by $\mathbb{1}$, parity $P = \text{diag}(1, -1, -1, -1)$, time reversal $T = \text{diag}(-1, 1, 1, 1)$, and $PT = -\mathbb{1}$. The component containing the identity, $\det \Lambda = +1$ and $\Lambda^0_0 \geq 1$, is the **proper orthochronous Lorentz group** $SO^+(1,3)$; the full group is the union of four cosets $SO^+(1,3)$, $P \cdot SO^+(1,3)$, $T \cdot SO^+(1,3)$, $PT \cdot SO^+(1,3)$.

47.2 The Lie Algebra $\mathfrak{so}(3,1)$

Writing $\Lambda = \mathbb{1} + \omega$ for an infinitesimal element of $SO^+(1,3)$, the defining condition $\Lambda^T \eta \Lambda = \eta$ linearises to $\omega^T \eta + \eta \omega = 0$, i.e. $\omega_{\mu\nu} := \eta_{\mu\rho} \omega^\rho_\nu$ is antisymmetric, $\omega_{\mu\nu} = -\omega_{\nu\mu}$: exactly $6 = \binom{4}{2}$ independent parameters, matching $\dim \mathfrak{so}(3,1) = n(2n-1)|_{n=2} = 6$ of Theorem 45.5 (with signature, rather than definiteness, of the quadratic form — the classification of Part D concerns the *complexified* algebra, which does not see the difference between $\mathfrak{so}(4)$ and $\mathfrak{so}(3,1)$).

Definition 47.2. The generators $M^{\mu\nu} = -M^{\nu\mu}$ ($\mu, \nu = 0, 1, 2, 3$) of $\mathfrak{so}(3,1)$ are defined, in the defining (vector) representation, by

$$(M^{\mu\nu})^\rho_\sigma = i(\eta^{\mu\rho} \delta^\nu_\sigma - \eta^{\nu\rho} \delta^\mu_\sigma),$$

so that a general infinitesimal Lorentz transformation is $\omega^\rho{}_\sigma = -\frac{i}{2}\omega_{\mu\nu}(M^{\mu\nu})^\rho{}_\sigma$ for antisymmetric parameters $\omega_{\mu\nu}$.

Proposition 47.1. The generators of Definition 47.2 satisfy

$$[M^{\mu\nu}, M^{\rho\sigma}] = i(\eta^{\nu\rho}M^{\mu\sigma} - \eta^{\mu\rho}M^{\nu\sigma} - \eta^{\nu\sigma}M^{\mu\rho} + \eta^{\mu\sigma}M^{\nu\rho}). \quad (10)$$

Proof. Direct matrix multiplication using $(M^{\mu\nu})^\rho{}_\sigma = i(\eta^{\mu\rho}\delta^\nu_\sigma - \eta^{\nu\rho}\delta^\mu_\sigma)$ expresses $(M^{\mu\nu}M^{\rho\sigma})^\tau{}_\lambda - (M^{\rho\sigma}M^{\mu\nu})^\tau{}_\lambda$ as a sum of four terms, each a product of two η 's and one δ ; collecting these using $\eta^{\mu\rho} = \eta^{\rho\mu}$ and re-contracting each pair back into a single generator, e.g. $\eta^{\nu\rho}(M^{\mu\sigma})^\tau{}_\lambda = \eta^{\nu\rho} \cdot i(\eta^{\mu\tau}\delta^\sigma_\lambda - \eta^{\sigma\tau}\delta^\mu_\lambda)$, reproduces exactly the right-hand side of Eq. (10); we omit the routine index bookkeeping. $\square$

It is more transparent to package the six generators as three rotations and three boosts,

$$J^i := \frac{1}{2}\varepsilon^{ijk}M^{jk} \quad (i, j, k = 1, 2, 3), \quad K^i := M^{0i},$$

in terms of which Eq. (10) becomes

$$[J^i, J^j] = i\varepsilon^{ijk}J^k, \quad [J^i, K^j] = i\varepsilon^{ijk}K^k, \quad [K^i, K^j] = -i\varepsilon^{ijk}J^k. \quad (11)$$

The first relation says the J^i alone generate an $\mathfrak{su}(2) \cong \mathfrak{so}(3)$ rotation subalgebra; the second says the boosts K^i transform as a vector under rotations; the third — the genuinely relativistic relation — says that two boosts do *not* commute into zero but into a *rotation* (the algebraic origin of Thomas precession). Up to the change of basis and normalisation between the anti-Hermitian generators θ_a used in Part D and the Hermitian generators $T_a = -i\theta_a$ used here (the sign fixed by requiring $[T_a, T_b] = +i\varepsilon_{abc}T_c$, matching the physics convention used throughout this Part, from $[\theta_a, \theta_b] = -\varepsilon_{abc}\theta_c$), Eq. (11) is exactly the abstract Lorentz-algebra relation already established in Part D (Section 36.3), now rederived from the explicit tensor generators $M^{\mu\nu}$.

47.3 Complexification and the (j, j') Classification

Exactly as in Part D, define the complex (non-Hermitian) combinations

$$A_i^\pm := \frac{1}{2}(J^i \pm iK^i).$$

A direct substitution into Eq. (11) gives

$$[A_i^+, A_j^+] = i\varepsilon_{ijk}A_k^+, \quad [A_i^-, A_j^-] = i\varepsilon_{ijk}A_k^-, \quad [A_i^+, A_j^-] = 0,$$

i.e. $\mathfrak{so}(3, 1)_\mathbb{C} \cong \mathfrak{su}(2)_\mathbb{C} \oplus \mathfrak{su}(2)_\mathbb{C}$, the same conclusion as Part D reached from the real generators θ_p, θ_{p+3} , now rederived from the explicit tensor generators $M^{\mu\nu}$. By Theorem 43.3 applied to each $\mathfrak{su}(2)$ factor separately, finite-dimensional representations of the (complexified, hence also the real) Lorentz algebra are labelled by a pair of spins (j, j') , of dimension $(2j+1)(2j'+1)$, with A_i^+ acting as spin j and A_i^- as spin j' .

A **scalar field** is, by definition, a field carrying the trivial $(0, 0)$ representation: under a Lorentz transformation $x \rightarrow x' = \Lambda x$, it transforms with no matrix at all, $\phi'(x') = \phi(x)$, i.e. $\phi'(x) = \phi(\Lambda^{-1}x)$ — the field's numerical *value* at a given spacetime point does not change, only the point being referred to.

47.4 The Weyl Spinor Representations

We now construct $(\frac{1}{2}, 0)$ and $(0, \frac{1}{2})$ by hand.

(j, j')	dimension	field
$(0, 0)$	1	scalar $\phi(x)$
$(\frac{1}{2}, 0)$	2	left-handed Weyl spinor χ_α
$(0, \frac{1}{2})$	2	right-handed Weyl spinor $\bar{\chi}_{\dot{\alpha}}$
$(\frac{1}{2}, \frac{1}{2})$	4	vector V^μ

Table 2: The simplest finite-dimensional representations of the Lorentz algebra.

Example 47.1. a) The representation $(\frac{1}{2}, 0)$. Here $A_i^- \equiv 0$ identically and $A_i^+ = \frac{1}{2}\sigma_i$ (the spin- $\frac{1}{2}$ representation of $\mathfrak{su}(2)$, Part C). Solving $J^i = A_i^+ + A_i^- = \frac{1}{2}\sigma_i$ and $K^i = -i(A_i^+ - A_i^-) = -\frac{i}{2}\sigma_i$ for the original Hermitian generators gives, for a finite group element built from a rotation angle $\vec{\theta}$ and boost rapidity $\vec{\varphi}$,

$$g_L = \exp\left(i\vec{\theta} \cdot \vec{J} + i\vec{\varphi} \cdot \vec{K}\right) = \exp\left[\frac{i}{2}(\vec{\theta} - i\vec{\varphi}) \cdot \vec{\sigma}\right],$$

acting on a two-component **left-handed spinor** χ_α ($\alpha = 1, 2$), matching Table 2. (Rotations, being compact, are represented unitarily, with the familiar factor of i multiplying the *Hermitian* generator $\vec{J}$; boosts, being non-compact, are represented by the non-unitary, growing/decaying exponential $\exp(i\vec{\varphi} \cdot \vec{K}) = \exp(\frac{i}{2}\vec{\varphi} \cdot \vec{\sigma})$, since $\vec{K} = -\frac{i}{2}\vec{\sigma}$ comes out *anti*-Hermitian in this non-unitary representation — consistent with the general fact, Part D, that finite-dimensional representations of a non-compact real form need not, and here cannot, be unitary.)

b) The representation $(0, \frac{1}{2})$. Now $A_i^+ \equiv 0$ and $A_i^- = \frac{1}{2}\sigma_i$, giving $J^i = \frac{1}{2}\sigma_i$ (the *same* rotation generators) but $K^i = +\frac{i}{2}\sigma_i$ (the *opposite* sign), so

$$g_R = \exp\left[\frac{i}{2}(\vec{\theta} + i\vec{\varphi}) \cdot \vec{\sigma}\right],$$

acting on a two-component **right-handed spinor** $\bar{\chi}_{\dot{\alpha}}$ ($\dot{\alpha} = 1, 2$; the dot is the traditional notation distinguishing the two inequivalent two-dimensional representations), matching Table 2.

Proposition 47.2 (Parity and spinors). Under parity $P = \text{diag}(1, -1, -1, -1)$, $\vec{J} \rightarrow \vec{J}$ (both indices of M^{jk} are spatial and flip sign together) while $\vec{K} \rightarrow -\vec{K}$ (one index of M^{0i} is spatial and flips sign, the other is temporal and does not). Hence $A_i^\pm = \frac{1}{2}(J^i \pm iK^i) \xrightarrow{P} \frac{1}{2}(J^i \mp iK^i) = A_i^\mp$: parity exchanges the two $\mathfrak{su}(2)$ factors, and therefore exchanges the representations $(\frac{1}{2}, 0) \leftrightarrow (0, \frac{1}{2})$, i.e. the right- and left-handed spinors.

Physical perspective

This is the group-theoretic root of parity violation in the weak interaction. A field theory built symmetrically out of χ_α and $\bar{\chi}_{\dot{\alpha}}$ (e.g. a Dirac fermion, which combines both into the four-component representation $(\frac{1}{2}, 0) \oplus (0, \frac{1}{2})$ with equal, parity-symmetric couplings) is parity invariant; but the Standard Model couples the $SU(2)_{\text{weak}}$ gauge bosons *only* to left-handed Weyl fermions χ_α (and their right-handed conjugates $\bar{\chi}_{\dot{\alpha}}$ transform as gauge singlets), an explicit, maximal violation of the exchange symmetry of Proposition 47.2 — exactly the effect whose experimental confirmation, in the β -decay of polarised ^{60}Co , earned Lee and Yang the 1957 Nobel Prize: they had *proposed* parity violation in 1956, and Wu and collaborators

confirmed it experimentally in 1957.

Remark. As with any $\mathfrak{su}(2)$ doublet (Part C), the antisymmetric tensor $\varepsilon_{\alpha\beta}$ ($\varepsilon_{12} = -\varepsilon_{21} = 1$) is invariant under $SU(2)$ and hence lets one raise and lower left-handed spinor indices, $\chi^\alpha = \varepsilon^{\alpha\beta} \chi_\beta$; the analogous $\varepsilon_{\dot{\alpha}\dot{\beta}}$ does the same for right-handed indices. A full account of this index calculus, and its extension to spinor representations in higher spacetime dimensions, is the starting point of a different (and considerably longer) course and lies beyond our present scope.