

Geometry of Mathematical Physics III

(Michaelmas 2024-25)

Andreas Braun

`andreas.braun@durham.ac.uk`

Department of Mathematical Sciences, Durham University

January 8, 2025

Contents

Introduction	3
1 Lie Groups and Lie Algebras	9
1.1 Motivating Examples	9
1.1.1 The group $U(1)$: three easy pieces.	9
1.1.2 Review: matrices, indices, summation, and all that.	13
1.1.3 The group $SU(2)$: three easy pieces revisited	15
1.1.4 $SO(3)$ vs. $SU(2)$	25
1.2 Differentiable Manifolds	30
1.3 Lie groups	41
1.4 Lie algebras	43
2 Representations	50
2.1 Generalities and Basic Examples	50
2.2 Representations of $U(1)$	57
2.3 Representations of Lie algebras	59
2.4 Representations of $SU(2)$ and $SO(3)$	63
2.5 Representations of $SU(n)$ and other Lie Groups *	69
2.6 Tensor Representations*	70

Preliminaries

In these lectures we will explore a modern formulation of several aspects of fundamental physics which features symmetries and geometry. As we will see, making use of such mathematical concepts not only deepens our understanding of many classical topics, but paves the way to many of the great conquests of 20th century physics. Our point of view will mostly be that of action principles and classical field theories (or just the mechanics of point particles sometimes) and we will review those aspects that are needed as we go along. Sometimes, it will prove worthwhile to change perspective and consider quantum mechanics instead, but we will limit ourselves to a few crucial topics for this.

There are many more advanced topics that are outside the scope of these lectures, but that will come too close for us to completely ignore. The short detours into this terrain that we will undertake are meant to stimulate your intellectual curiosity, but do **not** form part of the examinable content. The corresponding sections will be clearly marked with a *.

These lectures combine material that naturally belongs together, but can rarely be found all in one source. Furthermore, many standard texts on the subjects covered pitch their material at a somewhat higher level. While I hope you take a look at the references given, don't get discouraged if what you find seems very difficult. These notes form the core of the material you should use in this course and the exercises given reflect the level of understanding I expect you to achieve. Only what is discussed in the lecture notes (and does not have a *) forms the examinable material.

Introduction*

In (classical) theoretical physics, we typically specify which kind of physical system we are talking about by stating the physical degrees of freedom, e.g. some generalized coordinates q_i and their derivatives \dot{q}_i with respect to ‘time’, which uniquely determine the configuration of our system, together with an action $S[q_i, \dot{q}_i]$ that encodes the dynamics. We will introduce actions formally later, so if you are unfamiliar with actions or feel a little rusty, nevermind, all that matters for this discussion is that the action is a functional of $q_i(t)$ and $\dot{q}_i(t)$ which determines the equations of motion by demanding that the equations of motion are extrema of S .

Defining systems using an action S a sensible¹ definition of what is a symmetry might be

Definition 0.1. Let g be an invertible map

$$g : \begin{aligned} q_i &\rightarrow g(q_i) \\ \dot{q}_i &\rightarrow g(\dot{q}_i) \end{aligned} \quad (0.1)$$

such that

$$S[g(q_i), g(\dot{q}_i)] = S[q_i, \dot{q}_i]. \quad (0.2)$$

Then g is called a *symmetry* of the action S . Note that this must hold true no matter what q_i and \dot{q}_i are.

Given an action we can ask about the set G of all of its symmetries (and this is something we will frequently do), and our definition above has two immediate consequences for that:

- i) The identity map $g = 1$ is a (rather trivial) symmetry.
- ii) For any two symmetries g and g' we can form another symmetry by composition:

$$S[g(g'(q_i)), g(g'(\dot{q}_i))] = S[g'(q_i), g'(\dot{q}_i)] = S[q_i, \dot{q}_i]. \quad (0.3)$$

Here it is crucial that S stays invariant no matter what q_i and \dot{q}_i are. In particular, it does not matter to apply our definition that we act with g on $g'(q_i), g'(\dot{q}_i)$ instead of q_i, \dot{q}_i .

¹We will be more precise with these things later.

If you consider the above overly formal, we can unpack what it means intuitively: if you can map the system using both g and g' without changing S , you might as well first act with g' , after which S did not change (read the above equation from right to left). Then we subsequently apply g , and again the action does not change. We hence acted with both maps without changing the action.

The properties of the set of symmetries we have just uncovered are exactly what is called a *group* (or rather, this is why groups are defined the way they are). Recall the following from Linear Algebra I:

Definition 0.2. A *group* is a set G equipped with an operation $\circ : G \times G \rightarrow G$ such that

- i) there is an identity element e in G such that $x \circ e = e \circ x = x$ for all $x \in G$;
- ii) if $x, y \in G$ then $x \circ y \in G$;
- iii) $(x \circ y) \circ z = x \circ (y \circ z)$ for all $x, y, z \in G$;
- iv) for each $x \in G$, there exists an inverse x^{-1} in G with $x^{-1} \circ x = e$;²

Proposition 0.1. The symmetries of a classical action form a group.

Proof. The composition \circ is just the composition of maps here and the identity e is the identity map $g = \mathbb{1}$. We just seen the identity is a symmetry and that the composition of symmetries is a symmetry as well, taking care of i) and ii). Composition of maps satisfies associativity (in each case we simply apply three maps after each other), so we also have iii). Finally, we have assumed that symmetries are invertible maps, so we also have iv). \square

Exercise 1. Let \mathbb{C} be the complex numbers and $\mathbb{C}^* = \mathbb{C} \setminus \{0\}$. Which of these is a group under addition? Which of these is a group under multiplication?

Note that the way we introduced groups did not depend on groups acting in a physics context. More generally, groups naturally appear as symmetries of whatever object you are interested in. Take a moment to think about how you would argue that the defining properties of a group are present for any symmetries that you can think of.

²This implies that $x \circ x^{-1} = e$ as well. Let $y = x \circ x^{-1}$. Then

$$y = y^{-1} \circ y \circ y = y^{-1} \circ x \circ x^{-1} \circ x \circ x^{-1} = y^{-1} \circ y = e.$$

Note how we made use of associativity here.

REMARK: In the last years it has become widely appreciated that one needs to rethink the notion of symmetry somewhat when discussing quantum systems with extended objects, and things become more complicated than what we discuss here. Here, a suitable notion of symmetry is different to ‘it is a transformations that leaves the action invariant’, which in turn leads to the possibility that symmetries are no longer groups, which is exactly what happens. This is an active field of research at present, here is a popular science article about the topic [Hartnett, 2023].

Symmetries have remarkable consequences on the physics in that they imply conservation laws via what is called ‘Noether’s theorem’. Understanding which symmetries an action has hence immediately gives us conserved quantities which we can use in turn to constrain the dynamics. Here is a famous example without which we would not exist: the ‘Standard Model of particle physics’ classically has a symmetry which implies ‘baryon number conservation’. This in turn implies that the proton cannot decay into any of the particles lighter than itself, as this would violate the conservation law. Without this restriction, protons could decay into positrons (the electron’s anti particle) and the world as we know it would end in a flash. So we can explain that the proton is stable by a symmetry! If you try to come up with an extension of the Standard Model you better be careful not to violate this symmetry³.

Instead of investigating if a given action has some group of symmetries, we can hence turn things around and try to construct actions symmetric under a given group G if the consequences of the symmetry match experiments. This point of view is precisely what people mean when they say that ‘we understand’ fundamental physics using symmetries. Imagine we have good reasons to write down an action

$$S[q] = S_1[q] + \alpha S_2[q] \quad (0.4)$$

with a parameter α , and measurements tell us that $\alpha = 0$. You will immediately write a publication and receive much fame if you can find a symmetry G which leaves S_1 invariant, but not S_2 . In this case we simply add ‘invariance under G ’ as a fundamental requirement which forces $\alpha = 0$.

Having established the relevance of groups we of course want to study them in more detail. Two questions that immediately come to mind (and which correspond to topics 1 and 2) are

³It turns out a *tiny* little bit is okay, and that there are quantum effects that in fact do this. This is good news, as this is what is needed for baryogenesis in the early universe, i.e. is needed to explain why there is matter but no antimatter in the universe.

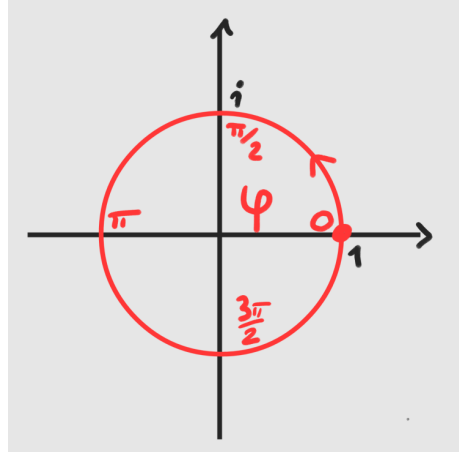


Figure 1: The group $U(1)$ is isomorphic to a circle with a distinguished point, the identity element.

‘**what are they like?**’ and ‘**how can they act?**’. To give you a feeling about the first question, consider the

Example 0.1. The group $U(1)$ is the group of complex numbers of unit modulus under multiplication. For $\varphi = 0..2\pi$ we can write any group element as

$$g = e^{i\varphi}. \quad (0.5)$$

However, it is not true that this group is isomorphic to the interval $[0..2\pi]$ as $\varphi = 0$ and $\varphi = 2\pi$ are one and the same group element. It is isomorphic to a circle S^1 , see figure 1. The fact that this group is topologically non-trivial will turn out to be the reason we can have magnetic monopoles in the second half of this course! In topic 1 we will examine several classes of groups that are also non-trivial spaces (‘manifolds’) in their own right, these are called ‘Lie groups’.

To appreciate the second question, consider a group that has elements which are $n \times n$ matrices (which ones is not really important here). If this group is supposed to act on the generalized coordinates in our theory, you might be tempted to say that they might be a vector with n components and the group acts as

$$\mathbf{q} \rightarrow M\mathbf{q}. \quad (0.6)$$

However, this is not the only possibility. We might say that the generalized coordinates in our theory are $n \times n$ matrices Q themselves and the group acts as

$$Q \rightarrow M^{-1}QM. \quad (0.7)$$

Note that set of $n \times n$ matrices is a vector space as well, but its dimension is n^2 , which is different from n . The general study of how groups can linearly act on vector spaces of various dimensions (and which dimensions can occur) is called ‘representation theory’ and will discuss

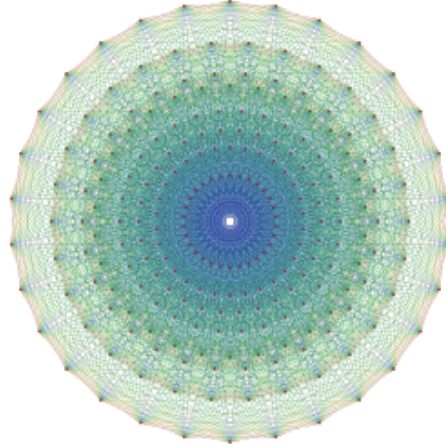


Figure 2: A projection of the root system of the Lie group E_8 , which can be used to construct its representations.

this in topic 2. As linear maps on vector spaces can always be written in terms of matrices, we can also say that representation theory is the question of how abstract relations such as

$$gg' = g'' \quad (0.8)$$

can be concretely realized using matrices. In figure 2 you can see a glimpse of the beautiful structures that emerge when asking (and answering) such questions.

Exercise 2. Consider the set S of real $n \times n$ matrices.

- a) Show that S this is a (real) vector space V of dimension n^2 .
- b) Let $U \subset S$ be the set of matrices with determinant 1. Is U a vector space as well?
- c) For any matrix Q in V define a map

$$g_M : Q \rightarrow M^{-1}QM$$

where M is a fixed invertible matrix. Show that g_M is a linear map on V .

In Epiphany term, we will start investigating what happens when we consider group actions that vary across space-time, i.e. we let

$$g = g(t, \vec{x}). \quad (0.9)$$

Such things are called gauge symmetries and form the underpinning for the interactions of the Standard Model of particle physics. Interestingly, demanding such symmetries forces us to have forces which are transmitted by ‘gauge bosons’ such as photons or gluons.

After spending some time formulating gauge theories as classical field theories, we will investigate the impact of the ‘gauge group’ G on the physics and will discover that the topology of G plays a central role.

Topic 1

Lie Groups and Lie Algebras

1.1 Motivating Examples

Before giving formal definitions of Lie groups, Lie Algebras, let us look at some motivating examples and explore their properties a bit. These examples will serve as templates for all there is to come, so make sure you understand them well¹.

1.1.1 The group $U(1)$: three easy pieces.

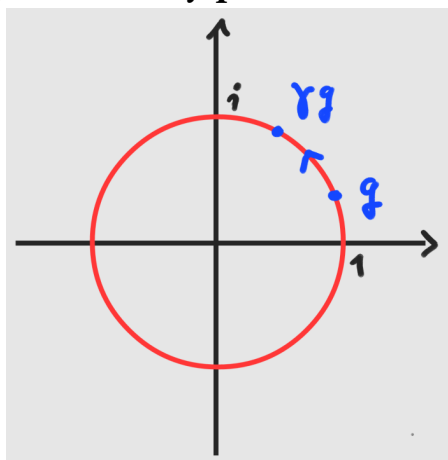


Figure 1.1: Group multiplication in $U(1)$.

¹The theoretical physicist Richard Feynman was famous for his approach of ‘example based research’: find a good example for what you want to study and understand it really well. Then develop the general theory such that the main features that ‘make it work’ are kept. Although it is not often presented like that, a lot of mathematics came about in this way.

1) The group $U(1)$ derives its name from **unitary** complex 1×1 matrices. Acting with g on a complex number c

$$z \rightarrow gz, \quad g \in U(1), \quad z \in \mathbb{C} \quad (1.1)$$

we require that the inner form $|c|^2$ remains unchanged. As

$$|z|^2 = \bar{z}z \rightarrow \bar{z}\bar{g}gz = |g|^2|z|^2 \quad (1.2)$$

this implies that g is a complex number of modulus one, $|g|^2 = 1$, so that we can write $g = e^{i\phi}$ (as before). Hence

Definition 1.1. $U(1)$ is the complex numbers of unit modulus.

Proposition 1.1. Using multiplication as the group operation, $U(1)$ is a group.

Proof. : The group operation can be written as

$$e^{i\phi}e^{i\phi'} = e^{i(\phi+\phi')}. \quad (1.3)$$

Let's check that this is a group indeed.

i) there is an identity element e in G such that $x \circ e = e \circ x = x$ for all $x \in G$;

\leftrightarrow

we simply use 1

ii) if $x, y \in G$ then $x \circ y \in G$;

\leftrightarrow

the product of any two complex numbers of unit modulus is again of unit modulus

iii) $(x \circ y) \circ z = x \circ (y \circ z)$ for all $x, y, z \in G$;

\leftrightarrow

multiplication of complex numbers is associative

iv) for each $x \in G$, there exists an inverse x^{-1} in G such that $x^{-1} \circ x = e$;

\leftrightarrow

for $g = e^{i\phi}$, $g^{-1} = e^{-i\phi}$.

Note that $gg' = g'g$ for all $g, g' \in U(1)$. This is a special property that has a name: 'abelian'. $U(1)$ is an example of an abelian group.

Definition 1.2. A group G is called **abelian** if $xy = yx$ for all elements x and y of G .

2) Already in the introduction we realized that the map from \mathbb{R} to $U(1)$ given by writing $g = e^{i\phi}$ is not a one-to-one map. If we are to do calculus on $U(1)$, we cannot simply do

calculus on \mathbb{R} and map that to $U(1)$ which is a circle. This is the easiest non-trivial example of what we call a manifold.

3) We already introduced $U(1)$ as acting on complex numbers as in (1.1). We might be interested in asking what happens when we perform an infinitesimal transformation, i.e. when ϕ is very close to 0. In this case we can approximate the exponential to linear order and get

$$z \rightarrow (1 + i\phi)z. \quad (1.4)$$

The approximation $e^{i\phi} = (1 + i\phi)$ is tangent to the group $U(1)$ at $g = 1$, see figure 1.2. We

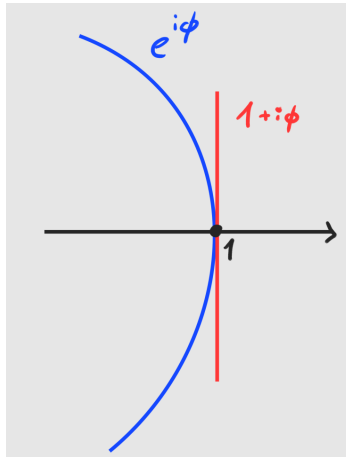


Figure 1.2: Infinitesimal transformations in $U(1)$ are tangent to the group at the identity element.

can try to reconstruct finite elements of $U(1)$ by successive infinitesimal transformations. Let us hence look at

$$(1 + i\phi)^2 = 1 + 2i\phi - \phi^2. \quad (1.5)$$

This fails to reproduce the expansion of the exponential

$$e^{i\phi} = 1 + i\phi + \frac{(i\phi)^2}{2} + \frac{(i\phi)^3}{3!} + \dots \quad (1.6)$$

but we can easily fix this by considering instead

$$(1 + i\phi/2)^2 = 1 + i\phi - \frac{\phi^2}{4}. \quad (1.7)$$

Now we at least have the linear term right, but the quadratic term is still way off. Now consider

$$(1 + i\phi/3)^3 = 1 + i\phi + \frac{(i\phi)^2}{3} + \frac{(i\phi)^3}{27}. \quad (1.8)$$

Again, we are forced to have the $1/3$ to get the linear term right and the quadratic term came closer to $1/2$. This continues as we consider $(1 + i\phi/n)^n$ for higher values of n , we always get the linear term right and the subsequent terms come closer to the expansion of the complex exponential. One can also understand the need for $1/n$ as follows: we are trying to reproduce a finite group element with phase ϕ by taking n consecutive infinitesimal group elements. For these to match up, we need the ‘phase’ of the infinitesimal group elements to be ϕ/n .

We might hence guess that we can recover a finite map from an infinitesimal one by looking at $(1 + i\phi/n)^n$ and letting n go to infinity, which turns out to be correct.

Proposition 1.2. $\lim_{n \rightarrow \infty} (1 + i\phi/n)^n = e^{i\phi}$

Proof. We can expand the powers to find

$$\begin{aligned} \lim_{n \rightarrow \infty} (1 + i\phi/n)^n &= \lim_{n \rightarrow \infty} \sum_{k=0}^n \frac{(i\phi)^k}{n^k} \binom{n}{k} \\ &= \lim_{n \rightarrow \infty} \sum_{k=0}^n \frac{(i\phi)^k}{k!} \frac{n!}{(n-k)!n^k} \end{aligned} \quad (1.9)$$

This already looks like the series of the exponential except for the factor

$$\frac{n!}{(n-k)!n^k} = \frac{n(n-1)(n-2) \cdots (n-k+1)}{n^k}. \quad (1.10)$$

There are exactly k factors in the numerator of this fraction which all approach n when $n \rightarrow \infty$. Hence this factor converges to 1 for any fixed k . In the sum, there are terms for which k approaches n so we cannot make the above approximation, but these terms multiply $(i\phi)^n/n!$ which is subleading to the rest of the expression when $n \rightarrow \infty$. In other words, for every n_0 we can choose a large enough n such that the first n_0 terms in the exponential series are approximated to any precision by the first n_0 terms in $(1 + i\phi/n)^n$. What this means is that

$$\lim_{n \rightarrow \infty} (1 + i\phi/n)^n = \sum_{k=0}^{\infty} \frac{(i\phi)^k}{k!} = e^{i\phi} \quad (1.11)$$

□

Exercise 3.

a) By working out the derivative of

$$\lim_{n \rightarrow \infty} (1 + i\phi/n)^n$$

with respect to ϕ , show that this expression satisfies the same differential equation as $e^{i\phi}$. You may assume that you can swap the order of the limit and taking the derivative.

As both functions have the same value at $\phi = 0$ this implies that they are equal by the uniqueness of solutions of ordinary differential equations.

- b) Consider a square matrix A and let $g = e^{iA}$, which is defined via the Taylor series of the exponential. Show that

$$\lim_{n \rightarrow \infty} (\mathbb{1} + iA/n)^n = e^{iA}.$$

Let's back up a bit and see what this example has taught us:

- a) We defined a continuous group by demanding that its action on complex numbers leaves the inner form invariant.
- b) This group does not have a one-to-one map to \mathbb{R} , it has a 'non-trivial topology' and is isomorphic to S^1 . Due to it being a group, it has something S^1 does not have: there is a special point, the identity element 1.
- c) We found that infinitesimal transformations are tangent to the group at the identity element. We can recover group elements (and in fact the whole group) by iterating infinitesimal transformations and taking a limit. This is the same as exponentiating the infinitesimal element.

1.1.2 Review: matrices, indices, summation, and all that.

For many of the things that follow, we'll need to be confident in manipulating expressions involving matrices by using index notation, so let's review this briefly. For a vector \mathbf{v} in \mathbb{R}^n we'll denote the n components (in the standard basis) by v_i with i running from 1 to n . Likewise, we denote the components of a square $n \times n$ matrix A by A_{ij} where i and j run from 1 to n . Here, the first index corresponds to the row and the second to the column of A .

Using this notation, we can write various familiar objects as follows

- scalar product: $\mathbf{v} \cdot \mathbf{w} = \sum_i v_i w_i$
- multiplication between a matrix and a vector: $\mathbf{w} = A\mathbf{v} \Leftrightarrow w_i = (A\mathbf{v})_i = \sum_j A_{ij} v_j$

- a matrix sandwiched between two vectors: $\mathbf{w}^T A \mathbf{v} = \sum_{ij} w_i A_{ij} v_j$
- matrix multiplication $C = AB \Leftrightarrow C_{ij} = (AB)_{ij} = \sum_k A_{ik} B_{kj}$
- the vector cross product in \mathbb{R}^3 : $(\mathbf{w} \times \mathbf{v})_i = \sum_{jk} \epsilon_{ijk} w_j v_k$ (recall $\epsilon_{ijk} = \pm 1$ if i, j, k are a cyclical/anti-cyclical permutation of 1,2,3 and zero else)

What makes such expressions easy to manipulate is that components of vectors, matrices, etc are just numbers, so we can freely rearrange them. Here are two examples:

$$\mathbf{v} \cdot \mathbf{w} = \sum_i v_i w_i = \sum_i w_i v_i = \mathbf{w} \cdot \mathbf{v} \quad (1.12)$$

or

$$\mathbf{v} \cdot (\mathbf{w} \times \mathbf{u}) = \sum_{ijk} \epsilon_{ijk} v_i w_j u_k = \sum_{ijk} \epsilon_{kij} u_k v_i w_j = \mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) \quad (1.13)$$

As you realize, whenever there is a sum in the above expressions, the sum runs over an index which appears twice. Of course one can consider objects which are not of this type, but this is the typical state of affairs for many ‘natural’ kinds of products² To save time, it is customary to use ‘**summation convention**’, i.e. to use the convention to sum whenever an index appears twice and drop the summation sign. If one wants to write an index twice that is *not* summed over, one then needs to specifically say what is going on.

Here are a few more things we’ll need:

Definition 1.3. The **transpose** of a matrix A is simply A with rows and columns swapped, i.e. we can write

$$(A^T)_{ij} := A_{ji} . \quad (1.14)$$

Definition 1.4. The **Hermitian conjugate** A^\dagger (‘A dagger’) of a matrix A is transpose T and complex conjugation applied at the same time:

$$A^\dagger := \bar{A}^T \Leftrightarrow (A^\dagger)_{ij} = \bar{A}_{ji} \quad (1.15)$$

Definition 1.5. The **trace** of a square matrix A is $\text{tr } A := A_{jj}$, the sum of its diagonal elements.

Definition 1.6. The **determinant** of a square matrix A is

$$\det A := \epsilon_{i_1, i_2, \dots, i_n} A_{1i_1} A_{2i_2} \cdots A_{ni_n} . \quad (1.16)$$

Here $\epsilon_{i_1, i_2, \dots, i_n} = 1$ for an even permutation of $\{1, 2, \dots, n\}$, $\epsilon_{i_1, i_2, \dots, i_n} = -1$ for an odd permutation of $1, 2, \dots, n$, and $\epsilon_{i_1, i_2, \dots, i_n} = 0$ else. Recall that an even/odd permutation of $\{1, 2, \dots, n\}$ is one that can be decomposed into an even/odd number of operations which swap two numbers only.

²The deeper reason for this is that all of these can be defined without making reference to a basis.

Note the use of summation convention in the last two definitions!

Proposition 1.3. The determinant has the useful properties:

a) $\det AB = \det A \det B$

b) $\det A^T = \det A$

Proof. see MATH1071: Linear Algebra I

Exercise 4. Show using index notation that

a) $(A + B)^T = A^T + B^T$

b) $(AB)^T = B^T A^T$

c) $\text{tr}(cA) = c \text{tr}(A)$

d) $\text{tr}(AB) = \text{tr}(BA)$

e) $\text{tr} A^T = \text{tr} A$

f) $\text{tr}(A + B) = \text{tr} A + \text{tr} B$

g) $(A\mathbf{v}) \cdot (B\mathbf{w}) = \mathbf{v} (A^T B) \mathbf{w}$

h) $\det A^\dagger = \overline{\det A}$

i) $\det cA = c^n \det A$

where A and B are complex $n \times n$ matrices, \mathbf{v} and \mathbf{w} are vectors with n components, and c is a number.

1.1.3 The group $SU(2)$: three easy pieces revisited

1) The group $SU(2)$ is the group of special unitary 2×2 matrices. Special refers to $\det g = 1$ and unitary means they keep the inner product (or canonical ‘length’) in \mathbb{C}^2 invariant when

multiplying a vector $\mathbf{z} \in \mathbb{C}^2$ by g :

$$\mathbf{z} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \rightarrow g\mathbf{z} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}. \quad (1.17)$$

The inner product

$$|\mathbf{z}|^2 = \bar{\mathbf{z}} \cdot \mathbf{z} = \bar{z}_1 z_1 + \bar{z}_2 z_2 \quad (1.18)$$

transforms as

$$|\mathbf{z}|^2 \rightarrow \bar{g} \begin{pmatrix} \bar{z}_1 \\ \bar{z}_2 \end{pmatrix} \cdot g \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \quad (1.19)$$

Let us write this out in components

$$\bar{g} \begin{pmatrix} \bar{z}_1 \\ \bar{z}_2 \end{pmatrix} \cdot g \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \bar{g}_{ij} \bar{z}_j g_{ik} z_k = \bar{g}_{ij} g_{ik} \bar{z}_j z_k = \bar{z}_j g_{ji}^\dagger g_{ik} z_k = \bar{\mathbf{z}}^T g^\dagger g \mathbf{z} = \mathbf{z}^\dagger g^\dagger g \mathbf{z}. \quad (1.20)$$

The above implies that we need $g^\dagger = g^{-1}$. We hence make the

Definition 1.7. $SU(2)$ is the group of complex 2×2 matrices g with $\det g = 1$ and $g^\dagger g = \mathbb{1}$.

Let us check that all of this makes sense, i.e. that this is a group indeed.

- i) there is an identity element e in G such that $x \circ e = e \circ x = x$ for all $x \in G$;

\leftrightarrow

we simply use $\mathbb{1}$ which satisfies $\mathbb{1}^\dagger \mathbb{1} = \mathbb{1}$ and $\det \mathbb{1} = 1$, so is part of $SU(2)$.

- ii) if $x, y \in G$ then $x \circ y \in G$;

\leftrightarrow

for two elements g, g' in $SU(2)$ we have $\det(gg') = \det g \det g' = 1$ and $(gg')^\dagger gg' = (g')^\dagger g^\dagger gg' = \mathbb{1}$.

- iii) $(x \circ y) \circ z = x \circ (y \circ z)$ for all $x, y, z \in G$;

\leftrightarrow

matrix multiplication is associative

- iv) for each $x \in G$, there exists an inverse x^{-1} in G such that $x^{-1} \circ x = e$;

\leftrightarrow

First note that as $g^\dagger g = \mathbb{1}$ g^\dagger plays the role of the inverse. The question hence is if g^\dagger is in $SU(2)$ whenever g is in $SU(2)$. We work out

$$\det g^\dagger = \det \bar{g}^T = \overline{\det g^T} = \overline{\det g} = 1. \quad (1.21)$$

and

$$(g^\dagger)^\dagger g^\dagger = gg^\dagger = \mathbb{1}. \quad (1.22)$$

With the operation of matrix multiplication we hence have defined a group. Note that even though we needed to work through some equations using properties of the † , we could have anticipated this by observing that we defined this group as a set of linear maps (to get associativity) that are a symmetry of the inner form on \mathbb{C}^2 .

2) Next, let us describe what $SU(2)$ looks like and how we can parametrize it. For any invertible complex 2×2 matrix

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad (1.23)$$

we can write

$$g^{-1} = \frac{1}{\det g} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \quad (1.24)$$

As $\det g = 1$, $g^{-1} = g^\dagger$ hence implies

$$\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} \bar{a} & \bar{c} \\ \bar{b} & \bar{d} \end{pmatrix} \quad (1.25)$$

i.e. the most general matrix in $SU(2)$ can be written as

$$g = \begin{pmatrix} a & b \\ -\bar{b} & \bar{a} \end{pmatrix} \quad (1.26)$$

and $\det g = 1$ implies $|a|^2 + |b|^2 = 1$. As a and b are complex numbers, we can write $a = x_1 + ix_2$ as well as $b = x_3 + ix_4$ and find

$$SU(2) : \{\mathbf{x} = (x_1, x_2, x_3, x_4) \in \mathbb{R}^4 | x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1\}. \quad (1.27)$$

This is the defining equation of a three-sphere³, i.e. $SU(2)$ is a space which looks like the three-sphere S^3 and $a = x_1 + ix_2$, $b = x_3 + ix_4$ define an embedding of S^3 into \mathbb{R}^4 .

3) In the $U(1)$ example we managed to write the whole group as the complex exponential of something simple, a real number, which generated the infinitesimal group elements. Let's try something similar here and write

$$g = e^{iA} \quad (1.28)$$

for a matrix A and a group element g of $SU(2)$. The exponential of a matrix is defined via the series

$$e^{iA} = \sum_{k=0}^{\infty} \frac{(iA)^k}{k!}. \quad (1.29)$$

³A sphere of dimension n is the set of points in \mathbb{R}^{n+1} for which $x_1^2 + x_2^2 + \dots + x_{n+1}^2 = 1$.

Let us first impose unitarity $g^\dagger = g^{-1}$ and see what this implies for A . As $(A^n)^\dagger = (A^\dagger)^n$ we have

$$g^\dagger = e^{-iA^\dagger}, \quad (1.30)$$

and as furthermore

$$g^{-1} = e^{-iA} \quad (1.31)$$

we find

$$A^\dagger = A. \quad (1.32)$$

Such matrices are called ‘Hermitian’ and they play an important role in quantum mechanics. You may be familiar with exponentials of Hermitian matrices giving unitary ones from there.

Next we investigate

$$\det g = \det e^{iA} \quad (1.33)$$

Proposition 1.4. For a 2×2 matrix A , we have $\det e^{iA} = e^{i \operatorname{tr} A}$.

Proof. : Let us first write

$$\begin{aligned} \det g &= \det e^{iA} = \det \lim_{n \rightarrow \infty} (\mathbb{1} + iA/n)^n = \lim_{n \rightarrow \infty} \det(\mathbb{1} + iA/n)^n \\ &= \lim_{n \rightarrow \infty} [\det(\mathbb{1} + iA/n)]^n \end{aligned} \quad (1.34)$$

We can write the determinant explicitly as

$$\det \left(\mathbb{1} + \frac{iA}{n} \right) = \left(1 + \frac{iA_{11}}{n} \right) \left(1 + \frac{iA_{22}}{n} \right) - \frac{i^2 A_{21} A_{12}}{n^2} = 1 + \frac{i \operatorname{tr} A}{n} + \dots, \quad (1.35)$$

where the dots stand for terms of order n^{-2} . In the limit $n \rightarrow \infty$ these terms are subleading and we hence have

$$\det g = \det e^{iA} = \lim_{n \rightarrow \infty} (1 + i \operatorname{tr} A/n)^n = e^{i \operatorname{tr} A}. \quad (1.36)$$

Exercise 5. For a general $k \times k$ matrix M show that

a) $\det e^M = e^{\operatorname{tr} M}$.

b) Use this to conclude that for $g = e^M$ we have $\log \det g = \operatorname{tr} \log g$. Here the \log of a matrix is defined as the inverse function of the exponential.

The requirement $\det g = 1$ now implies $e^{i \operatorname{tr} A} = 1$ which gives $\operatorname{tr} A = 0$ ⁴. If we can write $g \in SU(2)$ as a complex exponential, we hence have to use traceless hermitian 2×2 matrices

⁴We could also allow matrices with trace a multiple of 2π at this point. These can be ignored for the following reason: modulo the (traceless) Pauli matrices, any such matrix is of the form $n\pi \mathbb{1}$ for $n \in \mathbb{Z}$. This commutes with any other matrix and exponentiates to $\pm \mathbb{1}$. We hence get nothing new this way.

in the exponent. For $g \in SU(2)$ writing $g = e^{iA}$ implies that $A \in T$ where

$$T = \{A | A^\dagger = A, \text{tr} A = 0\} . \quad (1.37)$$

Whenever A, B are in T , then also $aA + bB$ for $a, b \in \mathbb{R}$ are in T : we have

$$\begin{aligned} \text{tr}(aA + bB) &= a \text{tr} A + b \text{tr} B = 0 \\ (aA + bB)^\dagger &= \bar{a}A^\dagger + \bar{b}B^\dagger = aA + bB . \end{aligned} \quad (1.38)$$

This means that T is a vector space over the real numbers. It is not too hard to convince yourself⁵ that T has dimension 3. This is a real vector space which contains (in general) complex matrices, so it really pays off to be able to think of vector spaces abstractly! We can make the following choice of basis vectors for T

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} , \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} , \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} . \quad (1.39)$$

These three matrices are known as the ‘Pauli matrices’⁶. By a direct computation one can work out that

Proposition 1.5. The Pauli matrices satisfy

$$[\sigma_i, \sigma_j] = 2i\epsilon_{ijk}\sigma_k \quad (1.40)$$

where $[a, b] \equiv ab - ba$ is the commutator.

Proof. :

Exercise 6. Show that

$$[\sigma_i, \sigma_j] = 2i\epsilon_{ijk}\sigma_k \quad (1.41)$$

where σ_i are the Pauli matrices.

Note in particular that different $g \in SU(2)$ in general do not commute, i.e. in general $g_1g_2 \neq g_2g_1$. Similarly, the Pauli matrices do not commute with each other, so that in general

$$\begin{aligned} e^{i\alpha_1\sigma_1}e^{i\alpha_2\sigma_2} &\neq e^{i\alpha_2\sigma_2}e^{i\alpha_1\sigma_1} \\ e^{i\alpha_1\sigma_1}e^{i\alpha_2\sigma_2} &\neq e^{i\alpha_1\sigma_1+i\alpha_2\sigma_2} \end{aligned} \quad (1.42)$$

$SU(2)$ is an example of a **non-abelian** Lie group.

⁵Try to write down the most general complex 2×2 matrix which obeys $A^\dagger = A$ and $\text{tr} A = 0$ in terms of real numbers.

⁶They are named after Wolfgang Pauli who introduced them in the 1920s to describe the spin of electrons. Why and how that works will be explained later.

We can form group elements of $SU(2)$ by exponentiating arbitrary real linear combinations of the Pauli matrices. Let $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ and write

$$g(\alpha) = e^{i\alpha_j \sigma_j} = e^{i\alpha \sigma} \quad \alpha_k \in \mathbb{R}. \quad (1.43)$$

As $\alpha_k \sigma_k$ is traceless and Hermitian for any real α , it follows that $g \in SU(2)$.

We hence get a group element for every vector $\alpha \in \mathbb{R}^3$. This map cannot possibly be injective as \mathbb{R}^3 is not the same as S^3 . To make contact with the earlier characterization $SU(2) \simeq S^3$ let us try to work out what kind of coordinates the α_k give us. As a warmup, let us first consider

$$g((\alpha_1, 0, 0)) = e^{i\alpha_1 \sigma_1} = \sum_{k=0}^{\infty} \frac{(i\alpha_1 \sigma_1)^k}{k!} = \sum_{k=\text{even}} \frac{(i\alpha_1 \sigma_1)^k}{k!} + \sum_{k=\text{odd}} \frac{(i\alpha_1 \sigma_1)^k}{k!} \quad (1.44)$$

As $\sigma_i^2 = \mathbb{1}$, all of the matrix powers in the sum over even k are equal to $\mathbb{1}$, and all of the matrix powers in the sum over odd k are equal to σ_1 . Hence

$$e^{i\alpha \sigma_1} = \sum_{k=\text{even}} \frac{(i\alpha_1)^k}{k!} \mathbb{1} + \sum_{k=\text{odd}} \frac{(i\alpha_1)^k}{k!} \sigma_1 = \cos(\alpha_1) \mathbb{1} + i \sin(\alpha_1) \sigma_1. \quad (1.45)$$

You can think of this as a generalization of Euler's formula $e^{i\phi} = \cos \phi + i \sin \phi$. Note in particular that $\alpha_1 + 2\pi$ maps to the same group element as α_1 , so we see the non-injectivity of the exponential map explicitly.

As σ_1 commutes with itself, we can write

$$g((\alpha_1, 0, 0))g((\alpha'_1, 0, 0)) = e^{i\alpha_1 \sigma_1} e^{i\alpha'_1 \sigma_1} = e^{i(\alpha_1 + \alpha'_1) \sigma_1} = g((\alpha_1 + \alpha'_1, 0, 0)) \quad (1.46)$$

so that matrices of this form are a subgroup of $SU(2)$:

Definition 1.8. For a group G , $H \subset G$ is called a **subgroup** if the elements of H form a group with the group composition of G .

Let us quickly check that elements of the form $g((\alpha_1, 0, 0))$ form a group in their own right (we already know all of them are in $SU(2)$). I copied the definition for you so we can check

- i) there is an identity element e in G such that $x \circ e = e \circ x = x$ for all $x \in G$;

\leftrightarrow

we simply use $\alpha = 0$ which gives $g((0, 0, 0)) = \mathbb{1}$

- ii) if $x, y \in G$ then $x \circ y \in G$;

\leftrightarrow

this works as $g((\alpha_1, 0, 0))g((\alpha'_1, 0, 0)) = g((\alpha_1 + \alpha'_1, 0, 0))$

iii) $(x \circ y) \circ z = x \circ (y \circ z)$ for all $x, y, z \in G$;

\leftrightarrow

matrix multiplication is associative

iv) for each $x \in G$, there exists an inverse x^{-1} in G such that $x^{-1} \circ x = e$;

\leftrightarrow

we have $g((\alpha_1, 0, 0))g((-\alpha_1, 0, 0)) = 1$

This subgroup ‘looks like’ or ‘works in the same way’ as $U(1)$ parametrized as $e^{i\phi}$ if we set $\alpha_1 = \phi$. A more precise way to say this is to use the word **group isomorphism**:

Definition 1.9. For two groups G and H , a **group homomorphism** is a map $f : G \rightarrow H$ such that

$$f(g_1 \circ_G g_2) = f(g_1) \circ_H f(g_2) \quad (1.47)$$

Note that in this definition we are using the group composition in G on the left side and the group composition in H on the right side. The map f is hence compatible with the group structures of G and H . Note further that this definition does not assume that f is injective or surjective.

Example 1.1. The map

$$g : x \rightarrow g(x) = e^x \quad (1.48)$$

is a group homomorphism from \mathbb{C} (with composition \circ_+ being addition) to \mathbb{C}^* (with composition \circ_* being multiplication). We can check that for every $x, y \in \mathbb{C}$ we find

$$g(x \circ_+ y) = g(x + y) = e^{x+y} = e^x e^y = g(x)g(y) = g(x) \circ_* g(y). \quad (1.49)$$

Saying that this is a homomorphism is just another way to express the ‘nice’ property of the exponential that $e^{x+y} = e^x e^y$.

Exercise 7. Let f be a homomorphism between two groups G and H . Show that

a) $f(e_G) = e_H$ where e_G and e_H are the unit elements of G and H , respectively.

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

Exercise 8. $U(2)$ is the group of complex 2×2 matrices g such that $g^\dagger = g^{-1}$, with the group composition being matrix multiplication. Let F be the map which sends

$$g \mapsto \det g. \quad (1.50)$$

Show that F is a group homomorphism from $U(2)$ to $U(1)$.

Definition 1.10. For two groups G and H , a **group isomorphism** is a map $f : G \rightarrow H$ which is one-to-one and a group homomorphism.

If two groups have a group isomorphisms, they are the same: they have the same elements that have the same group composition rule, i.e. in the present case we can identify

$$e^{i\phi} \leftrightarrow g((\phi, 0, 0)), \quad (1.51)$$

using the earlier presentation of $U(1)$. As the composition rule of the elements $g((\alpha_1, 0, 0))$ is the same as that of elements of $U(1)$ this is a group isomorphism. It is hence fair to say that there is a $U(1)$ sitting inside of $SU(2)$.

We can summarize the above observation as

Proposition 1.6. There is an injective group homomorphism $U(1) \rightarrow SU(2)$, the image of which is a $U(1)$ subgroup of $SU(2)$.

A similar computation shows that the same works not just for α of the form $\alpha = (\alpha_1, 0, 0)$ but for any subset of α s that can be written as $\alpha = t\alpha_0$, so that there are in fact infinitely many $U(1)$ subgroups of $SU(2)$. These deserve a special name:

Definition 1.11. A subgroup G_α of $SU(2)$ whose elements are of the form

$$G_\alpha = \{e^{it\alpha\sigma} | t \in \mathbb{R}\} \quad (1.52)$$

for some fixed α is called the **one-parameter subgroup generated by α** .

What is nice about the parametrization in terms of exponentials of matrices is that we can easily work out all infinitesimal elements. When α is very small, we can approximate by

$$g(\alpha) \simeq \mathbb{1} + i\alpha_j\sigma_j = \mathbb{1} + i\alpha \cdot \sigma. \quad (1.53)$$

Equivalently, the space of tangent vectors at $g = \mathbb{1}$ is spanned by the σ_j (times i), see figure 1.3 and a general vector is written as $i\alpha \cdot \sigma$. Convince yourself that this is indeed a vector space. As before, we can get back elements of $SU(2)$ by an infinite iteration of infinitesimal elements:

$$g(\alpha) = \lim_{n \rightarrow \infty} (\mathbb{1} + i\alpha \cdot \sigma/n)^n = e^{i\alpha\sigma}. \quad (1.54)$$

What is more, we can think of recovering any infinitesimal generator $i\alpha\sigma$ by considering a path

$$s(t) = e^{it\alpha\sigma} \quad (1.55)$$

in $SU(2)$ and taking a derivative w.r.t t evaluated at $t = 0$ (which corresponds to $\mathbb{1} \in SU(2)$):

$$\left. \frac{\partial s(t)}{\partial t} \right|_{t=0} = \left. \frac{\partial}{\partial t} e^{it\alpha\sigma} \right|_{t=0} = i\alpha\sigma. \quad (1.56)$$

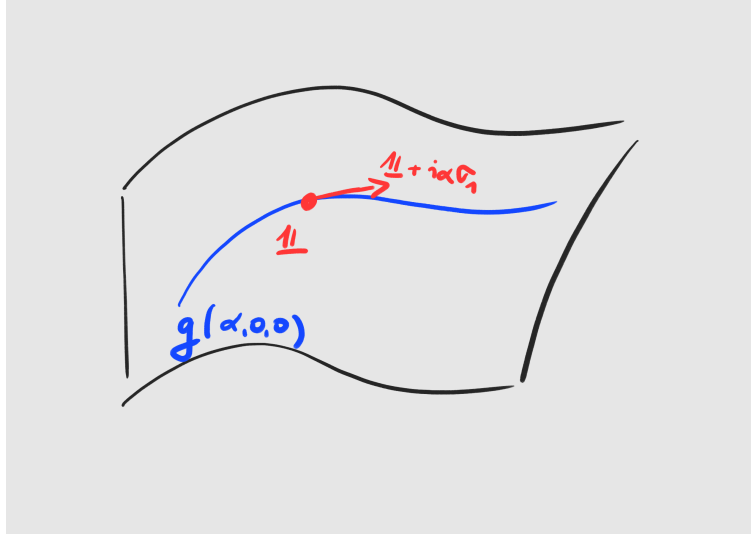


Figure 1.3: For the path $g((t\alpha_1, 0, 0))$ in S^3 , taking a derivative w.r.t t at $\mathbb{1}$ produces a tangent vector $\alpha_1\sigma_1$.

A natural question that is hopefully on your mind is the surjectivity of the map from $i\alpha\sigma$ to elements of $SU(2)$. Can we reach any element via the exponential? We can approach this by brute force and start working out

Exercise 9. a) Show that

$$\sigma_i\sigma_j + \sigma_j\sigma_i = 2\delta_{ij}\mathbb{1} \quad (1.57)$$

where σ_i are the Pauli matrices.

b) Show that

$$g(\alpha) = e^{i\alpha\sigma} = \begin{pmatrix} \cos(a) + i\sin(a)a_3/a & \sin(a)a_2/a + i\sin(a)a_1/a \\ -\sin(a)a_2/a + i\sin(a)a_1/a & \cos(a) - i\sin(a)a_3/a \end{pmatrix} \quad (1.58)$$

where $a = \sqrt{\alpha_1^2 + \alpha_2^2 + \alpha_3^2}$. [hint: write $\alpha = a\mathbf{n}$ with $|\mathbf{n}|^2 = 1$, i.e. $n_j = \alpha_j/a$]

As expressions of the type $AB + BA$ appear frequently they are given a special name:

Definition 1.12. For two matrices A, B , the **anti-commutator** $\{.,.\}$ is

$$\{A, B\} = AB + BA. \quad (1.59)$$

The remaining question is if we can choose an α that maps to any element of $SU(2)$. We will examine if every point on the S^3 that is $SU(2)$ is in the image of this map. We write a general group element as

$$g = \begin{pmatrix} x_1 + ix_2 & x_3 + ix_4 \\ -x_3 + ix_4 & x_1 - ix_2 \end{pmatrix} \quad (1.60)$$

where $x_i \in \mathbb{R}$ subject to $x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1$. First note that $x_1 \in [-1, 1]$. Fixing x_1 gives a slice of S^3 that is a two-sphere of radius $\sqrt{1 - x_1^2}$. Comparing to the form of $g(\alpha)$ we can always choose a such that $\cos(a) = x_1$. Note that this is not unique, but there always exists an a such that this is satisfied for any x_1 . The other variables are mapped as

$$\begin{pmatrix} x_2 \\ x_3 \\ x_4 \end{pmatrix} = \frac{\sin(a)}{a} \begin{pmatrix} \alpha_3 \\ \alpha_2 \\ \alpha_1 \end{pmatrix}. \quad (1.61)$$

The question is now if we can always find α such that this equation is satisfied for every x_1, x_2, x_3, x_4 subject to $x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1$. As we have already set $\cos(a) = x_1$, this fixes the length of α to be

$$\alpha^2 = \alpha_1^2 + \alpha_2^2 + \alpha_3^2 = a^2 \quad (1.62)$$

which is a two-sphere as well. The points on this two-sphere are mapped to the two-sphere

$$x_2^2 + x_3^2 + x_4^2 = 1 - x_1^2 \quad (1.63)$$

by (1.61), which is a one-to-one map. This is consistent as

$$x_2^2 + x_3^2 + x_4^2 = 1 - x_1^2 = 1 - \cos(a)^2 = \sin(a)^2 = \frac{\sin(a)^2}{a^2} (\alpha_1^2 + \alpha_2^2 + \alpha_3^2). \quad (1.64)$$

Hence we have shown that

Theorem 1.1. Every element of the group $SU(2)$ can be written as

$$g(\alpha) = e^{i\alpha \cdot \sigma} \quad (1.65)$$

where σ_j are a basis for infinitesimal transformations (which can be chosen as the matrices (1.39)) and $\alpha_j \in \mathbb{R}$.

Exercise 10. Let G be the set of complex 2×2 matrices of the form

$$g = \begin{pmatrix} \alpha & \beta \\ -\bar{\beta} & \bar{\alpha} \end{pmatrix}$$

for $\alpha, \beta \in \mathbb{C}$ and $|\alpha|^2 + |\beta|^2 \neq 0$.

- a) Show that G is a group using matrix multiplication as the group operation.
- b) Show that $SU(2)$ is a subgroup of G .
- c) Show that $V := \{\gamma | g = e^{i\gamma} \in G\}$ is a vector space (when ignoring the multi-valuedness of the logarithm) and find a basis for V .

Let's back up again and see what this example has taught us:

- a) We defined a continuous group by demanding that its action on vectors in \mathbb{C}^2 leaves the inner form invariant.
- b) This group does not have a one-to-one map to \mathbb{R}^3 , it has a 'non-trivial topology' and can be identified with S^3 . Again there is the special point $\mathbb{1}$ lying on this sphere.
- c) We found that infinitesimal transformations are tangent to the group at the identity element. We can recover group elements by iterating infinitesimal transformations and taking a limit. This is the same as exponentiating the infinitesimal element and gives us a surjective map to $SU(2)$. This is quite nice, as it means we can do all of our computations using the algebra of Pauli matrices σ_i instead of the group $SU(2)$.

1.1.4 $SO(3)$ vs. $SU(2)$

Definition 1.13. The group $SO(3)$ is the group of real 3×3 matrices S such that $S^T = S^{-1}$ and $\det S = 1$.

In this example, we will explore the global structure of $SO(3)$ by examining a clever map from $SU(2)$ to $SO(3)$. Let's take the components of \mathbf{v} and rearrange them in a 2×2 matrix $M_{\mathbf{v}}$:

$$M_{\mathbf{v}} = \begin{pmatrix} v_3 & v_1 - iv_2 \\ v_1 + iv_2 & -v_3 \end{pmatrix}. \quad (1.66)$$

This is just a funny way of writing \mathbb{R}^3 . Note that addition of vectors in \mathbb{R}^3 becomes addition of matrices $M_{\mathbf{v}}$, and multiplication of vectors in \mathbb{R}^3 by a real number c becomes multiplication of $M_{\mathbf{v}}$ by c .

Now consider

$$F(g) : M_v \rightarrow gM_v g^\dagger = F(g)[M_v] \quad (1.67)$$

for $g \in SU(2)$. What this means is that for every g in $SU(2)$, we get a map $F(g)$ acting on \mathbb{R}^3 .

Proposition 1.7. F is a group homomorphism from $SU(2)$ to $SO(3)$.

Proof. :

First of all, you might wonder if the map $F(g)$ preserves the structure of the matrix M_v . Note that M_v is the most general complex 2×2 matrix with the property that $M_v^\dagger = -M_v$ and $\text{tr } M_v = 0$. Under $F(g)$ it is mapped to $gM_v g^\dagger$ which also obeys

$$(gM_v g^\dagger)^\dagger = (g^\dagger)^\dagger M_v^\dagger g = gM_v g^\dagger \quad (1.68)$$

and

$$\text{tr } (gM_v g^\dagger) = g_{ij}(M_v)_{jk} g_{ki}^\dagger = g_{ki}^\dagger g_{ij}(M_v)_{jk} = \text{tr } (g^\dagger g M_v) = \text{tr } M_v = 0 \quad (1.69)$$

as $g^\dagger g = \mathbb{1}$ as $g \in SU(2)$. Next notice that $F(g)$ is a linear map on the coords of v_i of \mathbb{R}^3 as

$$g(M_v + M_{v'})g^\dagger = gM_v g^\dagger + gM_{v'} g^\dagger \quad (1.70)$$

This means that we could write F as

$$F(g) : \mathbf{v} \rightarrow \Phi(g)\mathbf{v} \quad (1.71)$$

for some 3×3 matrix $\Phi(g)$.

Next we observe that $\mathbf{v} \cdot \mathbf{v} = -\det M_v$, which is mapped under F as follows:

$$\det M_v \rightarrow \det gM_v g^\dagger = \det g \det M_v \det g^{-1} = \det M_v. \quad (1.72)$$

Hence this map leaves \mathbf{v}^2 invariant. As it is also a linear map, it must hence map to $O(3)$. The map $F(g)$ is a continuous map which maps $\mathbb{1} \in SU(2)$ to $\mathbb{1} \in O(3)$. As $SU(2)$ is connected (it is S^3), it only maps to the elements of $O(3)$ connected with the identity, i.e. to $SO(3)$.

For each $g \in SU(2)$ we hence have a way to assign an element of $SO(3)$ by sending g to $F(g)$, or equivalently $\Phi(g)$. And even better, composing two maps in $SO(3)$ is the same as composing two elements in $SU(2)$:

$$F(g'g)(M_v) = g'gM_v g^\dagger g'^\dagger = F(g') [F(g)[M_v]] \quad (1.73)$$

so that F is a group homomorphism.

REMARK: This homomorphism is not injective (it is not a group isomorphism), we have

$$F(g) = F(-g). \quad (1.74)$$

as in both cases we act in the same way on \mathbb{R}^3 :

$$\begin{aligned} F(g) : M_v &\rightarrow g M_v g^\dagger \\ F(-g) : M_v &\rightarrow (-g) M_v (-g)^\dagger = g M_v g^\dagger \end{aligned} \quad (1.75)$$

The element $g = -\mathbb{1}$ is in its kernel⁷,

$$F(-\mathbb{1}) [M_v] = -\mathbb{1} M_v (-\mathbb{1})^\dagger = M_v \quad (1.76)$$

which is $\mathbb{1} \in SO(3)$.

You might wonder if the group homomorphism from $SU(2)$ to $SO(3)$ is surjective. Consider the following simple example of how a one-parameter subgroup of $SU(2)$ is mapped:

$$g(0, 0, \theta) = \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix} = e^{i\theta\sigma_3}. \quad (1.77)$$

We have

$$\begin{aligned} g(0, 0, \theta) M_v g(0, 0, \theta)^\dagger &= \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix} \begin{pmatrix} v_3 & v_1 - iv_2 \\ v_1 + iv_2 & -v_3 \end{pmatrix} \begin{pmatrix} e^{-i\theta} & 0 \\ 0 & e^{i\theta} \end{pmatrix} \\ &= \begin{pmatrix} v_3 & e^{2i\theta}(v_1 - iv_2) \\ e^{-2i\theta}(v_1 + iv_2) & -v_3 \end{pmatrix} \end{aligned} \quad (1.78)$$

As

$$\begin{aligned} e^{2i\theta}(v_1 - iv_2) &= (\cos(2\theta) + i\sin(2\theta))(v_1 - iv_2) \\ &= v_1 \cos(2\theta) + v_2 \sin(2\theta) - i(v_2 \cos(2\theta) - v_1 \sin(2\theta)) \end{aligned} \quad (1.79)$$

this map sends

$$\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \mapsto \begin{pmatrix} v_1 \cos(2\theta) + v_2 \sin(2\theta) \\ v_2 \cos(2\theta) - v_1 \sin(2\theta) \\ v_3 \end{pmatrix} = \begin{pmatrix} \cos(2\theta) & \sin(2\theta) & 0 \\ -\sin(2\theta) & \cos(2\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}. \quad (1.80)$$

In other words

$$\Phi(g(0, 0, \theta)) = \begin{pmatrix} \cos(2\theta) & \sin(2\theta) & 0 \\ -\sin(2\theta) & \cos(2\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (1.81)$$

⁷For a group homomorphism, the kernel are those elements send to the identity element.

This g hence maps to a rotation around the v_3 axis by -2θ .⁸ Similarly, one can show that rotations around other axes are generated by using other group elements of $SU(2)$; e.g. for a rotation around the axis v_0 simply use $e^{i\theta v_0 \cdot \sigma}$.

If we can write any element of $SO(3)$ as a composition of rotations around fixed axis, such as v_1 , v_2 and v_3 , we have hence proven that the homomorphism from $SU(2)$ to $SO(3)$ is surjective. This is indeed true as the following proposition shows:

Proposition 1.8. Every element R of $SO(3)$ can be written as a product of three rotations $R_{v_i}(\phi_i)$ around fixed axis v_i by angles ϕ_i : $R = R_{v_1}(\phi_1)R_{v_2}(\phi_2)R_{v_3}(\phi_3)$.

Proof. : see [Sternberg, 1995]. You can also find a proof of this in many texts on mechanics.

Theorem 1.2. There is surjective group homomorphism from $SU(2)$ to $SO(3)$. It is not injective, and its kernel is the cyclic group with two elements: $\mathbb{Z}_2 = \{1, -1\} \in SU(2)$. As there are exactly two points in $SU(2)$ that are mapped to each point in $SO(3)$, this map is called a (double) covering map. We can write

$$SO(3) \simeq SU(2)/\mathbb{Z}_2 = S^3/\mathbb{Z}_2$$

where the \mathbb{Z}_2 acts by identifying antipodal points on the three-sphere.

Proof. : We have shown above that it is a non-injective homomorphism with kernel $\{1, -1\}$ and that any rotation around a fixed axis is in the image. As any element in $SO(3)$ can be written as a product of three rotations, any element of $SO(3)$ is in the image of this homomorphism, it is surjective. We have also seen that $SU(2)$ is isomorphic to S^3 . Antipodal points on S^3 correspond to sending $x_i \rightarrow -x_i$ for any solution to

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1. \quad (1.82)$$

As

$$g = \begin{pmatrix} x_1 + ix_2 & x_3 + ix_4 \\ -x_3 + ix_4 & x_1 - ix_2 \end{pmatrix} \quad (1.83)$$

g and $-g$ are hence antipodal points in $SU(2)$. We have seen that g and $-g$ are mapped to the same element of $SO(3)$. If we hence identify g and $-g$ in $SU(2)$, this group homomorphism becomes a group isomorphism. As the map from $SU(2)$ to $SO(3)$ is surjective, we conclude that $SO(3) \simeq SU(2)/\mathbb{Z}_2 = S^3/\mathbb{Z}_2$. \square

⁸Recall that θ going from 0 to 2π was a full $U(1)$ inside $SU(2)$. Under this map this is mapped to a rotation that goes with double speed from 0 to 4π . This is how it had to be such that -1 in $SU(2)$, which is $\theta = \pi$, is mapped to 1 in $SO(3)$ and is a consequence of the map F being two-to-one.

REMARK: * One may be forgiven to think that the group $SO(3)$ of rotations in \mathbb{R}^3 is a two-sphere by imagining all the positions that a vector v in \mathbb{R}^3 can be rotated to. This is NOT true, however, as there are non-trivial rotations that leave the chosen vector v invariant. The subgroup of $SO(3)$ that leaves any $v \neq 0$ invariant is a $U(1) = S^1$ which miraculously combines with S^2 to form S^3/\mathbb{Z}_2 . Studying the same for the double cover $SU(2)$ reveals that S^3 is in fact a fibration of S^1 over S^2 . This is called the **Hopf fibration** and is very pretty.

The topology of $SU(2)$ vs. that of $SO(3)$ *

As a final observation, let us examine the topology of those two groups by considering closed loops, i.e. continuous maps $\phi : [0, 1] \rightarrow G$ such that $\phi(0) = \phi(1)$. For $SU(2)$ there are no non-trivial such maps: any closed loop in S^3 can be shrunk to a point. Such spaces are called **simply connected**.

Now consider a path going from 1 to -1 in $SU(2)$. Under F , this maps to a closed path in $SO(3)$ that starts and ends at 1 . Let us see if we can shrink this curve in $SO(3)$. If we continuously deform this curve, it will still lift to an open curve in $SU(2)$, although now it may go from g to $-g$ in $SU(2)$. But this means there is no way of shrinking it! Hence $SO(3)$ is not simply connected. If we consider looping twice around any loop in $SO(3)$, we can lift to a closed curve in $SU(2)$, which we already know can be collapsed. We have hence shown that the fundamental group of $SO(3)$ contains a \mathbb{Z}_2 element (in fact, this is the whole fundamental group). For a given manifold with non-trivial fundamental group, there is a unique way to find a covering space (called universal cover) that is simply connected: $SU(2)$ is the universal cover of $SO(3)$.

In this example we have seen:

- a) Another continuous group defined by demanding that its action on vectors in \mathbb{R}^3 leaves the inner form invariant.
- b) This group has an even more interesting topology; it is isomorphic to S^3/\mathbb{Z}_2 .
- c) Curiously, we found this via a surjective homomorphism from $SU(2)$ to $SO(3)$. This in particular gave us an action of $SU(2)$ on vectors in \mathbb{R}^3 instead of the usual action on \mathbb{C}^2 . This came at the price that g and $-g$ act in the same way though.

In the motivating examples, we have mostly used a fairly pedestrian approach, but have already discovered many interesting things. In the following, these will be appropriately formalized and generalized.

1.2 Differentiable Manifolds

The groups we have investigated in our motivating examples were fundamentally different from vector spaces. Whereas we could cover parts using (subsets) of \mathbb{R}^n , we could not find one-to-one maps to these as a whole. As you might anticipate, such a behaviour is not exclusive to continuous groups, but gives rise to the more general notion of a differentiable manifold. In this section, we will introduce these objects and give some more elementary examples. The basic idea is that a differentiable manifold X can be covered by open sets for each of which we can find a continuous one-to-one map to an open set sitting in a vector space, i.e. X is sewn together from things that we know how to handle.

For ease of exposition and in order not to venture too far into the realms of topology⁹ we will restrict ourselves to manifolds that are given to us as subsets of \mathbb{R}^m , i.e. submanifolds of \mathbb{R}^m . If you like you can try to think about how what we are saying here might be generalized, or take a look at e.g. [Nakahara, 2003]. Considering subsets of \mathbb{R}^m is exactly the setup we need to study when talking about continuous groups of matrices. The set of all $k \times k$ matrices is isomorphic to \mathbb{R}^m with $m = k^2$, so that those $k \times k$ matrices forming some given group G will naturally sit inside \mathbb{R}^m .

Let us start by reviewing the notions of open and closed sets in \mathbb{R}^m . An open ball in \mathbb{R}^m centred at \mathbf{p} is the set

$$B_r(\mathbf{p}) = \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x} - \mathbf{p}\|^2 < r^2\}. \quad (1.84)$$

Using this, we can define open and closed sets as

Definition 1.14. A subset U of \mathbb{R}^n is **open** if for every point p in U there is an r such that $B_r(\mathbf{p})$ is fully contained in U .

Definition 1.15. A subset U of \mathbb{R}^n is **closed** if its complement $\mathbb{R}^n \setminus U$ is open.

Note that not every subset of \mathbb{R}^m has to be either closed or open. Furthermore, these properties are not mutually exclusive. Take some time to think of some examples that are open but not closed, closed but not open, not open and not closed, or both open and closed. Defining what we mean by open and closed subsets of a given set is called defining a **topology**, and the above is called the **standard topology** of \mathbb{R}^m .

Exercise 11. Which of the following sets are closed in the standard topology of \mathbb{R}^m ? Which are open?

⁹For a lively introduction to elementary topology I can recommend the book [Jänich and Levy, 1995].

- a) $\{0 < x < \pi\} \subset \mathbb{R}$ with coordinate x
- b) $\{x_1 < -2\} \subset \mathbb{R}^2$ with coordinates (x_1, x_2)
- c) $\{0 < x \leq \pi\} \subset \mathbb{R}$
- d) $\{0 < x_1 < 1\} \subset \mathbb{R}^2$ with coordinates (x_1, x_2)
- e) $\mathbb{R}^n \subseteq \mathbb{R}^n$
- f) $\{(x_1, x_2) \in \mathbb{R}^2 \mid x_1^2 \leq 42 - x_2^2\} \subset \mathbb{R}^2$ with coordinates (x_1, x_2)
- g) $\{(x_1, x_2, x_3) \mid x_1^2 + x_2^2 = 1\} \subset \mathbb{R}^3$

An important property of the notion of openness is that

Proposition 1.9. Arbitrary unions and finite intersections of open sets in \mathbb{R}^m are open again. For closed sets of \mathbb{R}^m it consequently works the opposite way: arbitrary intersections and finite unions are closed again.

Proof. :

Exercise 12. Prove that arbitrary unions and finite intersections of open sets in \mathbb{R}^m are again open. Why is the intersection of an infinite number of open sets not open in general ?

We now want to be able to talk about spaces X sitting inside \mathbb{R}^m . First we need to know what are open sets of X , i.e. introduce a topology. This is easy: we can simply inherit the notions of open and closed from \mathbb{R}^m :

Definition 1.16. For a subset X of \mathbb{R}^m we define the **induced topology** by declaring that $V \subset X$ is open if $V = U \cap X$ with U open in \mathbb{R}^m . Closed sets of X are then defined as complements (in X) of open sets of X . This turns X into a **topological space**.

Example 1.2. Let us consider $S^1 \subset \mathbb{R}^2$ defined by $x_1^2 + x_2^2 = 1$. By intersecting S^1 with open balls we obtain open segments on S^1 . Writing $x_1 = \cos \phi$, $x_2 = \sin \phi$, these are all of the form $\phi_1 < \phi < \phi_2$ for some ϕ_1, ϕ_2 . Of course this is what one would have naively considered open sets anyway. Arbitrary unions and finite intersections of these are then again open, as is \emptyset and all of S^1 .

Given the notion of a topological space allows us to define what we mean by continuous:

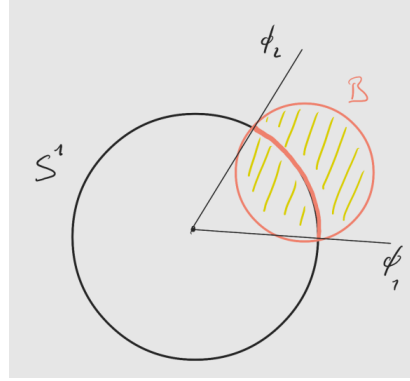


Figure 1.4: An open set of $S^1 \subset \mathbb{R}^2$ obtained by intersecting with an open Ball. Note that this intersection is not an open set in \mathbb{R}^2 .

Definition 1.17. A map $f : X \rightarrow Y$ between topological spaces X and Y is called **continuous** if the set $f^{-1}(U)$ is open in X whenever U is open in Y .

Similarly, a map $f : U_X \rightarrow U_Y$ between open sets $U_X \subset X$ and $U_Y \subset Y$ is called **continuous** if for all $V \subseteq U_Y$ the set $f^{-1}(V)$ is open in X whenever V is open in Y .

For maps $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ and using the standard topology, this agrees with the usual $\epsilon\delta$ definition from analysis.

Definition 1.18. A one-to-one map $f : X \rightarrow Y$ between topological spaces X and Y is called a **homeomorphism** if both f and f^{-1} are continuous.

These are the maps that preserve the structure of topological spaces.

We are now ready to define differentiable manifolds sitting inside \mathbb{R}^m .

Definition 1.19. A subset X of \mathbb{R}^m with the induced topology is an n -dimensional **differentiable manifold** if the following conditions are met

- i) X is covered by open sets $U_i \subseteq X$ and homeomorphisms ϕ_i that map U_i to an open subset $\phi_i(U_i)$ of \mathbb{R}^n . These are called **coordinate charts** or **patches**. The collection of patches (U_i, ϕ_i) is called an **atlas**.
- ii) We only need countably many U_i to cover all of X .
- iii) The **coordinate changes** $\phi_i \circ \phi_j^{-1}$ and their inverses $\phi_j \circ \phi_i^{-1}$ are C^∞ ('smooth') in their domains, i.e. they are continuous one-to-one maps that have infinitely many continuous derivatives.

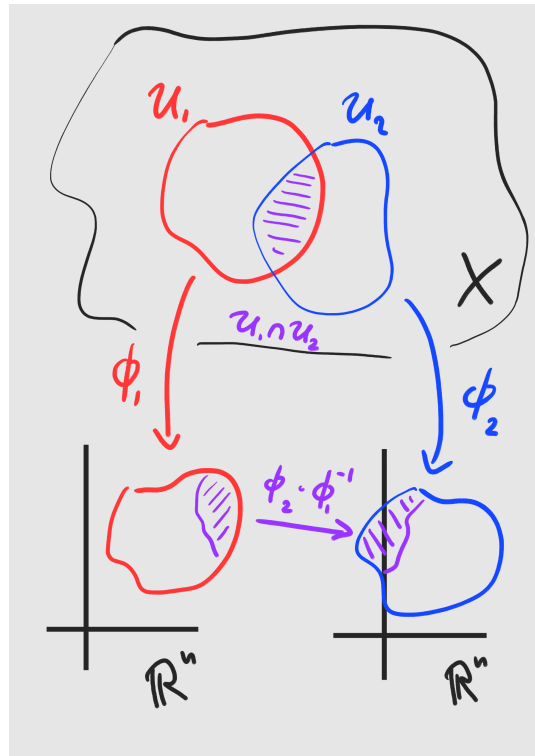


Figure 1.5: Using the ϕ_i we can make a map of X using open subsets of \mathbb{R}^n .

The property that we can cover X by open sets, each of which 'looks like' an open set in \mathbb{R}^n poses certain restrictions on what X can look like. For example, $X = \{xy = 0\} \subset \mathbb{R}^2$, the union of two lines $x = 0$ and $y = 0$ meeting at the origin, is not a manifold. Using the topology

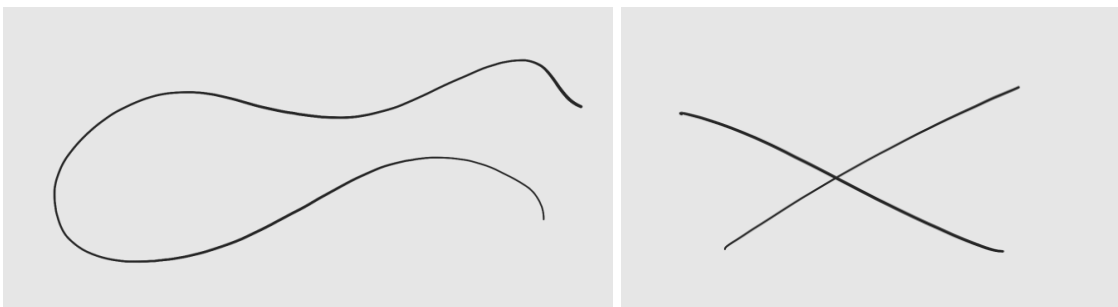


Figure 1.6: The shape on the left is a manifold, the one on the right is not.

induced from \mathbb{R}^2 , there is no issue to define coordinates away from the point $(y, x) = (0, 0)$, we just cut out a little branch and map it to an open set in \mathbb{R} . However any open set U containing the point $(y, x) = (0, 0)$ also contains (a small piece at least from) both branches. Hence these open sets look like a cross, which is radically different from any open subset of \mathbb{R} . There

cannot be any homeomorphism to an open subset of \mathbb{R} for such a U .

We can make a slightly more detailed argument about why that is as follows: choose a point p_a on the line $x = 0$, and an open interval on $xy = 0$ which connects it to $(0, 0)$, and then to a second point p_b on the line $x = 0$ beyond $(0, 0)$. Using that we want a continuous map to \mathbb{R} , this interval must be mapped to an open interval in \mathbb{R} and $(0, 0)$ goes to $0 \in \mathbb{R}$ (say). The image of the interval on one branch gives us an open interval in \mathbb{R} . Its inverse image must be an open set as well, as we need our coordinate map to be a homeomorphism. The open sets containing $(0, 0)$ all contain points on the other branch as well, so it needs to be mapped to our interval $\subset \mathbb{R}$ as well. But this cannot be as we need a 1-1 map. Note that this problem disappears as soon as you either drop that our map and its inverse are continuous, or that it is 1-1.

Exercise 13. Consider the sets of points in \mathbb{R}^2 with coordinates (x, y) defined implicitly by the following relations

a) $y = x^3$

b) $xy = c$

c) $x^2 + y^4 = 0$

d) $x \geq y$

e) $y^2 + x^3 - 3x - c = 0$

Using the induced topology from \mathbb{R}^2 , decide in each case if it is a differentiable manifold. For relations containing a constant c , decide for which value of c it is a differentiable manifold.

Example 1.3. Let us try to see how we can make a circle into a differentiable manifold. We already saw that we can write

$$g = e^{i\psi} \tag{1.85}$$

which however does not give us good coordinates: $g(0) = g(2\pi) = 1$ for example. Note that it would be a bad idea to let $\psi \in (-\pi.. \pi]$. This would not be an open set, so we cannot use this as a coordinate patch. Here is why coordinate patches are defined that way: even though this would give a one-to-one map to $U(1)$, we still cannot do calculus on $U(1)$ by doing calculus on $(-\pi.. \pi]$. A smooth function on $(-\pi.. \pi]$ would not even give us a continuous function on $U(1)$ unless the limit of its value when going $-\pi$ equals its value at π .

Let us hence get rid of the multivaluedness of the coordinate ψ by restricting its range such

that $\psi \in (-\pi.. \pi)$. Now we have a one-to-one map from the open interval $(-\pi.. \pi)$ to all of $U(1)$ except the point $g = -1$. To make sure we can cover all of $U(1)$ by coordinates, we hence need a second patch. Let us set $g = e^{i\pi+i\theta}$ and again let $\theta \in (-\pi.. \pi)$. Now we can cover all of $U(1)$ except $g = 1$. In summary, we have

$$\begin{aligned} g &= e^{i\psi}, \quad \psi \in (-\pi.. \pi) \\ g &= e^{i\pi+i\theta}, \quad \theta \in (-\pi.. \pi) \end{aligned} \tag{1.86}$$

However, we can now describe all points except $1, -1$ in two ways using either ψ or θ . The coordinate changes are

$$\begin{aligned} \psi &= \pi + \theta, \quad \theta < 0 \\ \psi &= -\pi + \theta, \quad \theta > 0 \end{aligned} \tag{1.87}$$

Now we can use the open intervals described by ψ and θ to construct functions on $U(1)$, if the functions we consider agree on the overlap region $g \neq 1, -1$. If it bothers you that the overlap region is composed of two disconnected pieces, it is not difficult to introduce more patches such that the overlap between each pair is either empty or a single interval. Try to write this down clearly for 3 or 4 patches. Note in particular that the choice above in which every patch covers all but a single point is somewhat special.

Example 1.4. * Manifolds and the implicit function theorem.

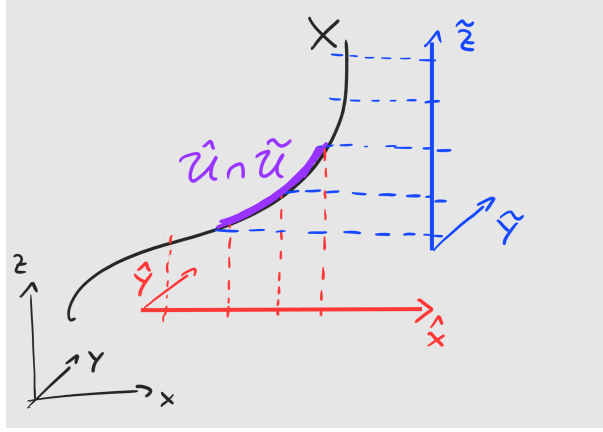


Figure 1.7: Using the implicit function theorem naturally recovers the abstract notion of a manifold for implicit surfaces $f(x, y, z) = 0$.

We can describe a subspace X of \mathbb{R}^3 given by the vanishing of a scalar function $f(x, y, z)$:

$$X : \{(x, y, z) \in \mathbb{R}^3 \mid f(x, y, z) = 0\}. \quad (1.88)$$

as a manifold as follows. By the implicit function theorem (see AMV II) we can find a function $g(x, y)$ such that $f(x, y, g(x, y)) = 0$ in a neighborhood $V \subset \mathbb{R}^3$ of a point x_0, y_0, z_0 where $\partial f / \partial z(x_0, y_0, z_0) \neq 0$. Let us call $\hat{U} = V \cap X$ and use \hat{x}, \hat{y} as coordinates in \mathbb{R}^2 . For a point $p = (x, y, z)$ in U we set

$$\hat{\phi} : (x, y, z) \rightarrow (\hat{x}, \hat{y}) \quad (1.89)$$

If $\partial f / \partial z(x_0, y_0, z_0) = 0$ but e.g. $\partial f / \partial x(x_0, y_0, z_0) \neq 0$ we can use the same theorem for $x = h(y, z)$ in a patch \tilde{U} :

$$\tilde{\phi} : (x, y, z) \rightarrow (\tilde{y}, \tilde{z}). \quad (1.90)$$

Recalling that $z = g(x, y)$ and $h = (y, z)$, the coordinate changes are given by

$$\hat{\phi} \circ \tilde{\phi}^{-1}(\tilde{y}, \tilde{z}) = \hat{\phi}(h(y, z), y, z) = (h(y, z), z) \quad (1.91)$$

and

$$\tilde{\phi} \circ \hat{\phi}^{-1}(\hat{x}, \hat{y}) = \tilde{\phi}(x, y, g(x, y)) = (y, g(x, y)). \quad (1.92)$$

A sketch of this situation is given in figure 1.7. Exercise: using the above strategy, find coordinate patches and coordinate changes on the two-sphere S^2 where $f = x^2 + y^2 + z^2$.

The only points (x_0, y_0, z_0) at which this strategy fails is when

$$\begin{aligned} f(x_0, y_0, z_0) &= 0 \\ \partial f / \partial x(x_0, y_0, z_0) &= \partial f / \partial y(x_0, y_0, z_0) = \partial f / \partial z(x_0, y_0, z_0) = 0. \end{aligned} \tag{1.93}$$

Hence X can not be given the structure of a differentiable manifold at such points. These points are called **singularities** of the surface $f(x, y, z) = 0$.

Another concept which we encountered in the examples we studied was that of a tangent vector. The way we constructed these was simple and can be immediately generalized. First we need to introduce paths.

Definition 1.20. A **path** is a continuous map S from an open interval $(a, b) \subset \mathbb{R}$ to X . Letting $t \in (a, b) \subset \mathbb{R}$ we can write this as

$$S : t \mapsto \mathbf{q}(t) \in X, \quad (1.94)$$

where $\mathbf{q}(t)$ is a description of our path using the coordinates on $\mathbb{R}^m \supset X$. We furthermore demand that $\mathbf{q}(t)$ is a differentiable function from S to \mathbb{R}^m .

Definition 1.21. A **tangent vector** at \mathbf{p} is the derivative of a path passing through \mathbf{p} with respect to its parameter t , evaluated at \mathbf{p} . Assuming that t_0 is such that $\mathbf{q}(t_0) = \mathbf{p}$ we can write

$$T_p(S) := \left. \frac{\partial \mathbf{q}(t)}{\partial t} \right|_{t=t_0}. \quad (1.95)$$

In the above definitions, we have used that the manifold X in question is realized as a sub-manifold of \mathbb{R}^m , which allows us to write tangent vectors conveniently as sitting inside \mathbb{R}^m . Given the definition of a manifold with coordinate charts we could also use the coordinates obtained this way to describe tangent vectors. While this is the superior perspective for more abstract and far-reaching applications to the notion of tangent vectors and, more generally, tensors, we'll stick to this more pedestrian approach.

Proposition 1.10. The tangent vectors $T_p(S)$ at a point \mathbf{p} form a real n -dimensional vector space $T_p X$, the tangent space at \mathbf{p} .

Proof. : problem class 2

Example 1.5. The tangent vector on a circle: For a path $(x_1, x_2) = (\cos t, \sin t)$ on the circle $x_1^2 + x_2^2 = 1$, the tangent vector at $t = t_0$ is

$$T_{t_0}(S^1) = \left. \frac{\partial}{\partial t} \begin{pmatrix} \cos t \\ \sin t \end{pmatrix} \right|_{t=t_0} = \begin{pmatrix} -\sin t_0 \\ \cos t_0 \end{pmatrix} \quad (1.96)$$

Example 1.6. The tangent vector of $SO(3)$ associated to the path

$$t \mapsto \begin{pmatrix} \cos t & \sin t & 0 \\ -\sin t & \cos t & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (1.97)$$

at $t = 0$. Note that we end up at the identity element for $t = 0$, so this gives a tangent vector at $\mathbf{1}$. We work out

$$\left. \frac{\partial}{\partial t} \begin{pmatrix} \cos t & \sin t & 0 \\ -\sin t & \cos t & 0 \\ 0 & 0 & 1 \end{pmatrix} \right|_{t=0} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (1.98)$$

Example 1.7. Let P be the surface given implicitly by $x^2 + y^2 + z = 0$ in \mathbb{R}^3 . Let's find the tangent space at the point at $(x, y, z) = (0, 0, 0)$. Here we need to work a little bit more and first need to construct a sufficient amount of paths passing through $(x, y, z) = (0, 0, 0)$. As we have one equation in \mathbb{R}^3 , P will turn out to be real two-dimensional, so we need at least two paths/two tangent vectors to generate the tangent space.

We can take $S_1 : (x, y, z) = (t, 0, -t^2)$ which gives $T_{(0,0,0)}(S_1) = (1, 0, 0)$ and $S_2 : (x, y, z) = (0, t, -t^2)$ which gives $T_{(0,0,0)}(S_2) = (0, 1, 0)$. Hence the tangent space $T_{(0,0,0)}(P)$ is simply the $z = 0$ plane. This is no surprise as P is an upside down paraboloid.

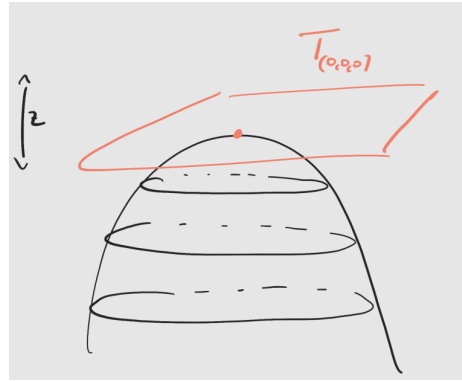


Figure 1.8: The paraboloid $x^2 + y^2 + z = 0$ with its tangent space at $(0, 0, 0)$ in red.

Example 1.8. Tangent vectors of $SU(2)$ at $p = \mathbb{1}$ have the form $i \sum_j \alpha_j \sigma_j$ with σ_j the Pauli matrices and $\alpha_j \in \mathbb{R}$. This can be seen by writing down a path

$$S : t \rightarrow e^{it\alpha_j \sigma_j} . \quad (1.99)$$

which passes through $\mathbb{1}$ at $t = 0$.

We now have

$$T_{\mathbb{1}}(S) = \left. \frac{\partial}{\partial t} e^{it\alpha_j \sigma_j} \right|_{t=0} = i\alpha_j \sigma_j . \quad (1.100)$$

Exercise 14. Find the tangent space of $SO(3)$ at the identity.

Exercise 15. Consider $y = f(x)$ sitting inside \mathbb{R}^2 with coordinates x, y and convince yourself that the notion of tangent introduced here is the same as the usual one.

Exercise 16. $O(1, 1)$ are the real 2×2 matrices O which leave the bilinear form $x_1^2 - x_2^2$ invariant when acting on $x = (x_1, x_2)$ as

$$x \rightarrow Ox .$$

a) Show that $O(1, 1)$ is a group using matrix multiplication.

b) Find the general form of elements of $O(1, 1)$.

- c) *Explain why $O(1, 1)$ is a differentiable manifold and write down coordinate charts.*
- d) *Find the tangent space of $O(1, 1)$ at the identity element.*

1.3 Lie groups

We are now ready to formally welcome Lie groups to these lectures. The idea here is simple: Lie groups unite the structures of groups and differentiable manifolds in a compatible way:

Definition 1.22. A **Lie group** is a group that is also a differentiable manifold such that the group operations

$$\begin{aligned} \circ : G \times G &\rightarrow G & (x, y) &\rightarrow x \circ y \\ ^{-1} : G &\rightarrow G & x &\rightarrow x^{-1} \end{aligned} \quad (1.101)$$

are differentiable maps.

Example 1.9. The group $\mathbb{C}^* \equiv \mathbb{C} \setminus \{0\}$ is a Lie group under multiplication. The map

$$(x, y) \rightarrow xy \quad (1.102)$$

is a differentiable map from $\mathbb{C}^* \times \mathbb{C}^*$ to \mathbb{C}^* , and $x \rightarrow 1/x$ is a differentiable map from \mathbb{C}^* to \mathbb{C}^* . This is an example of an abelian group.

Proposition 1.11. The group $GL(n, \mathbb{R})$ of real invertible $n \times n$ matrices is a Lie group under matrix multiplication. Note that these naturally sit inside \mathbb{R}^m with $m = n^2$.

Proof. : Recall the definition (from AMV II): a map is differentiable if it can locally be approximated by a linear map. Let us see if this is true for matrix multiplication. For two matrices $P, Q \in GL(n, \mathbb{R})$, the group operation is the map

$$(P, Q) \rightarrow PQ. \quad (1.103)$$

To examine if this can be approximated by a linear map we change P to $P + \Delta_P$ and Q to $Q + \Delta_Q$:

$$\begin{aligned} (P + \epsilon \Delta_P, Q + \epsilon \Delta_Q) &\rightarrow (P + \epsilon \Delta_P)(Q + \epsilon \Delta_Q) = PQ + P\epsilon \Delta_Q + \epsilon \Delta_P Q + \epsilon^2 \Delta_P \Delta_Q \\ &\simeq PQ + \epsilon(P\Delta_Q + \Delta_P Q) \end{aligned} \quad (1.104)$$

which is manifestly linear in both Δ_P and Δ_Q . Cramer's rule for constructing inverse matrices similarly shows that $P \rightarrow P^{-1}$ is differentiable \square

Theorem 1.3. A closed subgroup H of $GL(n, \mathbb{R})$ is again a Lie group.

REMARK: The word 'closed' does not refer to the group operation on elements being closed in H (this must be true for H to be a subgroup anyway) but is meant in the topological sense: H is a closed subset of $GL(n, \mathbb{R})$.

Proof. : An elementary proof of this result can be found in [Howe, 1983].

Definition 1.23. Lie groups that are closed subgroups of $GL(n, \mathbb{R})$ are called **matrix Lie groups**.

REMARK: Not all Lie groups are of this type, but in this course we will only study those. For the Lie groups that are most interesting to us, this only excludes a handful of cases.

We can hence get our hands on a lot of examples by finding closed subgroups of $GL(n, \mathbb{R})$. We start with

Definition 1.24. The **orthogonal group** $O(n)$ is the group of real $n \times n$ matrices g such that

$$g^T g = \mathbb{1}. \quad (1.105)$$

The **special orthogonal group** $SO(n)$ is the subgroup of matrices in $O(n)$ that have determinant $\det g = 1$.

REMARK: The group $O(n)$ consists of those invertible maps acting on a real vector space \mathbb{R}^n such that the canonical inner form stays invariant:

$$\mathbf{x} \cdot \mathbf{y} \rightarrow \mathbf{x}' \cdot \mathbf{y}' = \mathbf{x}^T g^T g \mathbf{y} = \mathbf{x} \cdot \mathbf{y}. \quad (1.106)$$

Corollary 1.1. $O(n)$ and $SO(n)$ are Lie groups.

Proof. : One can quickly check that these are indeed groups, they are obviously subgroups of $GL(n, \mathbb{R})$. The conditions that g has to satisfy in order to be in $O(n)$ only hold true on the closed subset where the defining relation $g^T g = \mathbb{1}$ holds true. To be more precise, for any matrix that does not satisfy these equations, we can find a little ball $GL(n, \mathbb{R})$ such that $g^T g \neq \mathbb{1}$ for every member of this ball. The complement of $O(n)$ in $GL(n, \mathbb{R})$ is hence open, which means that $O(n)$ is closed.

For $SO(n)$ we can repeat a similar argument.

Exercise 17. Make the argument above precise, i.e. show that for every $g \in GL(n, \mathbb{R}) \setminus O(n)$, i.e. $g \in GL(n, \mathbb{R})$ such that $g^T g \neq \mathbb{1}$, there is an open set U_g containing g such that U_g is entirely contained in $GL(n, \mathbb{R}) \setminus O(n)$.

hint: $GL(n, \mathbb{R})$ inherits its topology from the vector space $V_{n \times n}$ of real $n \times n$ matrices, which is isomorphic to \mathbb{R}^{n^2} : the n^2 entries of such a matrix are just the components of a vector in \mathbb{R}^{n^2}

from this perspective. We can hence describe the open ball of radius r around a matrix M with components M_{ij} as

$$B_r(M) \left\{ N \in V_{n \times n} \mid \sum_{ij} (N_{ij} - M_{ij})^2 < r \right\}. \quad (1.107)$$

REMARK: For $g \in O(n)$ it follows that $\det(g^T g) = (\det g)^2 = \det \mathbb{1} = 1$. As g is a real matrix, we hence have $\det g = \pm 1$. The space of such matrices is hence disjoint with two components, the one that contains the identity (which has $\det \mathbb{1} = 1$) is called $SO(n)$ and is a subgroup. The other component is not a subgroup.

REMARK: Conditions such as $g^T g = \mathbb{1}$ and $\det g = 1$ are typically called ‘closed conditions’ as the sets they define are closed sets in the vector space of all matrices.

Exercise 18. $GL(n, \mathbb{C})$ is the group of invertible complex $n \times n$ matrices. Show that $GL(n, \mathbb{C})$ is a Lie group.

Definition 1.25. The **unitary group** $U(n)$ is the group of complex $n \times n$ matrices g such that $U^\dagger U = \mathbb{1}$. The **special unitary group** $SU(n)$ is the subgroup of matrices in $U(n)$ that have determinant $\det g = 1$.

REMARK: The group $U(n)$ consists of those invertible maps acting on a complex vector space \mathbb{C}^n such that the canonical inner form stays invariant:

$$\bar{x} \cdot y \rightarrow \bar{x}' \cdot y' = \bar{x} \bar{g}^T g y = \bar{x} g^\dagger g y = \bar{x} \cdot y. \quad (1.108)$$

Corollary 1.2. The unitary and special unitary groups are Lie groups.

Proof. : These are both closed subgroups of $GL(n, \mathbb{R})$ by identifying \mathbb{C} with \mathbb{R}^2 . \square

1.4 Lie algebras

The idea of a Lie algebra is to formalize the notion of infinitesimal transformation. We first define Lie algebras abstractly.

Definition 1.26. A **Lie algebra** \mathfrak{g} is a vector space together with a bilinear map (‘**Lie bracket**’)

$$[\cdot, \cdot] : \mathfrak{g} \times \mathfrak{g} \rightarrow \mathfrak{g} \quad (1.109)$$

that is antisymmetric $[x, y] = -[y, x]$, and satisfies the Jacobi identity:

$$[x, [y, z]] + [y, [z, x]] + [z, [x, y]] = 0. \quad (1.110)$$

for all $x, y, z \in \mathfrak{g}$.

REMARK: This definition does not say if we should think of \mathfrak{g} as a real or complex vector space, so one and the same algebra can have different ‘real’ or ‘complex’ forms (or even forms over other fields).

Theorem 1.4. Every Lie group comes equipped with a Lie algebra which is equal to its tangent space at the identity element.

$$\mathfrak{g} = T_1 G. \quad (1.111)$$

Proof. : This is already a vector space by construction. For matrix Lie groups, we can simply take the bilinear form $[\cdot, \cdot]$ to be the commutator, which clearly satisfies the Jacobi identity. We will show in lemma 1.1 below that the commutator of two Lie algebra elements indeed returns a Lie algebra element. The general case requires some more technology we have not introduced, see [Harris et al., 1991] for details.

Corollary 1.3. The dimension of the Lie algebra (as a vector space) is equal to the dimension of its Lie group (as a differentiable manifold).

Exercise 19. Find the dimension of the group $SO(n)$ by finding the dimension of its Lie algebra.

Example 1.10. The Lie algebra $\mathfrak{u}(1)$ of $U(1)$ are the purely imaginary numbers and $[\gamma, \gamma'] = 0$ for all $\gamma, \gamma' \in \mathfrak{u}(1)$.

Example 1.11. The Lie algebra of \mathbb{C}^* are the complex numbers and $[\gamma, \gamma'] = 0$ for all γ, γ' in the Lie algebra $*$ of \mathbb{C}^* .

Example 1.12. The Lie algebra $\mathfrak{su}(n)$ of $SU(n)$ was found for the case $n = 2$ in Section 1.1.3. It is a real vector space with basis the complex $n \times n$ matrices γ such that $\gamma^\dagger = -\gamma$ and $\text{tr} \gamma = 0$. We can use i times the three Pauli matrices σ_j as basis vectors for $n = 2$. Note that while these have complex entries, this is a *real* vector space !

Example 1.13. As we have seen before, the group $SU(2)$ is a double cover of $SO(3)$. This means that a small neighborhood of the identity $\in SU(2)$ is isomorphic to a small neighborhood of the identity $\in SO(3)$, so that these two groups have isomorphic Lie algebras. You can also check this explicitly by working out the Lie algebra of $SO(3)$, see problem classes 2 and 3. We can hence have different groups that have the same Lie algebras.

Definition 1.27. For any Lie algebra, we can choose a **basis** $\{t_a\}_{a=1}^{\dim \mathfrak{g}}$ of so called **generators** t_a . In this basis the Lie bracket reads

$$[t_a, t_b] = f_{ab}^c t_c \quad (a, b, c = 1, \dots, \dim \mathfrak{g}) \quad (1.112)$$

where the f_{ab}^c are called **structure constants**, which express the component of the Lie bracket $[t_a, t_b]$ along the generator t_c . Repeated indices are summed over.

REMARK: While there are reasons for putting one index up in the expression f_{ab}^c , you can completely ignore this for now. Just think about f_{ab}^c as producing a number for any a, b, c and think about the positioning of the indices as a pure convention.

The Jacobi identity,(1.110), implies that

$$f_{ab}^d f_{dc}^e + f_{bc}^d f_{da}^e + f_{ca}^d f_{db}^e = 0 \quad (1.113)$$

for the structure constants.

Example 1.14. A basis of the Lie algebra $\mathfrak{su}(2)$ of $SU(2)$ is given by $t_a = i\sigma_a$ for σ_a the Pauli matrices. We can work out

$$[i\sigma_a, i\sigma_b] = -[\sigma_a, \sigma_b] = -2\epsilon_{abc}i\sigma_c \quad (1.114)$$

so that we conclude that $f_{ab}^c = -2\epsilon_{abc}$ for $\mathfrak{su}(2)$.

Definition 1.28. The **exponential map** is a map $\exp : \mathfrak{g} \rightarrow G$ sending $\gamma \in \mathfrak{g}$ to

$$e^\gamma := \sum_k \frac{\gamma^k}{k!} \in G. \quad (1.115)$$

REMARK: For every matrix $\gamma \in \mathfrak{g}$, one can show that the above indeed converges and that it is indeed in G if γ is in \mathfrak{g} , see [Hall, 2003] or the non-examinable box below for a sketch of a proof.

More on the exponential map*

Let us explain how the exponential map comes about by taking a slightly more geometric perspective. Elements of the Lie algebra are associated with elements of the tangent space of G at the identity and we can think of both Lie algebra elements and Lie group elements as matrices, which can be multiplied. For any $\gamma \in T_1(G)$, it turns out that

$$L(\gamma)|_g = g\gamma. \quad (1.116)$$

is a tangent vector at g for any point $g \in G$. This defines what is called a vector field $L(\gamma)$, i.e. something that attaches a tangent vector to any point on G . The vector fields we have just defined are called **left-invariant** vector fields and have the nice property that

$$g'L(\gamma)|_g = g'g\gamma = L(\gamma)|_{g'g}. \quad (1.117)$$

Now what's important about vector fields is that one can flow along with them. E.g. flowing out from the identity is done by solving the differential equation

$$\frac{\partial g(t)}{\partial t} = L(\gamma)|_{g(t)} = g(t)\gamma. \quad (1.118)$$

The solution to this flow is a path

$$g(t) = e^{t\gamma}, \quad (1.119)$$

and the fact that we constructed this as a flow shows that the exponential actually lands in the group.

You can also understand convergence of the series we used to define the exponential by observing that any power of γ will produce a matrix with entries that are polynomials in the components of γ . For $k \rightarrow \infty$ the $k!$ then grows faster at some point which makes the sum converge to a finite value.

Exercise 20. Consider the set G of matrices

$$G = \left\{ \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} \mid a, b, c \in \mathbb{R}, ac \neq 0 \right\} \quad (1.120)$$

a) Show that G is a Lie group using matrix multiplication as the group composition.

b) Find the Lie algebra \mathfrak{g} of G .

c) Compute the exponentials of the basis elements of the Lie algebra you have found.

Lemma 1.1. Let G be a Lie group and \mathfrak{g} be its Lie algebra. We then have

i) $g\gamma g^{-1} \in \mathfrak{g}$ for all $\gamma \in \mathfrak{g}$ and $g \in G$.

ii) $[\gamma, \delta] \in \mathfrak{g}$ for all $\gamma, \delta \in \mathfrak{g}$

Proof. : To see the first part, let's try to construct a path that gives us $g\gamma g^{-1}$ as a tangent vector upon differentiating. We could try

$$e^{tg\gamma g^{-1}} = \sum_k \frac{(tg\gamma g^{-1})^k}{k!} = g \left(\sum_k \frac{(t\gamma)^k}{k!} \right) g^{-1} = ge^{t\gamma} g^{-1}. \quad (1.121)$$

As all of the factors on the rhs are in G , it follows that $ge^{t\gamma} g^{-1} \in G$. As $ge^{t\gamma} g^{-1}$ is a path in G that passes through $\mathbb{1}$ at $t = 0$ and

$$\left. \frac{\partial}{\partial t} ge^{t\gamma} g^{-1} \right|_{t=0} = g\gamma g^{-1} \quad (1.122)$$

it follows that $g\gamma g^{-1} \in \mathfrak{g}$. Although we are talking about matrix Lie algebras in this course where we can just multiply elements $g \in G$ with elements $\gamma \in \mathfrak{g}$, you might feel a little uneasy about just multiplying them. In this case, you can read the above statement as a *definition* of what $g\gamma g^{-1}$ is: it is the Lie algebra element you get from the path $ge^{t\gamma}g^{-1}$.

For the second part, consider $e^{t\gamma}\delta e^{-t\gamma}$ for $\delta \in \mathfrak{g}$. It follows from i) that this is in \mathfrak{g} for all t . As a tangent space, the Lie algebra is in particular an n -dimensional vector space which sits inside the vector space of $n \times n$ matrices. As such it is closed under taking limits. Hence

$$\lim_{t \rightarrow 0} (e^{t\gamma}\delta e^{-t\gamma} - \delta)/t = \gamma\delta - \delta\gamma = [\gamma, \delta]. \quad (1.123)$$

is in \mathfrak{g} . \square

A natural question about the exponential map concerns its injectivity and surjectivity. Clearly, it cannot be injective for every group G . We have already seen it is not for $U(1)$ and $SU(2)$, which in turn had to be like that, because this is how these groups become topologically non-trivial.

It can also not always be surjective as the following simple counter-example shows.

Example 1.15. Elements of the Lie algebra $\mathfrak{sl}(2, \mathbb{R})$ of $SL(2, \mathbb{R})$ must obey

$$e^\gamma = g \quad (1.124)$$

for $g \in SL(2, \mathbb{R})$. This implies that γ is real by taking complex conjugation. Furthermore

$$\det e^\gamma = e^{\text{tr}\gamma} = 1 \quad (1.125)$$

implies that γ is traceless. Finally, e^γ always maps to $SL(2, \mathbb{R})$ if the above conditions are met: the inverse is simply $e^{-\gamma}$. The Lie algebra $\mathfrak{sl}(2, \mathbb{R})$ hence contains traceless real matrices. Now consider the matrix

$$g = \begin{pmatrix} -4 & 0 \\ 0 & -1/4 \end{pmatrix} \in SL(2, \mathbb{R}) \quad (1.126)$$

We claim there is no element $\gamma \in \mathfrak{sl}(2, \mathbb{R})$ s.t. $e^\gamma = g$.

Proof. : If such an element exists, we can immediately write down a square root of g as $\sqrt{g} = e^{\frac{1}{2}\gamma}$. But as we show now, no such square root (in $SL(2, \mathbb{R})$) exists. The eigenvalues of g are 4 and $1/4$, so there is one eigenvalue of \sqrt{g} that is $\pm 2i$ and another one that is $\pm \frac{1}{2}i$. However, for \sqrt{g} to be in $SL(2, \mathbb{R})$ it must be a real matrix, so that the eigenvalues are given by an equation of the form $\lambda^2 + p\lambda + q = 0$ with p, q real. Hence

$$\lambda_{\pm} = -p/2 \pm \sqrt{(p/2)^2 - q} \quad (1.127)$$

so that there are two eigenvalues which are either real or complex conjugates of each other. However $\pm 2i$ is not real and never the complex conjugate of $\pm \frac{1}{2}i$. Hence there is no \sqrt{g} such that $(\sqrt{g})^2 = g$. But this implies that there cannot be a γ with $g = e^\gamma$ as could write down such a \sqrt{g} otherwise. \square

Although it does not hold in general, there are favourable circumstances where the exponential map is surjective. We have already seen this for $SU(2)$ and $SO(3)$ already (see also exercises). To spell out the general result clearly, we need one more

Definition 1.29. A subset of \mathbb{R}^n is called **compact** if it is closed and bounded, i.e. one can find a ball of finite size that entirely contains it.

Example 1.16. $U(1)$ and $SU(2)$ are both compact.

Fact 1.1. The orthogonal and unitary groups $O(n)$, $SO(n)$, $U(n)$, $SU(n)$ are all compact.

Theorem 1.5. If G is a connected, compact matrix Lie group, the exponential map for G is surjective.

Proof. : We only give a sketch, details are found in [Hall, 2003]. The main idea is to observe that the exponential map is surjective for $U(1)$ and then try to replicate this setting. The crucial step is to show for any compact matrix Lie group G that every element $g \in G$ lies inside some $U(1)$ subgroup of G . Once this ‘torus theorem’ is established, we can simply use the generators of the $U(1)$ to reach g by the exponential map. \square

REMARK: It is not true that the exponential map is surjective for compact groups only, e.g. \mathbb{C}^* is not compact but we can write every element in \mathbb{C}^* as e^z for some complex number z .

Classification of compact Lie algebras*

Definition 1.30. Lie algebras of compact Lie groups are called **compact Lie algebras**.

Definition 1.31. An **ideal** of a Lie algebra is a subset $I \subset \mathfrak{g}$ such that $[\iota, x] \subset I$ for all $\iota \in I$ and all $x \in \mathfrak{g}$.

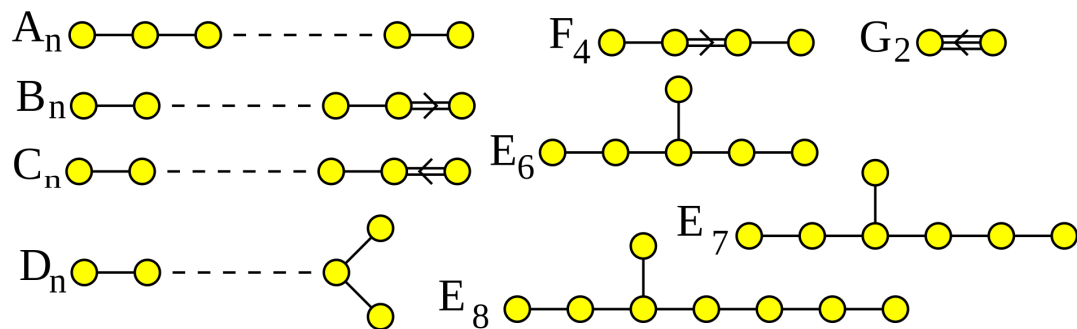
Definition 1.32. A **simple** Lie algebra is a Lie algebra that has no non-trivial ideals.

Theorem 1.6. Any compact Lie algebra can be decomposed into the direct sum of $u(1)$ Lie algebras and of simple Lie algebras:

$$\mathfrak{g} = u(1) \oplus \cdots \oplus u(1) \oplus \mathfrak{g}_1 \oplus \cdots \oplus \mathfrak{g}_l. \quad (1.128)$$

Proof. : [Hall, 2003].

Simple Lie algebras were in turn classified by Killing and Cartan, and this classification was put in its definitive form by Dynkin. The matrix Lie algebras form four infinite series, the so called classical Lie algebras $A_n = su(n+1)$, $B_n = so(2n+1)$, $C_n = usp(2n)$, $D_n = so(2n)$. But there are a few more exceptional Lie algebras which are not of matrix type: E_6 , E_7 , E_8 , F_4 , G_2 . See [Cahn, 2014] for a down-to-earth introduction to the subject and [Harris et al., 1991, Hall, 2003] for a more advanced perspective. The structure of these algebras boils down to the following pictures, called Dynkin diagrams, which determine the structure constants in a suitable basis of \mathfrak{g} .



Topic 2

Representations

Although we have introduced groups in a rather concrete form as subspaces of $GL(n, \mathbb{R})$ or $GL(n, \mathbb{C})$, we can also drop this association and just keep their abstract structure. Taking this point of view, we will explore how groups can act on vector spaces, and how their structure prefers some vector spaces over others, in this section. Such questions belong to a subject called representation theory, and this is a vast field that we will only scratch the surface of. There is dedicated lecture, MATH4241: REPRESENTATION THEORY IV that gives a more detailed and general account of this subject. Here, we will take a practical approach and mostly only explore those aspects of direct use to us.

2.1 Generalities and Basic Examples

Definition 2.1. For a vector space V , we will denote the group of invertible linear maps acting on V by $GL(V)$. Hence for $V = \mathbb{R}^n$, $GL(V) = GL(n, \mathbb{R})$ and for $V = \mathbb{C}^n$, $GL(V) = GL(n, \mathbb{C})$.

Definition 2.2. A **representation** of a group G is a group homomorphism $r : G \rightarrow GL(V)$, where V is a finite-dimensional (real or complex) vector space.

REMARK: What this means is that we ‘represent’ the group G by matrices in $GL(V)$. Given r , we can hence act with the group G on vectors in V using linear maps. This is often expressed as

- ‘ G acts on (elements of) V in the representation r ’
- ‘elements of V transform under G in the representation r ’

- ‘elements of V live in the representation r ’

Although a representation is defined as the map r which takes elements of G to elements of $GL(V)$, it is quite common to speak about the elements in V that the image of G in $GL(V)$ acts on as a ‘representation’ in the physics literature.

When we define a representation we have essentially two options, and we will see examples of both:

- We explicitly write some matrices $M(g)$ in $GL(V)$ for any $g \in G$.
- We describe how $M(g)$ acts on a vector in V without explicitly giving $M(g)$.

REMARK: We only ask a representation to be a homomorphism, i.e. r need not be injective. This means some aspects of G can get lost in a representation.

Given that we defined $SU(n)$ as a group of matrices acting on \mathbb{C}^n , we can just use this as an example of a representation.

Definition 2.3. The **defining representation** (also called **fundamental representation**) of $SU(n)$ is the representation where the matrices defining this groups are allowed to be themselves: $r(g) = g$. The fundamental representation of $SU(n)$ is complex n -dimensional and is denoted by ‘the \mathbf{n} of $SU(n)$ ’.

REMARK: We may construct another representation of $SU(n)$ that is also n -dimensional and is called ‘the $\bar{\mathbf{n}}$ of $SU(n)$ ’ by acting with \bar{g} instead of g . This representation is isomorphic to the \mathbf{n} representation.

Definition 2.4. The **defining representation** of $SO(n)$ is the representation where the matrices defining this groups are allowed to be themselves: $r(g) = g$. The defining representation of $SO(n)$ is real n -dimensional and is denoted by ‘the \mathbf{n} of $SO(n)$ ’.

Even though matrix groups sit inside vector spaces, they themselves are not vector spaces, so a group action on itself is not a representation (in general). Here is an example of a non-trivial representation that exists for all Lie groups. It exploits the fact that Lie groups come equipped with an intrinsic vector space: each one of them has its own Lie algebra, which is a vector space.

Definition 2.5. The **adjoint** representation is a map $Ad : G \rightarrow GL(\mathfrak{g})$ which sends a group element g to the $GL(\mathfrak{g})$ element $Ad(g)$, which is defined by its action on \mathfrak{g} :

$$Ad(g) : \gamma \mapsto g\gamma g^{-1}. \quad (2.1)$$

REMARK: Don't get confused, there are two maps we need to distinguish. There is the map $Ad(g)$ sending g to the $GL(\mathfrak{g})$ element $Ad(g)$, but this itself is a map acting on the vector space \mathfrak{g} . The definition above works by telling us how $Ad(g)$ acts on any $\gamma \in \mathfrak{g}$ given g , i.e. $\gamma \rightarrow g\gamma g^{-1}$ is the linear map in $GL(\mathfrak{g})$ that is the image of $g \in G$ under the representation Ad . That this is well-defined follows directly from theorem 1.1 part i) where we showed that $g\gamma g^{-1} \in \mathfrak{g}$. As it is also a linear non-degenerate map, it is hence in $GL(\mathfrak{g})$. Taken as a vector space, the Lie algebra \mathfrak{g} of a Lie group G is hence acted on by a representation of G .

Example 2.1. As $e^{i\phi}i\theta e^{-i\phi} = i\theta$ for all $g \in U(1)$, the adjoint representation of $U(1)$ is trivial.

Example 2.2. The adjoint representation of $SU(2)$ is precisely the map we used to map it to $SO(3)$. Recall that it acts on its own Lie algebra here, which is a real three-dimensional vector space, so this makes perfect sense.

Example 2.3. The adjoint representation of $SU(3)$ acts on a real eight-dimensional vector space: the matrices in $\mathfrak{su}(3)$ are traceless anti-hermitian 3×3 matrices. Their number of real components is 2 from the diagonal plus 3×2 from off-diagonal terms. This is the reason there are eight different gluons in strong interactions.

Exercise 21. Writing \mathbb{R}^3 as

$$M_{\mathbf{v}} = \begin{pmatrix} v_3 & v_1 - iv_2 \\ v_1 + iv_2 & -v_3 \end{pmatrix}. \quad (2.2)$$

we considered the action of $g \in SU(2)$ on \mathbb{R}^3 defined by

$$F(g) : M_{\mathbf{v}} \mapsto gM_{\mathbf{v}}g^\dagger \quad (2.3)$$

in the lectures. Show that this is a representation, and that this representation is the adjoint representation of $SU(2)$.

Exercise 22. The adjoint action representation defines a linear map $r(g)$ acting on \mathfrak{g} and as such can be written as a matrix M acting on a column vector after choosing a basis for \mathfrak{g} . Make this explicit for the action of

$$g = \begin{pmatrix} e^{i\phi} & 0 \\ 0 & e^{-i\phi} \end{pmatrix} \in SU(2). \quad (2.4)$$

in the adjoint representation. Is the adjoint representation faithful?

Exercise 23. Let $\mathbf{q} \in \mathbb{C}^n$ be acted on in the fundamental representation of $SU(n)$ and γ in the adjoint representation of $SU(n)$ (this is often expressed as \mathbf{q} 'lives' in the fundamental and γ 'lives' in the adjoint of $SU(n)$.)

By acting with $SU(n)$ simultaneously on γ and \mathbf{q} , describe the action of $SU(n)$ on

i) $\mathbf{v} = \gamma\mathbf{q}$

ii) $\bar{\mathbf{q}}$

iii) A matrix Q with components $Q_{ij} = q_i q_j$

and decide in each case if this defines a representation.

Exercise 24. Let $g \in SO(3)$ be given by

$$g = \begin{pmatrix} \cos \phi & \sin \phi & 0 \\ -\sin(\phi) & \cos(\phi) & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Find the action of g in the adjoint representation and describe it using a basis of the vector space $\mathfrak{so}(3)$. As $\mathfrak{so}(3)$ is the same as \mathbb{R}^3 , we can describe its elements as column vectors after having chosen a basis. Using the basis you have chosen, write the adjoint action as a 3×3 matrix acting on a column vector.

Exercise 25. Let P be a homogeneous polynomial in two complex variables z_1 and z_2 of degree d , i.e. we can write

$$P(\mathbf{z}) = \sum_{k=0}^d \alpha_k z_1^k z_2^{d-k} \quad (2.5)$$

for complex numbers α_k .

There is a natural action of $SU(2)$ on $\mathbf{z} = (z_1, z_2)$, which is just

$$\mathbf{z} \mapsto g\mathbf{z}. \quad (2.6)$$

For a polynomial $P(\mathbf{z})$, we can then define an action by $SU(2)$ as

$$r_d(g) : P(\mathbf{z}) \mapsto P(g^{-1}\mathbf{z}). \quad (2.7)$$

Show that this defines a representation of $SU(2)$.

[remark: in the above formula, g^{-1} does not act on the argument of P but on \mathbf{z} , i.e. the action on $P(A\mathbf{z})$ for a 2×2 matrix A would be $r_d(g) : P(A\mathbf{z}) \mapsto P(Ag^{-1}\mathbf{z})$.]

Definition 2.6. A representation is called faithful if r is injective.

Example 2.4. We have seen examples of $SU(2)$ acting (faithfully) on \mathbb{C}^2 in Section 1.1.3 and (non-faithfully) on \mathbb{R}^3 in example 1.1.4.

Example 2.5. We can act with $SU(2)$ (faithfully) on \mathbb{C}^4 by using the block-diagonal representation

$$r : g \rightarrow \begin{pmatrix} g & 0 \\ 0 & g \end{pmatrix} \quad (2.8)$$

Clearly this seems a bit redundant and we want to distinguish between such cases and those that truly give us something new. One way to phrase this is in terms of invariant subspaces.

Definition 2.7. A subspace $W \subseteq V$ is called **invariant** if $r(g)w \in W$ for all $g \in G$ and all $w \in W$.

Example 2.6. Coming back to the example 2.5, we can decompose $V = \mathbb{C}^2 \oplus \mathbb{C}^2 \oplus \mathbb{C}^2 \oplus \dots$ and every one of those summands is an invariant subspace.

Definition 2.8. A representation $r : G \rightarrow GL(V)$ is **irreducible** if the only invariant subspaces are V and $\{0\}$. Otherwise it is called **reducible**.

Exercise 26. Let G be a Lie group and H be a subgroup of G that is also a Lie group.

- a) Explain why any representation $r(G)$ of G also gives us a representation $r(H)$ of H .
- b) Let's assume $r(G)$ is irreducible. Can you think of an example where the representation $r(H)$ is reducible? Can you think of an example where the representation $r(H)$ is irreducible?

Example 2.7. Mapping all $g \in G$ to $1 \in GL(V)$ is a group homomorphism, so this trivial representation always exists. This is as un-faithful as possible and reducible: every subspace of V is an invariant subspace. Objects transforming in this representation are called **scalars** or **singlets**. They are often referred to as 'living in the 1 of G '.

Definition 2.9. A representation $r : G \rightarrow GL(V)$ is **unitary** if V has an inner form¹ $\langle \cdot, \cdot \rangle$ and $\langle x, y \rangle = \langle r(g)x, r(g)y \rangle$ for all $g \in G$ and all $x, y \in V$.

Example 2.8. The defining representation of $SU(n)$ is faithful, irreducible, and unitary.

Given that we have introduced most matrix groups as preserving some inner form, this seems like a natural concept. It's power lies in the following

Theorem 2.1. Let $r : G \rightarrow GL(V)$ be a finite-dimensional unitary representation. Then it can be completely decomposed into irreducible representations $r_i(G)$:

$$r(G) = \bigoplus_i r_i(G) , \quad V = \bigoplus_i V_i , \quad r_i(G) \in GL(V_i) . \quad (2.9)$$

If you like, you can think of $r(G)$ as respecting the same block-diagonal form for all $g \in G$ in an appropriate basis of V . We have already seen a reducible representation that can be decomposed into irreducible ones, see example 2.5.

Proof. :

¹A **inner form** is a symmetric bilinear map $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R}$ s.t. $\langle v, v \rangle \geq 0$ for all $v \in V$ and $\langle v, v \rangle = 0$ if and only if $v = 0$.

Let $r(G)$ be a reducible representation (otherwise there is nothing to prove) and consider any of its invariant subspaces W . The main step of the proof is to show that the orthogonal complement

$$W^\perp := \{v \in V \mid \langle v, w \rangle = 0 \ \forall w \in W\} \quad (2.10)$$

is an invariant subspace as well.

For any $r(g)$ we can define its dual $r^*(g)$ by

$$\langle v, r(g)u \rangle = \langle r^*(g)v, u \rangle. \quad (2.11)$$

for all $v, u \in V$. It follows that

$$\langle v, u \rangle = \langle r(g)v, r(g)u \rangle = \langle r^*(g)r(g)v, u \rangle \quad (2.12)$$

so that $r^*(g)r(g) = \mathbb{1}$.

Now for all $w \in W$, $v \in W^\perp$ and all $g \in G$ we have

$$0 = \langle v, w \rangle = \langle v, r(g)w \rangle = \langle r^*(g)v, w \rangle = \langle (r(g))^{-1}v, w \rangle \quad (2.13)$$

where $(r(g))^{-1}$ is the inverse of the matrix $r(g) \in GL(V)$. As every element in G has an inverse and $(r(g))^{-1} = r(g^{-1})$ (as shown in problems), we can just write

$$0 = \langle r(g)v, w \rangle \quad (2.14)$$

for all $w \in W$, $v \in W^\perp$ and all $g \in G$. This means whatever g acts on $v \in W^\perp$, we stay in W^\perp . Hence W^\perp is an invariant subspace as well, which is what we wanted to show.

Now we can decompose

$$r(G) = r_W(G) \oplus r_{W^\perp}(G), \quad V = W \oplus W^\perp. \quad (2.15)$$

as both W and W^\perp are invariant subspaces. If both $r_W(G)$ and $r_{W^\perp}(G)$ are irreducible we are done. Otherwise, we can simply run the same argument again to achieve a finer decomposition. This iteration must terminate as V is finite dimensional. \square

For unitary representations all is hence nice and easy. But what can we do when we do not have an inner form that is respected by $r(g)$? Using ‘Weyl’s unitarity trick’ we can just cook one up (if G is compact)!

Theorem 2.2. *Let G be a compact Lie group and $r(G)$ a finite-dimensional representation on a vector space with inner form $\langle \cdot, \cdot \rangle$. Then there exists an inner form that is invariant under $r(G)$ and hence the same statement as in Theorem 2.1 holds.*

Proof. : Let $\langle \cdot, \cdot \rangle$ be some inner form on V . As G is a compact group, $\langle r(g)v, r(g)w \rangle$ is bounded for fixed v, w : this expression cannot diverge for $g \rightarrow \hat{g}$ anywhere on G as such a \hat{g} cannot be in G . But it follows from G being topologically closed that any sequence of group elements $g_i \in G$ has a limit that is also in G . Hence there must be a maximal value of $\langle r(g)v, r(g)w \rangle$ for fixed v and w and we can use that as the bound.

Furthermore, G is some bounded subspace in \mathbb{R}^m for some m for the matrix Lie groups we are treating in these lectures, and as such has a finite volume. We can then integrate a bounded function over it and receive a finite answer. In particular, we can use any realization of G as a subset of \mathbb{R}^n to define

$$\langle v, w \rangle_G := \int_G \langle r(g)v, r(g)w \rangle dV. \quad (2.16)$$

What is happening here is that we are averaging over the action of the group on $\langle v, w \rangle$. Let's act with a group element h on v and w

$$\langle r(h)v, r(h)w \rangle_G = \int_G \langle r(g)r(h)v, r(g)r(h)w \rangle dV = \int_G \langle r(gh)v, r(gh)w \rangle dV. \quad (2.17)$$

where we have used that r is a group homomorphism. Now if g sweeps out the whole group, so does gh for any $h \in G$. In particular, every group element g' can be uniquely written as $g' = gh$ for some g , just take $g = g'h^{-1}$. Hence

$$\langle r(h)v, r(h)w \rangle_G = \int_G \langle r(gh)v, r(gh)w \rangle dV = \int_G \langle v, w \rangle dV = \langle v, w \rangle_G. \quad (2.18)$$

and we are done. \square

An important feature of complex irreducible representations is **Schur's lemma**:

Theorem 2.3. *Let r be an irreducible representations of G on a finite-dimensional complex vector space V , and let $T : V \rightarrow V$ be a linear map such that*

$$r(g)T = T r(g) \quad (2.19)$$

for all $g \in G$. Then

a) $T = 0$

or

b) $T = c\mathbb{1}$ for some complex number c

Proof. : First observe that $\ker T$ is an invariant subspace: if $v \in \ker T$ we have

$$0 = Tv = r(g)Tv = Tr(g)v \quad (2.20)$$

so $r(g)v \in \ker T$ as well. As we have assumed that r is irreducible, $\ker T = V$ or $\ker T = \{0\}$. If $\ker T = V$ it follows that $T = 0$, so case a) is realized and we are done.

Let us hence assume from now on that $\ker T = \{0\}$. As a complex matrix, T has at least one non-zero eigenvalue, let that eigenvalue be c and the associated eigenvector be v_c . Now consider the map $\hat{T} := T - c\mathbb{1}$ for which $v_c \in \ker \hat{T}$. We have

$$r(g)\hat{T} = \hat{T}r(g) \quad (2.21)$$

as the identity commutes with every matrix. Now we can again observe as above that $\ker \hat{T}$ is an invariant subspace and hence must be $\{0\}$ or V . We already know that $\ker \hat{T} \neq 0$, so it must be that $\ker \hat{T} = V$ which implies $\hat{T} = 0$, i.e. $T = c\mathbb{1}$. \square

2.2 Representations of $U(1)$

The complex representations of $U(1)$ can be found by more or less elementary considerations.

Theorem 2.4. *Complex irreducible representations of $U(1)$ are all unitary, take $U(1)$ to $GL(1, \mathbb{C})$, i.e. act on \mathbb{C} , and only depend on an integer n . For $g = e^{i\phi}$ we can write the homomorphism $f_n : U(1) \rightarrow GL(1, \mathbb{C})$ as*

$$f_n(g) = g^n = e^{in\phi}. \quad (2.22)$$

Proof. : We first show that all complex irreducible representations of $U(1)$ are one-dimensional. Consider such a representation $f(g)$ of $U(1)$ for which $f(g) \in GL(m, \mathbb{C})$. As $U(1)$ is an abelian group we have

$$f(g)f(h) = f(gh) = f(hg) = f(h)f(g). \quad (2.23)$$

for all $g, h \in U(1)$. Now let us fix h for the time being. Then we can set $r(g) = f(g)$ in Schur's lemma, and use $T = f(h)$ to conclude that $f(h)$ must be proportional to the identity map. This is true for all $h \in U(1)$ (different h might produce different c however) which implies that the representation $f(g)$ is one-dimensional: any subspace of \mathbb{C}^m is an invariant subspace, and the only subspaces giving irreducible ones are complex one-dimensional ones.

As f is a homomorphism we need

$$f(hg) = f(h)f(g). \quad (2.24)$$

Differentiating both sides w.r.t h and setting $h = 1$ we find

$$gf'(g) = f'(1)f(g). \quad (2.25)$$

Let us write the constant $f'(1)$ by $n = f'(1)$. The the unique solution to the above differential equation that satisfies $f(1) = 1$ (a consequence of f being a group homomorphism) is

$$f(g) = g^n = e^{i\phi n}. \quad (2.26)$$

Letting $\phi = 2\pi$ yields $g = 1$, so that $f(1) = 1$ additionally requires $e^{2\pi i n} = 1$, i.e. $n \in \mathbb{Z}$. For $f(g)$ we have $f(g)^\dagger = f(g)^{-1}$, so these are all unitary (using the standard inner form on \mathbb{C}). \square

REMARK: The integer n is often called the ‘charge’ in physics, and you will see later that it deserves this name when we study electromagnetism, but this also shows up in other contexts with a $U(1)$ action. The realization that n is an integer has the profound consequence that charges are quantized, i.e. they are multiples of some fundamental charge (corresponding to $n = 1$). This doesn’t explain why protons and electrons have the same charges, but already implies that their charges must satisfy

$$\frac{q_{\text{electron}}}{q_{\text{proton}}} = \frac{n_{\text{electron}}}{n_{\text{proton}}} \quad (2.27)$$

i.e. the ratio must be a rational number.

Exercise 27.

a) Describe a $U(1)$ subgroup of $SU(2)$. Is $U(1) \times U(1)$ a subgroup of $SU(2)$ as well?

b) Let A be an element of the vector space that is acted on by the adjoint representation of $SU(2)$. For the $U(1)$ subgroup of $SU(2)$ you identified above, find the action on A and use this to decompose the action of $U(1)$ into irreducible representations.

Exercise 28. Consider the map $r_\kappa : U(1) \rightarrow GL(3, \mathbb{C})$ defined by

$$r_\kappa(e^{i\phi}) = e^{\phi\lambda\kappa}$$

where $\kappa \in \mathbb{C}$ and

$$\lambda = \begin{pmatrix} 0 & i & 0 \\ i & 0 & i \\ 0 & i & 0 \end{pmatrix}$$

For which values of κ is r_κ a representation of $U(1)$? [hint: think about what happens to eigenvectors of λ and use the classification theorem for complex representations of $U(1)$.]

2.3 Representations of Lie algebras

If we have found a representation of a Lie group G on some vector space V , every element $g \in G$ is assigned an element $r(g) \in GL(V)$. We can think the same way about representations of Lie algebras. The only difference is that we want to preserve the algebra structure.

Definition 2.10. A **Lie algebra homomorphism** is a linear map $f : \mathfrak{g} \rightarrow \mathfrak{h}$ between Lie algebras \mathfrak{g} and \mathfrak{h} such that $[f(\gamma), f(\delta)] = f([\gamma, \delta])$.

Definition 2.11. A **representation** of a Lie algebra \mathfrak{g} is a Lie algebra homomorphism $\rho : \mathfrak{g} \rightarrow \mathfrak{gl}(V)$ for a finite-dimensional vector space V .

Definition 2.12. A representation of a Lie algebra \mathfrak{g} is called **reducible** if there exists an invariant subspace, i.e. there exists a $W \subset V$ with $W \neq \{0\}$ and $W \neq V$ s.t.

$$\rho(\gamma)w \in W \quad \forall w \in W \quad \forall \gamma \in \mathfrak{g}.$$

In the same way that a path $g(t)$ passing through $\mathbb{1}$ determines an element γ of the Lie algebra, we can use $r(g(t))$ to determine an associated representation $\rho(\gamma)$: all we need to do is consider $r(g(t))$ instead of $g(t)$ and do the same computation.

Proposition 2.1. Given a finite-dimensional representation r of a Lie group G , there is a unique associated representation ρ of its Lie algebra \mathfrak{g} such that

$$r(e^{t\gamma}) = e^{t\rho(\gamma)}, \quad (2.28)$$

we can compute this by working out

$$\rho(\gamma) = \left. \frac{\partial}{\partial t} r(e^{t\gamma}) \right|_{t=0}. \quad (2.29)$$

Proof. : First of all (2.29) shows how to compute the map $\rho(\gamma)$ from $r(g)$, i.e. it is uniquely given once $r(g)$ is fixed.

Nect we check that we satisfy the definition. Consider the path

$$g(t) := e^{t\gamma} e^{t\delta}. \quad (2.30)$$

On the one hand we have

$$\left. \frac{\partial}{\partial t} r(g(t)) \right|_{t=0} = \left. \frac{\partial}{\partial t} r(e^{t\gamma} e^{t\delta}) \right|_{t=0} = \left. \frac{\partial}{\partial t} r(e^{t\gamma}) r(e^{t\delta}) \right|_{t=0} = \rho(\gamma) + \rho(\delta). \quad (2.31)$$

Now consider

$$\begin{aligned} e^{t\gamma} e^{t\delta} &= \sum_k \frac{(t\gamma)^k}{k!} \sum_l \frac{(t\delta)^l}{l!} = (1 + t\gamma + t^2\gamma^2/2 + \dots)(1 + t\delta + t^2\delta^2/2 + \dots) \\ &= \left(\sum_k \frac{t^k(\gamma + \delta)^k}{k!} \right) + \frac{t^2}{2}(\gamma\delta - \delta\gamma) + t^3(\dots) + \dots. \end{aligned} \quad (2.32)$$

Such a relation can be given more concisely as what is called the ‘Baker-Campbell-Hausdorff formula’, see [Hall, 2003] for details.

Hence

$$\left. \frac{\partial}{\partial t} r(e^{t\gamma} e^{t\delta}) \right|_{t=0} = \left. \frac{\partial}{\partial t} r(e^{t(\gamma+\delta)} + t^2(\dots)) \right|_{t=0} = \left. \frac{\partial}{\partial t} r(e^{t(\gamma+\delta)}) \right|_{t=0} = \rho(\gamma + \delta). \quad (2.33)$$

In the first step, we have used the relation established above, the second step is just the chain rule, and the third step uses the definition of a Lie algebra representation.

Hence we have shown that

$$\rho(\gamma) + \rho(\delta) = \rho(\gamma + \delta). \quad (2.34)$$

Furthermore using the chain rule we have that

$$\begin{aligned} \rho(c\gamma) &= \left. \frac{\partial}{\partial t} r(e^{tc\gamma}) \right|_{t=0} \\ &= c \left. \frac{\partial}{\partial t} r(e^{t\gamma}) \right|_{t=0} = c\rho(\gamma) \end{aligned} \quad (2.35)$$

Hence we have shown that ρ is a linear map of \mathfrak{g} .

Now we check that it respects the algebra $[\cdot, \cdot]$, i.e. is a Lie algebra homomorphism. Recall that we shown earlier that

$$e^{tg\gamma g^{-1}} = g e^{t\gamma} g^{-1} \quad (2.36)$$

so we find

$$r(e^{tg\gamma g^{-1}}) = r(g e^{t\gamma} g^{-1}) = r(g) r(e^{t\gamma}) r(g^{-1}). \quad (2.37)$$

Taking a derivative w.r.t t on both sides and then setting $t = 0$ we get

$$\rho(g\gamma g^{-1}) = r(g)\rho(\gamma)r(g^{-1}). \quad (2.38)$$

Now comes the final trick: we set $g = e^{t\delta}$ in the above equation, take another derivative w.r.t. t and set $t = 0$ again. The rhs becomes

$$\left. \frac{\partial}{\partial t} e^{t\delta} \rho(\gamma) e^{-t\delta} \right|_{t=0} = \rho(\delta)\rho(\gamma) - \rho(\gamma)\rho(\delta) = [\rho(\delta), \rho(\gamma)] \quad (2.39)$$

For the lhs, recall that ρ is a linear map between vector space, so that²

$$\frac{\partial}{\partial t} \rho(\kappa(t)) = \rho\left(\frac{\partial}{\partial t} \kappa(t)\right), \quad (2.40)$$

²Think of how you would differentiate $A\mathbf{v}(t)$ for a constant matrix A and t dependent vector $\mathbf{v}(t)$.

The lhs hence becomes

$$\left. \frac{\partial}{\partial t} \rho(e^{t\delta} \gamma e^{-t\delta}) \right|_{t=0} = \rho \left(\left. \frac{\partial}{\partial t} e^{t\delta} \gamma e^{-t\delta} \right|_{t=0} \right) = \rho([\delta, \gamma]) \quad (2.41)$$

Hence we have shown that

$$\rho([\delta, \gamma]) = [\rho(\delta), \rho(\gamma)], \quad (2.42)$$

i.e. we have defined a Lie algebra homomorphism. \square

REMARK: The converse to the above theorem is not always true, i.e. given a representation ρ of a Lie algebra, there does not need to be a group representation that relates to it via (2.28). We have actually encountered this before. Recall that $SO(3)$ and $SU(2)$ have isomorphic Lie algebras. Hence I can think of the Lie algebra of $SU(2)$ as a (Lie algebra) representation of $\mathfrak{so}(3)$. However, if I exponentiate the Lie algebra $\mathfrak{su}(2)$, I do not get a representation of $SO(3)$ but just $SU(2)$ in the defining representation. We will examine this in a little more detail later.

Exercise 29. Consider the Lie group G of upper triangular 2×2 matrices

$$G = \left\{ \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} \mid a, b, c \in \mathbb{R}, ac \neq 0 \right\} \quad (2.43)$$

a) Let $\mathbf{v} \in \mathbb{R}^3$, $\mathbf{v} = (v_1, v_2, v_3)$. Define an action of G on \mathbf{v} by writing

$$\mathbf{v}_m := \begin{pmatrix} v_1 & v_2 \\ 0 & v_3 \end{pmatrix} \quad (2.44)$$

and letting $g \in G$ act as

$$r(g)\mathbf{v}_m := g\mathbf{v}_m g^{-1}. \quad (2.45)$$

Convince yourself that this is a representation of G . Write the action of $r(g)$ on \mathbf{v} defined above in terms of a 3×3 matrix acting on \mathbf{v} :

$$r(g)\mathbf{v} = M(g)\mathbf{v} \quad (2.46)$$

for a 3×3 matrix $M(g)$ acting on the vector $\mathbf{v} \in \mathbb{R}^3$ in the usual way.

b) Writing $r(g)$ in terms of the matrices $M(g)$, work out the associated representation ρ of the Lie algebra \mathfrak{g} of G .

c) Check that they give a Lie algebra representation of the Lie algebra \mathfrak{g} of G , i.e. find a Lie algebra homomorphism between the Lie algebra \mathfrak{g} of G and the Lie algebra representation $\rho(\mathfrak{g})$ associated with $r(G)$.

Example 2.9. The adjoint representation of the Lie algebra is a map $ad : \mathfrak{g} \rightarrow GL(\mathfrak{g})$ which maps $\delta \in \mathfrak{g}$ to a linear map acting on \mathfrak{g} that acts on $\gamma \in \mathfrak{g}$ as

$$ad(\delta) : \gamma \rightarrow [\delta, \gamma], \quad (2.47)$$

i.e. $ad(\delta)$ is in $GL(\mathfrak{g})$. We can work out this representation is associated to the usual adjoint representation of the group, Definition (2.5), using (2.29):

$$ad(\delta)(\gamma) = \left. \frac{\partial}{\partial t} Ad(e^{t\gamma})(\gamma) \right|_{t=0} = \left. \frac{\partial}{\partial t} e^{t\delta} \gamma e^{-t\delta} \right|_{t=0} = [\delta, \gamma]. \quad (2.48)$$

Earlier, we mentioned that we can characterise Lie algebras by their structure constants f_{ab}^c once a basis $\{t_a\}$ was chosen, (1.112).

Proposition 2.2. The structure constants define a representation by setting

$$(\rho_{adj}(t_a))^b{}_c = f_{ac}^b. \quad (2.49)$$

This representation is the adjoint representation written in the basis $\{t_a\}$, i.e. the adjoint action in the basis $\{t_a\}$ is given by matrices $\rho_{adj}(t_a)$ with components $\rho_{adj}(t_a)^b{}_c = f_{ac}^b$.

Proof. :

Exercise 30.

a) Check that (2.49) defines a representation of \mathfrak{g} .

b) Show that the adjoint action in the basis $\{t_a\}$ is given by the matrices $\rho_{adj}(t_a)$ with components f_{ac}^b by showing that

$$ad(t_a)(\gamma^b t_b) = (\rho_{adj}(t_a))^b{}_c \gamma^c t_b. \quad (2.50)$$

Theorem 2.5. Let r be a complex representation of a compact Lie group G acting on V , and let ρ be the associated Lie algebra representation. Writing the basis elements of the Lie algebra \mathfrak{g} of G as $\{it_a\}$, we can choose a basis of V such that $\rho(t_a)$ are Hermitian matrices, $\rho(t_a)^\dagger = \rho(t_a)$.

Proof. :

As we have seen when using Weyl's unitarity trick, we can choose an inner form $\langle \cdot, \cdot \rangle$ on the complex vector space V such that

$$\langle r(g)v, r(g)w \rangle = \langle v, w \rangle. \quad (2.51)$$

As this is an inner form on a complex vector space, we have that for any $c \in \mathbb{C}$

$$\langle v, cw \rangle = c \langle v, w \rangle \quad \langle cv, w \rangle = \bar{c} \langle v, w \rangle. \quad (2.52)$$

With this inner form, we can choose a basis e_i on V such that

$$\langle e_i, e_j \rangle = \delta_{ij}. \quad (2.53)$$

Using this basis, we can write $r(g)$ as matrices $r(g)_{ij}$.

For $v = v_i e_i$ and $w = w_j e_j$ we now work out

$$\begin{aligned} \bar{v}_i w_j &= \langle v, w \rangle = \langle r(g)v_i e_i, r(g)w_j e_j \rangle = \langle r(g)_{ik} v_k e_i, r(g)_{jl} w_l e_j \rangle \\ &= \overline{r(g)_{ik} v_k} r(g)_{jl} w_l \langle e_i, e_j \rangle = \bar{v}_k r^\dagger(g)_{ki} r(g)_{jl} w_l \delta_{ij} = \bar{v}_k r^\dagger(g)_{ki} r(g)_{il} w_l \end{aligned} \quad (2.54)$$

i.e. $r^\dagger(g)_{ki} r(g)_{il} = \delta_{kl}$ so that $r(g)^\dagger = r(g)^{-1}$.

As we have assumed that $\{it_a\}$ is a basis of \mathfrak{g} , we can write

$$g = e^{it_a \gamma^a} \quad (2.55)$$

for some real numbers γ^a . By definition this implies that $\rho(t_a)$ satisfies

$$r(g) = e^{i\rho(t_a)}. \quad (2.56)$$

Now $r(g)^\dagger = r(g)^{-1}$ gives us $\rho(\gamma)^\dagger = \rho(\gamma)$. \square

2.4 Representations of $SU(2)$ and $SO(3)$

Representations of $SU(2)$

Here is a neat way to explicitly construct representations of $SU(2)$, and as we will see, it gives us all the irreducible ones. $SU(2)$ naturally acts \mathbb{C}^2 in the the defining representation. Given some complex polynomial $P(\mathbf{z})$ in two variables $\mathbf{z} = (z_1, z_2)$, we can then let $SU(2)$ act on $P(\mathbf{z})$ in this way. This is particularly nice if $P(\mathbf{z})$ is a homogenous polynomial of degree d , i.e. we can write

$$P(\mathbf{z}) = \sum_{k=0}^d a_k z_1^k z_2^{d-k}. \quad (2.57)$$

The space of such polynomials is a vector space Π_d of dimension $d + 1$. You can think of the $a_k \in \mathbb{C}$ as the components of the vector and the monomials as the basis vectors. Letting $SU(2)$ act on \mathbb{C}^2 , we have a corresponding induced action on the vector space of polynomials.

Proposition 2.3. The map

$$r_d(g)P := P(g^{-1}\mathbf{z}). \quad (2.58)$$

where $g^{-1} \in SU(2)$ acts on $\mathbf{z} = (z_1, z_2)$ as

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \rightarrow g^{-1} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \quad (2.59)$$

defines a representation r_d of $SU(2)$ on the complex vector space Π_d of dimension $d + 1$.

Proof. : exercises

We can now figure out the representations ρ_d of $su(2)$ that are associated with the r_d described in proposition 2.3. Let us choose $\ell_j \equiv \frac{i}{2}\sigma_j$ as the generators of the Lie algebra $su(2)$:

$$\ell_1 = \frac{1}{2} \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \quad \ell_2 = \frac{1}{2} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \ell_3 = \frac{1}{2} \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}. \quad (2.60)$$

Their action on the monomials $z_1^l z_2^{d-k}$ is then (see problem class 4):

$$\begin{aligned} \ell_1 : z_1^k z_2^{d-k} &\rightarrow -\frac{i}{2} (k z_1^{k-1} z_2^{d-k+1} + (d-k) z_1^{k+1} z_2^{d-k-1}) \\ \ell_2 : z_1^k z_2^{d-k} &\rightarrow \frac{1}{2} (-k z_1^{k-1} z_2^{d-k+1} + (d-k) z_1^{k+1} z_2^{d-k-1}) \\ \ell_3 : z_1^k z_2^{d-k} &\rightarrow i(d/2 - k) z_1^k z_2^{d-k} \end{aligned} \quad (2.61)$$

Theorem 2.6. *For every integer $d \geq 0$, there is a single finite-dimensional irreducible representations r_d of $SU(2)$ on a complex vector space Π_d of dimension $d+1$. These are all of the complex irreducible finite-dimensional representations of $SU(2)$.*

Representations of $SU(2)$: proof of the main theorem *

Before taking on the theorem, let me prove a little lemma that will be quite useful:

Lemma 2.1. *Let r be a complex representation of $SU(2)$ acting on V . Then all eigenvalues of $\rho(\sigma_i)$, for ρ the associated representation of $\mathfrak{su}(2)$, are real.*

Proof. (of the lemma):

Let us denote $\exp(i\rho(\sigma_j)) \equiv r_j$ and $\rho(\sigma_j) \equiv \rho_j$, so $\exp(i\rho_j) = r_j$. As $SU(2)$ is compact, we can choose an inner form $\langle \cdot, \cdot \rangle$ on the complex vector space V such that

$$\langle r_j v, r_j v \rangle = \langle v, v \rangle. \quad (2.62)$$

When using Weyl's unitarity trick in Theorem 2.5, we further found that we can always choose a basis such that $\rho_i^\dagger = \rho_i$. Now let v be an eigenvector of ρ_j with eigenvalue e_v and work out

$$e_v \langle v, v \rangle = \langle v, \rho_j v \rangle = \langle \rho_j v, v \rangle = \bar{e}_v \langle v, v \rangle. \quad (2.63)$$

so that e_v must be real. \square

Now we are ready to prove the theorem:

Proof. :

Let's start by slightly enlarging the scope and study representation of the Lie algebra $\mathfrak{sl}(2, \mathbb{C}) = \mathfrak{su}(2) \otimes \mathbb{C} = \mathfrak{su}_{\mathbb{C}}(2)$. As we have seen in problem class 4, irreducible representations of $\mathfrak{sl}(2, \mathbb{C})$ are in one-to-one correspondence with irreducible representations of $\mathfrak{su}(2)$. What this means is that we consider a complex instead of a real vector space over the Pauli matrices as the algebra under consideration. This allows us to define

$$H \equiv \frac{1}{2} \rho_d(\sigma_3) , \quad L_{\pm} = \frac{1}{2} (\rho_d(\sigma_1) \pm i \rho_d(\sigma_2)) . \quad (2.64)$$

These obey the algebra

$$[H, L_{\pm}] = \pm L_{\pm} , \quad [L_+, L_-] = 2H . \quad (2.65)$$

Let us start by assuming that w_n is an eigenvector of H with eigenvalue n , so that $Hw_n = nw_n$. Then

$$HL_+w_n = (L_+H + [H, L_+])w_n = (L_+H + L_+)w_n = (L_+n + L_+)w_n = (n+1)(L_+w_n) . \quad (2.66)$$

This equation means that L_+w_n is another eigenvector of H , but now the eigenvalue is $n+1$. Hence L_+ is a 'raising operator' that increases the eigenvalue n by one. A similar computation reveals that L_-w_n has eigenvalue $n-1$, so L_- is a 'lowering operator'.

As we only care about representations of $SU(2)$ we can use the lemma above and conclude that H only has real eigenvalues. As we only consider finite-dimensional vector spaces, one of these eigenvalues must be the largest. Let us call this eigenvalue m and w_m the associated eigenvector³. Then we must have

$$L_+w_m = 0 . \quad (2.67)$$

Otherwise L_+w_m would be another eigenvector with eigenvalue $m+1$, which violates the assumption that we have chosen the largest.

We can then repeatedly act with L_- to produce more eigenvectors with smaller eigenvalues. As we are looking for finite-dimensional representations, this must terminate at some point, i.e. for some $d \in \mathbb{Z}$, $(L_-)^{d+1}w_m = 0$. A basis of our representation are hence the vectors

$$w_{m-l} \equiv (L_-)^l w_m , l = 0 \cdots d . \quad (2.68)$$

³Here, I have made the tacit assumption that w_m is unique, i.e. there is only a single eigenvector with the maximal eigenvalue m . You can try to see what will happen if you repeat the following argument without this assumption, and you will find that this results in a reducible representation.

and its dimension is $d + 1$. To find out which values can appear, we introduce

$$\Delta \equiv \frac{1}{4} (\rho_d(\sigma_1)^2 + \rho_d(\sigma_2)^2 + \rho_d(\sigma_3)^2) = \frac{1}{2} (L_+ L_- + L_- L_+) + H^2 \quad (2.69)$$

We already know that $\sigma_i^2 = \mathbb{1}$ in the defining representation. That $\Delta = c\mathbb{1}$ here as well follows from Schur's lemma after observing that

$$[\Delta, H] = [\Delta, L_{\pm}] = 0. \quad (2.70)$$

This does not imply that $c = 3/4$ however, as we are not in the defining representation! To fix c , observe that

$$\Delta w_m = \left(\frac{1}{2} (L_+ L_- + L_- L_+) + H^2 \right) w_m = (L_- L_+ + H(H + 1)) w_m = m(m + 1) w_m. \quad (2.71)$$

Hence $c = m(m + 1)$. As $L_- w_{m-d} = 0$ and furthermore $L_+ L_- = \Delta - H(H - 1)$ we have that

$$\begin{aligned} 0 &= (\Delta - H(H - 1)) w_{m-d} = (m(m + 1) - (m - d)(m - d - 1)) w_{m-d} \\ &= (1 + d)(2m - d) w_{m-d} \end{aligned} \quad (2.72)$$

which implies $2m - d = 0$. As d is an integer, this implies that m takes half-integer values. By construction, these are finite-dimensional irreducible representations of $\mathfrak{sl}(2, \mathbb{C})$.

We can easily restrict all of the matrices appearing in this representation to anti-hermitian ones to find a representation of $\mathfrak{su}(2)$. As we have that $\rho(\sigma_i) = \rho(\sigma_i^\dagger)$, we get a representation of $\mathfrak{sl}(2, \mathbb{C})$ as

$$\sum_j a_j \rho(\sigma_j) \text{ for } a_j \in \mathbb{C} \quad (2.73)$$

and a representation of $su(2)$ as

$$\sum_j i a_j \rho(\sigma_j) \text{ for } a_j \in \mathbb{R}. \quad (2.74)$$

In problem class 4 we have seen using the above that irreducible representations of $\mathfrak{sl}(2, \mathbb{C})$ are in one-to-one correspondence with irreducible representations of $\mathfrak{su}(2)$, so we can think of the representations just found as representations of $\mathfrak{su}(2)$.⁴

In order to compare this with the representations ρ_d of $su(2)$ associated to the representations r_d of $SU(2)$ that we already know exist, it is convenient to rescale the basis vectors w_j as follows. We define $v_m \equiv w_m$ and

$$L_- v_k = (m + k) v_{k-1} \quad (2.75)$$

⁴Recall that the punchline here was that these representations and their irreducibility is both determined solely by representing the Pauli matrices, i.e. fixing $\rho(\sigma_j)$. In the case of $\mathfrak{sl}(2, \mathbb{C})$ its representation is a complex vector space with basis $\{\rho(\sigma_j)\}$, in the case of $\mathfrak{su}(2)$ you get a real vector space with basis $\{i\rho(\sigma_j)\}$.

which implies

$$\begin{aligned} L_+ v_k &= \frac{1}{m+k+1} L_+ L_- v_{k+1} = \frac{1}{m+k+1} (\Delta - H(H-1)) v_{k+1} \\ &= \frac{1}{m+k+1} (m(m+1) - k(k+1)) = (m-k) v_{k+1}. \end{aligned} \quad (2.76)$$

In this basis, the action of the ℓ_i is

$$\begin{aligned} \ell_1 v_{m-k} &= \frac{i}{2} (L_+ + L_-) v_{m-k} = \frac{i}{2} (k v_{m-k+1} + (d-k) v_{m-k-1}) \\ \ell_2 v_{m-k} &= \frac{1}{2} (L_+ - L_-) v_{m-k} = \frac{1}{2} (k v_{m-k+1} - (d-k) v_{m-k-1}) \\ \ell_3 v_{m-k} &= i H v_{m-k} = i(m-k) v_{d/2-k} \end{aligned} \quad (2.77)$$

where $m = d/2$ for an integer d . Comparing with (2.61) we see that these representations are identified if we associate

$$z_1^k z_2^{d-k} \simeq (-1)^k v_{m-k}. \quad (2.78)$$

This means all the representations of $\mathfrak{su}(2)$ we have found are the associated representations of the group representations r_d we already know exist.

This representation ρ_d is irreducible as can be seen as follows. Take any invariant subspace V of Π_d . By assumption the action of the ℓ_k maps any vector of V to another vector of V . As V is a complex vector space, complex linear combinations are again in V . This implies that if $P \in V$, we also have that any linear combination of

$$\ell_+^n := (z_2 \frac{\partial}{\partial z_1})^n P, \quad \ell_-^p := (z_1 \frac{\partial}{\partial z_2})^p P \quad (2.79)$$

is in V (these are just powers of the polynomial versions of raising and lowering operators). We can hence apply a suitable power of ℓ_- to map P to the single monomial z_1^d . This monomial is hence in V , which implies that any complex multiple of it is in V as well. But now we can use ℓ_+ to conclude the same for any other monomial. As the monomials are a basis of Π_d , it follows that $V = \Pi_d$. The Lie algebra representations ρ_d are hence irreducible.

This implies that r_d is irreducible as well. If $W \in \Pi_d$ is an invariant subspace of r_d , then it must be invariant under $e^{t\rho(\gamma)}$ for all t and $\gamma \in \mathfrak{su}(2)$, so in particular under $\partial/\partial t e^{t\rho(\gamma)}$ and hence under $\rho_d(\mathfrak{su}(2))$. But the Lie algebra representation ρ is irreducible as we already know.

So now we know all irreducible representations of $SU(2)$: if there were others, the associated Lie algebra representation would have had to show up in our analysis. \square

Representations of $SO(3)$

We are now ready to discuss representations of $SO(3)$. As the Lie algebra of $SO(3)$ is the same as the Lie algebra of $SU(2)$, it has the same irreducible representations. Coming to the groups, recall that there is a 2 to 1 map from $SU(2)$ to $SO(3)$ that we investigated in problem class 1 which mapped both $\mathbb{1} \in SU(2)$ and $-\mathbb{1} \in SU(2)$ to $\mathbb{1} \in SO(3)$. We can hence construct representations of $SO(3)$ from representations of $SU(2)$ if $r(-\mathbb{1}) = \mathbb{1}$. Let us look at the action of $r_d(-\mathbb{1})$ on a monomial

$$r_d(-\mathbb{1}) : z_1^k z_2^{d-k} \rightarrow (-1)^d z_1^k z_2^{d-k}. \quad (2.80)$$

This map is the identity only if d is an even integer, i.e. $m = d/2$ is an integer. We have seen that every representation of a Lie group gives us an associated representation of its Lie algebra. The above shows that the converse is not true, the representations of $so(3)$ where m is half-integer cannot come from any representations of $SO(3)$. On the other hand, we can lift any finite-dimensional representation R of $SO(3)$ to one of $SU(2)$:

Exercise 31. Show that any irreducible complex representation of $SO(3)$ also defines an irreducible complex representation of $SU(2)$.

Hence we have

Theorem 2.7. The r_d for $d = 2m$, $m \in \mathbb{Z}$ are all of the finite-dimensional complex irreducible representations of $SO(3)$.

$SO(3)$, $SU(2)$, and Spin

In physics in \mathbb{R}^3 , the half-integer m is called the spin: if there is a physical object that transforms in the representation r_d , we say it has spin $m = d/2$. This applies both to field theories, where $SO(3)$ acts on the components of a field, and to quantum mechanics, where $SO(3)$ acts on states. If $d = 0$ we have a one-dimensional representation, e.g. a scalar field, that does not transform at all, this is the spin 0 case. An ordinary vector in \mathbb{R}^3 transforms in the three-dimensional representation r_2 of $SU(2)$, so you would call a field $\phi = (\phi_1, \phi_2, \phi_3)$ transforming like a vector in \mathbb{R}^3 a ‘vector field’ as well. Here $m = 1$, so this is ‘spin 1’.

The representations of $SO(3)$ show up in most courses on quantum mechanics when treating the hydrogen atom. Using wavefunctions gives us a very concrete version of these representations: the ‘spherical harmonics’.

It is a fact of nature that there are particles of ‘spin 1/2’, e.g. the electron or quarks. You might find this irritating as we might want to classify particles according to how they transform

under space-time symmetries, i.e. $SO(3)$ for rotations, and for $m = 1/2$ we do not get a representation of $SO(3)$, but only one of its Lie algebra. One way to explain this is that in quantum mechanics, multiplying any state vector by a non-zero complex number does not change the state we are in. Taking this into account means studying projective representations, which for $SO(n)$ are in one-to-one correspondence with ordinary representations of the associated ‘spin groups’: $Spin(3) = SU(2)$.

Definition 2.13. The **spinor representation** is the **2** of $SU(2)$, and objects transforming in this representation are called spinors (of $SO(3)$). The covering group $SU(2)$ of $SO(3)$ is likewise called the ‘**spin group**’ $Spin(3)$.

REMARK: We saw earlier how to map $SU(2)$ to $SO(3)$. For the element of $SU(2)$ of the form

$$g_{SU(2)} = \begin{pmatrix} e^{i\phi/2} & 0 \\ 0 & e^{-i\phi/2} \end{pmatrix} \quad (2.81)$$

the corresponding element in $SO(3)$ was

$$g_{SO(3)} = \begin{pmatrix} \cos(\phi) & \sin(\phi) & 0 \\ -\sin(\phi) & \cos(\phi) & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (2.82)$$

Let us assume we are performing a rotation by 360° using the (usual) rotation group $SO(3)$ in \mathbb{R}^3 , i.e. we let ϕ go from 0 to 2π in the above matrix $g_{SO(3)}$. In the corresponding $SU(2)$ matrix $g_{SU(2)}$ this takes us from $\mathbb{1}$ to $-\mathbb{1}$, i.e. we do not come back to where we started from, and need to let ϕ go from 0 to 4π to return to $\mathbb{1}$. In this sense, a spinor needs to be rotated by 720° for a full rotation!

2.5 Representations of $SU(n)$ and other Lie Groups *

For more general Lie groups such as $SU(n)$, you will not be surprised to hear that there is a richer representation theory. We know already two representations: the defining and the adjoint. Instead of developing the general theory, we will only try to sketch how one might go about creating such representations. Given a representation of a group G , this implies representations of any subgroup H by simply restricting the homomorphism $r : G \rightarrow GL(V)$ to $H \subset G$. Every group $SU(n)$ for $n > 2$ contains many copies of $SU(2)$ as subgroups, and we have seen how we could construct representations of $SU(2)$ using the operators H, L_+, L_- . This motivates to try and lift the method used for $SU(2)$ to that of $SU(n)$ by writing the (complexified) Lie algebra of $\mathfrak{su}(n)$ in terms of number operators H_i , lowering operators and raising operators. This is called a ‘Cartan-Weyl basis’ and leads to what are called ‘root systems’ which can in turn be used to classify certain classes of Lie algebras. Such a root system is shown in figure 2.

2.6 Tensor Representations*

Definition 2.14. Given two vector spaces V and W we can form their **tensor product** $V \otimes W$. Let $e_i, i = 1.. \dim V$, be a basis of V and $f_j, j = 1.. \dim W$, be a basis of W . Then $V \otimes W$ is a vector space with basis consisting of tuples (e_i, f_j) (also written as $e_i \otimes f_j$).

REMARK: It follows from the definition that $\dim V \otimes W = \dim V \cdot \dim W$. Computing with tensor products works almost the same as with usual products, we have

$$\begin{aligned} v \otimes w + v' \otimes w &= (v + v') \otimes w \\ v \otimes w + v \otimes w' &= v \otimes (w + w'), \end{aligned} \tag{2.83}$$

and for $c \in \mathbb{R}$ (or \mathbb{C})

$$c(v \otimes w) = (cv) \otimes w = v \otimes (cw). \tag{2.84}$$

However $v \otimes w \neq w \otimes v$: the first slot is reserved for vectors from V and the second for vectors from W , so writing $w \otimes v$ does not even make sense if $v \in V$ and $w \in W$. Not every vector in $V \otimes W$ can be written as a product, e.g. $v \otimes w + v' \otimes w'$ for $v \neq v'$ and $w \neq w'$.

Example 2.10. Consider $\mathbb{R}^3 \otimes \mathbb{R}^3$ and let e_1, e_2, e_3 be a basis of the first \mathbb{R}^3 and f_1, f_2, f_3 of the second. A basis of $\mathbb{R}^3 \otimes \mathbb{R}^3$ is then

$$\begin{aligned} e_1 \otimes f_1, e_1 \otimes f_2, e_1 \otimes f_3, \\ e_2 \otimes f_1, e_2 \otimes f_2, e_2 \otimes f_3, \\ e_3 \otimes f_1, e_3 \otimes f_2, e_3 \otimes f_3, \end{aligned} \tag{2.85}$$

whereas $\mathbb{R}^3 \oplus \mathbb{R}^3$ has a basis

$$e_1, e_2, e_3, f_1, f_2, f_3. \tag{2.86}$$

Note that whereas $\mathbb{R}^3 \oplus \mathbb{R}^3$ is six-dimensional, $\mathbb{R}^3 \otimes \mathbb{R}^3$ is 9-dimensional. Note that you can naturally think of $\mathbb{R}^3 \otimes \mathbb{R}^3$ as the vector space of real 3×3 matrices: we can write any element of $\mathbb{R}^3 \otimes \mathbb{R}^3$ as

$$v = \sum_{ij} a_{ij} e_i \otimes f_j. \tag{2.87}$$

What makes tensor products interesting in the present context is that we can form new representations out of old ones by tensoring the vector spaces they act on.

Example 2.11. $2 \otimes \bar{2}$ Let's say you have a vector space \mathbb{C}^2 that 'lives' in the fundamental representation of $SU(2)$, and one \mathbb{C}^2 that lives in the anti-fundamental and you form their tensor product. The question we are asking is: how does $SU(2)$ act on the tensor product? For a vector in \mathbb{C}^2 in the defining representation we have

$$z \rightarrow gz \tag{2.88}$$

and in the anti-fundamental

$$z \rightarrow \bar{g}z. \quad (2.89)$$

This is how we would things down using a chosen fixed basis, $v = (z_1, z_2)$, so we might also write this more abstractly as (for the first case):

$$v = \sum_i z_i e_i \rightarrow g_{ij} z_j e_i \quad (2.90)$$

We can think of this as either acting with g on z (this is called the **active interpretation**) or as acting with g^T on the tuple of basis vectors (this is called the **passive interpretation**):

$$e_j \rightarrow g_{ij} e_i = g_{ji}^T e_i \quad (2.91)$$

i.e.

$$\begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \rightarrow g^T \begin{pmatrix} e_1 \\ e_2 \end{pmatrix}. \quad (2.92)$$

We can use either if we like it, and this will help us to figure out how to act on elements of $\mathbb{C}^2 \otimes \mathbb{C}^2$. As the first copy transforms with g and the second with \bar{g} we have

$$v = \sum_i a_{ij} e_i \otimes f_j \rightarrow \sum_{ijkl} a_{ij} (g_{ki} e_k) \otimes (\bar{g}_{lj} f_l) = \sum_{ijkl} g_{ki} a_{ij} \bar{g}_{lj} e_k \otimes f_l \quad (2.93)$$

so in summary the vectors in $\mathbb{C}^2 \otimes \mathbb{C}^2$ behave as

$$a_{ij} \rightarrow \sum_{kl} g_{ik} a_{kl} g_{lj}^\dagger, \quad (2.94)$$

i.e. if we collect the a_{ij} in a matrix A we get

$$A \rightarrow g A g^\dagger. \quad (2.95)$$

We can repeat the same logic to find out how arbitrary representations acting on vector spaces V and W act on $V \otimes W$:

Definition 2.15. Let $r_V(G) \in GL(V)$ and $r_W(G) \in GL(W)$, and let the components of these matrices be $r_V(G)_{ij}$ and $r_W(G)_{ab}$. Then the **tensor product representation** $r_{V \otimes W}$ acts on a vector $U \in V \otimes W$ with components U_{ia} as

$$U'_{ia} := r_V(G)_{ij} r_W(G)_{ab} U_{jb}. \quad (2.96)$$

Example 2.12. $2 \otimes \bar{2} = 1 \oplus 3$

Continuing example 2.11 we know that we can decompose the representation acting on $\mathbb{C}^2 \otimes \mathbb{C}^2$ into irreducible representations. But which ones? As we have seen, $SU(2)$ acts on a_{ij} as

$$A \rightarrow A' \quad a'_{ij} = g_{ik} a_{kl} (g^\dagger)_{lj} \quad (2.97)$$

or in matrix notation

$$A \rightarrow A' = gAg^\dagger. \quad (2.98)$$

The trace of A hence transforms as

$$\text{tr} A \rightarrow \text{tr} gAg^{-1} = \text{tr} A. \quad (2.99)$$

Now what this implies is that the representation $\mathbf{2} \otimes \bar{\mathbf{2}}$ is reducible, as we can never map matrices with a vanishing trace to ones with a non-vanishing trace. Let's try to understand this a bit more clearly. The matrices A have the form

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \quad (2.100)$$

and we think of the four complex components a_{ij} as components of a vector in a vector space V isomorphic to \mathbb{C}^4 that we chose to write as a matrix. Within this vector space there is a complex three-dimensional vector subspace W defined by $a_{11} + a_{22} = 0$, and as (2.99) shows, the group action on V maps vectors in W again to vectors in W , i.e. W is an invariant subspace. More concretely, W is the subspace of matrices of the form

$$W = \left\{ A \mid A = \begin{pmatrix} z_1 & z_2 \\ z_3 & -z_1 \end{pmatrix}, (z_1, z_2, z_3) \in \mathbb{C}^3 \right\}. \quad (2.101)$$

You might want to convince yourself that this is indeed a vector subspace. Similarly W^\perp is the one-dimensional subspace containing matrices of the form

$$W^\perp = \left\{ A \mid A = \begin{pmatrix} z_4 & 0 \\ 0 & z_4 \end{pmatrix}, z_4 \in \mathbb{C} \right\}. \quad (2.102)$$

which again form an invariant subspace under the group action (2.99). The inner form under which this is $^\perp$ is just the standard inner form on \mathbb{C}^4 , which we can write as $\langle A, A' \rangle = \sum_{i,j} \bar{a}_{ij} a'_{ij}$ using two matrices A, A' . Also note that for any A we can write

$$A = \begin{pmatrix} z_1 & z_2 \\ z_3 & -z_1 \end{pmatrix} + \begin{pmatrix} z_4 & 0 \\ 0 & z_4 \end{pmatrix}. \quad (2.103)$$

The above shows that the representation $\mathbf{2} \otimes \bar{\mathbf{2}}$ is not irreducible, but decomposes into a one-dimensional and a three-dimensional complex representation, i.e. we can write $\mathbf{2} \otimes \bar{\mathbf{2}} = \mathbf{1}^\perp \oplus \mathbf{1}$. The only remaining thing to show is hence that $\mathbf{1}^\perp$ transforms in the $\mathbf{3}$ of $SU(2)$. The action here is the same as the adjoint representation of $SU(2)$, except that we are acting on a complex vector space of dimension three instead of a real one. The irreducibility of the adjoint representation implies that there is no invariant complex subspace if we act on \mathbb{C}^3 instead of \mathbb{R}^3 , so this is the $\mathbf{3}$ of $SU(2)$.

Exercise 32. Consider the representation $\mathbf{n} \otimes \bar{\mathbf{n}}$ of $SU(n)$. Explain why this is always reducible. Can you identify the irreducible representations and invariant subspaces?

Proposition 2.4. $\mathbf{2} \otimes \mathbf{2} = \mathbf{1} \oplus \mathbf{3}$

Proof. :

Exercise 33. a) Find the transformation of elements of $\mathbf{2} \otimes \mathbf{2}$.

b) Show that the representations $\mathbf{2}$ and $\bar{\mathbf{2}}$ are isomorphic by showing they are related by a change of basis

$$z' = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} z \quad (2.104)$$

[Note: of course, \bar{v} transforms also as $\bar{v} \rightarrow \bar{g}\bar{v}$ if $v \rightarrow gv$. In a complex vector space, complex conjugation is not a change of basis however!]

c) Use the above to argue that $\mathbf{2} \otimes \mathbf{2} = \mathbf{1} \oplus \mathbf{3}$. Can you identify the invariant subspaces?

REMARK: More generally, tensor products can be decomposed into irreducible representations, i.e. we may write

$$r_{V \otimes W}(G) = \oplus_i r_{V_i}(G) \quad (2.105)$$

whenever we know any representation of G can be decomposed into irreducible representations. The change of basis relating the natural basis of the tensor product to a basis showing the decomposition on the right hand side of the above equation is a well-known problem, and the coefficients appearing in the change of basis are called 'Clebsch-Gordan coefficients'. Details on this can be found in most books on quantum mechanics or [Sternberg, 1995], [Hall, 2003], and [Chaichian and Hagedorn, 1998] for the more mathematically minded.

REMARK: * There are a number of examples in physics in which $\mathbf{2} \otimes \bar{\mathbf{2}} = \mathbf{1} \oplus \mathbf{3}$ and $\mathbf{2} \otimes \mathbf{2} = \mathbf{1} \oplus \mathbf{3}$ plays an important role in organizing degrees of freedom of a theory. Two important ones are explained below:

spin $\frac{1}{2}$

The rotation group in 3D is the group $SO(3)$. As we have seen, $SO(3) = SU(2)/\mathbb{Z}_2$, and it turns out that the relevant group describing the behavior of rotations acting on fermions (such as electrons, protons, etc ..) is in fact is $SU(2)$ and not $SO(3)$. The relation between usual rotations and maps in $SU(2)$ is exactly given by the homomorphism we constructed earlier, more on this topic will be discussed later. This means we can write down the wavefunction of a fermion as

$$\psi = \begin{pmatrix} \psi_+ \\ \psi_- \end{pmatrix} \quad (2.106)$$

which lives in the $\mathbf{2}$ representation of $SU(2)$ under rotations. If you have a system composed of two fermions ψ_1 and ψ_2 , the total wavefunction Ψ is a tensor product of the two wavefunctions⁵

$$\Psi = \psi_i \otimes \psi_j \quad (2.107)$$

and hence lives in the $\mathbf{2} \otimes \mathbf{2}$. Decomposing this into irreducible representations, we find a singlet and a triplet of wavefunctions. This is what physicists sometimes call ‘**addition of angular momentum**’ and it explains why a Helium atom or Positronium have a singlet (‘para-’) or triplet (‘ortho-’) behavior under rotations. Note that the triplet we found behaves just as the adjoint of $SU(2)$, which corresponds to the usual defining (‘vector’) representation of $SO(3)$.

quarks

The two lightest quarks are the up and the down quark. They have nearly identical masses, but differ in their electric charges. They can form bound states which are called Baryons, and many different of these were found since the 40s. Until the discovery of quarks, people were bewildered by how many there were and how to organize them into some sort of pattern. The simplest bound states, called mesons, contain only two quarks. The force binding them together is the strong nuclear force which is a lot stronger than electromagnetism, and from the perspective of the strong force, the up and the down quark look identical if we forget the small mass difference. We can combine their wavefunctions into one

$$\psi_q = \begin{pmatrix} \psi_u \\ \psi_d \end{pmatrix} \quad (2.108)$$

and the statement that they are identical in strong interactions means there is an $SU(2)$ symmetry acting in the $\mathbf{2}$ on ψ_q . Because this looks exactly like the $SU(2)$ action of rotations on fermions, it was called ‘**iso-spin**’ which is a terrible name there is no other relation to spin than this. But this means that bound states of e.g. a quark and an anti-quark transform in the $\mathbf{2} \otimes \bar{\mathbf{2}} = \mathbf{1} \oplus \mathbf{3}$ representation. The triplet are called ‘pions’ (π^+ , π^- , π^0).

In fact, there is a third quark called the strange quark which also has (almost) the same mass as up and down. This enhances the $SU(2)$ to an $SU(3)$ and we should be studying $\mathbf{3} \otimes \bar{\mathbf{3}} = \mathbf{1} \oplus \mathbf{8}$ for mesons and

$$\mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3} = \mathbf{1} \oplus \mathbf{8} \oplus \mathbf{8} \oplus \mathbf{10} \quad (2.109)$$

for Baryons which are made up of three quarks.

Historically, this way of thinking was in fact used to motivate the existence of quarks by Murray Gell-Mann and Yuval Ne’eman in 1961 as they saw that observed particles could be fit into this pattern. They called it the ‘Eightfold way’ [Gell-Mann and Ne’eman, 1964] in a nod to Buddhism and since the adjoint of $SU(3)$ is eight-dimensional. However, one particle in the $\mathbf{10}$ had not been seen in experiments yet, so they predicted it. It is now called the Ω^- and was discovered in 1964 which among other things earned Gell-Mann a Nobel price in 1969. If you want to read more about this story, [Griffiths, 1987] is a good starting point.

⁵In fact, it is an antisymmetric version (under exchange) if you have twice the same particle. I will ignore this in the present discussion as it does not really alter the conclusions.

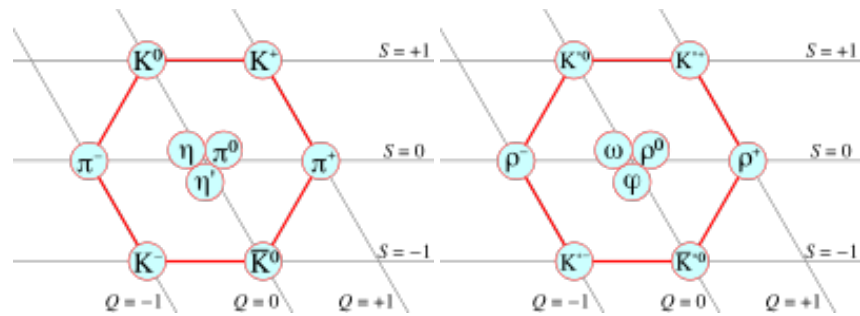


Figure 2.1: Mesons whose structure can be understood as representations of $\mathbf{3} \otimes \bar{\mathbf{3}}$ of $SU(3)$. As quarks are also fermions, mesons transform as $\mathbf{2} \oplus \bar{\mathbf{2}}$ under rotations, which makes them split into a singlet ('scalar') and triplet ('vector'). The particles on the left are scalar mesons and the ones on the right are vector mesons.

Bibliography

- Cahn, R. (2014). *Semi-Simple Lie Algebras and Their Representations*. Dover Publications.
- Chaichian, M. and Hagedorn, R. (c1998.). *Symmetries in quantum mechanics* : Graduate students series in physics. Institute of Physics Pub., Bristol.
- Gell-Mann, M. and Ne'eman, Y. (1964). *The Eightfold Way*. Frontiers in Physics : a lecture note and reprint series. W.A. Benjamin.
- Griffiths, D. (1987). *Introduction to Elementary Particles*. John Wiley and Sons, New York, USA.
- Hall, B. (2003). *Lie groups, Lie algebras, and representations: an elementary introduction*. Springer.
- Harris, W., Fulton, W., and Harris, J. (1991). *Representation Theory: A First Course*. Graduate Texts in Mathematics. Springer New York.
- Hartnett, K. (2023). A new kind of symmetry shakes up physics. www.quantamagazine.org/a-new-kind-of-symmetry-shakes-up-physics-20230418/.
- Howe, R. (1983). Very basic lie theory. *The American Mathematical Monthly*, 90(9):600–623.
- Jänich, K. and Levy, S. (1995). *Topology*. Undergraduate Texts in Mathematics. Springer New York.
- Nakahara, M. (2003). *Geometry, topology and physics*. Taylor and Francis.
- Peskin, M. and Schroeder, D. (1995). *An Introduction To Quantum Field Theory*. Frontiers in Physics. Avalon Publishing.

Scheck, F. (2010). *Mechanics: From Newton's Laws to Deterministic Chaos*. Graduate Texts in Physics. Springer Berlin Heidelberg.

Sternberg, S. (1995). *Group Theory and Physics*. Cambridge University Press.

Woit, P. (2017). *Quantum Theory, Groups and Representations: An Introduction*. Springer International Publishing.