Part III: Symmetries, Fields and Particles
Examples Sheet 1 Solutions

Please send all comments and corrections to mjb318@cam.ac.uk and jmm232@cam.ac.uk.

1. The dihedral group D_4 describes the symmetries of a square and is generated by a $90°$ rotation $r = R\left(\frac{\pi}{2}\right)$ and a reflection s (say, about the vertical symmetry axis).

(a) Write the group multiplication table for the dihedral group D_4.

(b) What are the subgroups of D_4?

(c) What are the conjugacy classes of D_4?

(d) Which of the subgroups of D_4 are normal?

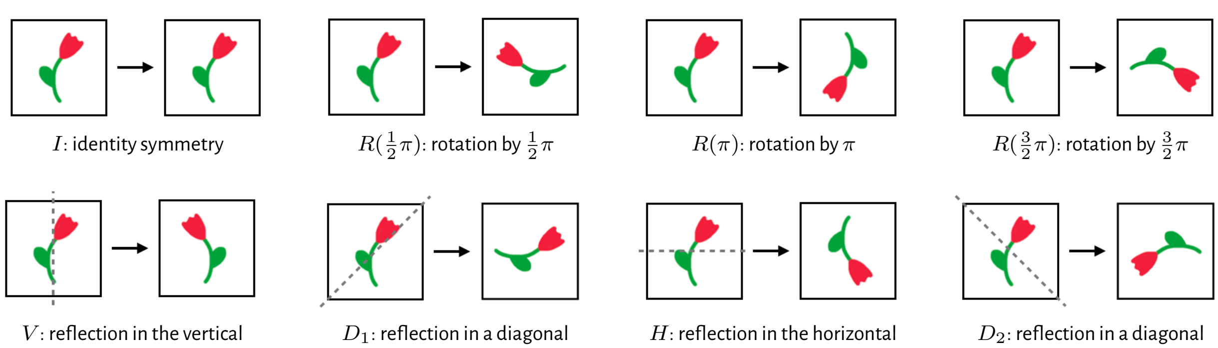
(e) Can D_4 be written as the non-trivial direct product of some of its subgroups?

➾ **Solution:** Before we start this question, it's useful to provide some background on the dihedral groups D_n. We begin with the following geometrical definition of the dihedral group D_n:

> **Definition:** The *dihedral group* D_n is the set of all rigid symmetries of a regular plane polygon with n vertices, together with the operation of composition of symmetries.

The term 'rigid' in this definition means we are not allowed to squash or bend the polygon, restricting us entirely to *reflections* and *rotations* of the polygon. Note that this definition of D_n as the 'set of symmetries' of an object automatically makes it a group, since sets of symmetries of objects always form groups.

Specialising to this question, the dihedral group D_4 is the group of rigid symmetries of a square. All of the possible elements of D_4 - namely the rotational and reflectional symmetries of a square - are shown in the figure below. By convention, the rotations are always assumed to be anticlockwise; for reflections, the line of reflection is shown.



In the diagrams above, we included a picture of a flower to help visualise what each of the symmetries actually does. The flower isn't something that needs to be preserved by a symmetry, just the square itself, else the symmetry group is trivial!

Using the pictures above, it's possible to calculate compositions of the symmetries. For example, consider composing a rotation followed by a reflection:

$$D_1 \circ R(\tfrac{3}{2}\pi) = V.$$

To see that the left hand side is equal to the right hand side, follow the pictures! First, we look at the picture corresponding to $R(\frac{3}{2}\pi)$; in this picture, the flower has been rotated by $\frac{3}{2}\pi$. Next, we carefully imagine what would happen to the rotated flower if we applied the symmetry D_1, i.e. a reflection in one of the diagonals. We then look for the picture that matches - in this case, a reflection in the vertical, V.

It is possible to calculate all products of group elements in this way, but it is extremely cumbersome. It would be much more useful to have a purely *algebraic* way of multiplying group elements together. This motivates the idea of a *group presentation*, which comprises:

- A small set of elements S, called the *generators* of the group, in terms of which all other elements of the group can be written.
- A set of *relations* R which tell us how the generators of the group interact with one another.

A group specified in this 'group presentation' way is usually written as $\langle S|R\rangle$.

We can write the dihedral group D_4 in 'group presentation' form as follows:

- First, we note that all the rotations can be written as powers of the rotation by $\frac{1}{2}\pi$; explicitly, we have $R(\frac{1}{2}\pi)$, $R(\pi) = R(\frac{1}{2}\pi)^2$ and $R(\frac{3}{2}\pi) = R(\frac{1}{2}\pi)^3$. This suggests that we define a generator of D_4 via $r := R(\frac{1}{2}\pi)$, in terms of which all the rotations can be written as r, r^2 and r^3. We can also think of the identity as a rotation, if we define the zeroth power of a group element to be the identity, $e = r^0$.
- The rotations also obey a non-trivial *relation*. The group cannot contain r^4, r^5, etc, as distinct elements - these are equivalent to rotations we already have. Hence we demand $r^4 = e$; that is, four rotations by $\frac{1}{2}\pi$ are equivalent to doing nothing.
- Turning to the reflections, we note that the reflections D_1, H and D_2 above are equivalent to performing the rotation given in the corresponding position in the first row, and then performing the reflection V. Hence we can write $V = VI$, $D_1 = VR(\frac{1}{2}\pi)$, $H = VR(\pi)$ and $D_2 = VR(\frac{3}{2}\pi)$. This suggests that another useful generator for our group is the reflection $a := V$, in terms of which all other reflections can be written: a, ar, ar^2, ar^3.
- The reflections obey a non-trivial relation too. Since a^2, a^3, etc, are equivalent to existing reflections (considering $e = a^0$ to be a reflection), we must enforce the condition $a^2 = e$.
- At this point, we have written down all group elements in terms of the generators a, r; explicitly, we have

  $$D_4 = \{e, r, r^2, r^3, a, ar, ar^2, ar^3\}.$$

  It is also possible to compute certain products of elements using the relations we already have, for example: $ar^2 \cdot r^3 = ar^5 = ar$, using the relation $r^4 = e$.

  However, some products cannot be evaluated, for example $r^2 \cdot a$ cannot be simplified using the relations $a^2 = e$ and $r^4 = e$. We need to be able move r's past a's to evaluate such products. This gives rise to our final relation:

  $$ra = ar^{-1}.$$

  At first sight, this looks like a very strange condition to impose. However, it has a very simple geometric interpretation. On the left hand side, we have performed a reflection, a, followed by a $\frac{1}{2}\pi$-rotation, r. This relation tells us that this is equivalent to a rotation by $-\frac{1}{2}\pi$-rotation, r^{-1}, followed by the same reflection, a. All it's telling us is that a reflection followed by a rotation is the same as a rotation by the same amount, but with the opposite orientation, followed by the same reflection!

Hence, we have constructed a group presentation for the dihedral group,

$$D_4 = \langle a, r \,|\, r^4 = e, a^2 = e, ra = ar^{-1}\rangle\,,$$

which is the form of the dihedral group you met in lectures. Generalising to an n-vertex regular plane polygon again, we can write:

$$D_n = \langle a, r \,|\, r^n = e, a^2 = e, ra = ar^{-1}\rangle\,.$$

(a) Now we have reviewed the algebraic form of the dihedral group - and how this form connects with the geometric form of the group - we can begin the question proper. First, we are asked to construct a *group multiplication table* for D_4. Recall that a group multiplication table for the group G consists of rows labelled by the group elements $g \in G$, and columns also labelled by the group elements $h \in G$, with the entry in the (g, h)th position given by the product of elements gh.

Hence, to construct a group multiplication table for D_4, we simply need to evaluate each of the products gh for each $g, h \in D_4$. Most of these products are easy to evaluate using the algebraic form of the group we derived above, together with the relations obeyed by each of the elements. For example, we have:

$$r^2 \cdot ar^2 = r \cdot ar^{-1}r^2 = a \cdot r^{-2}r^2 = a,$$

by repeated application of the relation $ra = ar^{-1}$.

Carrying out the multiplications for all possible combinations of the group elements, we arrive at the group multiplication table:

	e	r	r^2	r^3	a	ar	ar^2	ar^3
e	e	r	r^2	r^3	a	ar	ar^2	ar^3
r	r	r^2	r^3	e	ar^3	a	ar	ar^2
r^2	r^2	r^3	e	r	ar^2	ar^3	a	ar
r^3	r^3	e	r	r^2	ar	ar^2	ar^3	a
a	a	ar	ar^2	ar^3	e	r	r^2	r^3
ar	ar	ar^2	ar^3	a	r^3	e	r	r^2
ar^2	ar^2	ar^3	a	ar	r^2	r^3	e	r
ar^3	ar^3	a	ar	ar^2	r	r^2	r^3	e

(b) Our next task is to compute all subgroups of the group D_4. Recall that a *subgroup* H of a group G is a subset $H \subseteq G$ that constitutes a group in its own right (with the same group operation as G). If H is a subgroup of G, we write $H \leq G$.

In order to find the subgroups then, we could just check whether each subset $H \subseteq G$ of G constitutes a group - but this would take a very long time! Fortunately, for finite groups G, there is a very useful theorem we can apply - *Lagrange's theorem* - which tells us how large subgroups $H \leq G$ can be. Recall the statement of Lagrange's theorem from lectures:

Lagrange's Theorem: Let G be a finite group, and let $H \leq G$ be a subgroup of G. Then the order of H divides the order of G; or, in symbols, $|H| \mid |G|$.

Hence, in our hunt for subgroups of D_4, we need only consider subsets of size $1, 2, 4$ and 8, since these are the only factors of $|D_4| = 8$. Let's begin:

- Clearly $\{e\}$ and $\{e, r, r^2, r^3, a, ar, ar^2, ar^3\}$ are the trivial subgroups of D_4 of orders 1 and 8 respectively.
- There are five subgroups of order 2, given by $\{e, a\}$, $\{e, ar\}$, $\{e, ar^2\}$, $\{e, ar^3\}$, $\{e, r^2\}$. These are the only possible subgroups of order 2 because no other elements in D_4 are self-inverse, as we can easily see from the group multiplication table.

- The only remaining subgroups are subgroups of order 4. To build them up, let's think about what our subgroup H would look like if certain elements were included. We must always include $e \in H$, the identity. Now consider adding the following elements:

  – Suppose that $r \in H$. Then $r^2, r^3 \in H$ by closure. Hence the minimum size of H is $H = \{e, r, r^2, r^3\}$ which is indeed a subgroup of size 4 (it's isomorphic to the cyclic group, C_4). Adding any of the reflections must then generate the whole group, since the only allowed subgroup size bigger than 4 is 8 by Lagrange's theorem.

    A similar thing happens if $r^3 \in H$. We must have $r^3 \cdot r^3 = r^2 \in H$ and $r^3 \cdot r^2 = r \in H$ by closure. Thus we generate $H = \{e, r, r^2, r^3\}$ by closure again.

  – Suppose that $r^2 \in H$. This time, multiplying the group element by itself lots of times doesn't help us, because $r^2 \cdot r^2 = e$. Hence we don't generate any new elements this way. Instead, we have the freedom to include some other element. We already know that if we include one of the other rotations, r, r^3, we will generate $H = \{e, r, r^2, r^3\}$. Hence we choose instead to include one of the reflections.

    Choosing $a \in H$, we generate the subgroup $\{e, r^2, a, ar^2\}$. Choosing $ar \in H$, we generate the subgroup $\{e, ar, r^2, ar^3\}$. Both of these subgroups are isomorphic to the direct product of cyclic groups $C_2 \times C_2$, which is sometimes called the *Klein four-group* and written V_4.

    Choosing $ar^2 \in H$ or $ar^3 \in H$ just generates the subgroups $\{e, r^2, a, ar^2\}$ and $\{e, ar, r^2, ar^3\}$ again, so we don't find anymore this way. Hence we have found all subgroups of size four containing $r^2 \in H$.

  – Finally, suppose that no rotations are included in H. That is, H consists of e and three reflections, chosen from a, ar, ar^2 and ar^3. But note that if we include any two distinct reflections, we must include a rotation, since:

    $$ar^n ar^m = r^{-n} a^2 r^m = r^{m-n},$$

    is a non-trivial rotation for $m \neq n$ modulo 4.

  It follows that the only possible subgroups of size four are $\{e, r, r^2, r^3\}$, $\{e, r^2, ar, ar^3\}$, $\{e, r^2, a, ar^2\}$.

Hence we have enumerated all the subgroups: there is 1 of order 1, there are 5 of order 2, there are 3 of order 4 and there is 1 of order 8. In total there are 10 subgroups.

An aside: subgroups and symmetry breaking

In the above analysis, we treated the determination of subgroups as a purely algebraic task. But what is the *geometric* interpretation of subgroups? If we suppose that the group G is the symmetry group of some object X, then we can think of any subgroup $H \leq G$ as the symmetry group of X after it has been made 'less symmetric' in some fashion. When we make X 'less symmetric' in some way, we say that we have *spontaneously broken* the symmetry group G to the symmetry group H.

The idea of making an object X 'less symmetric' in some way has important applications in physics, as it turns out. Consider for example a symmetric double-well potential in one dimension. The resulting theory has a $\mathbb{Z}_2$ symmetry that results from reflecting the potential and swapping the wells. However, when we add a particle to the system, its vacuum state will reside in precisely *one* of the wells, breaking the symmetry of the theory. We say that the $\mathbb{Z}_2$ symmetry is *spontaneously broken* to the trivial symmetry group $\{e\} \leq \mathbb{Z}_2$.[1]

[1] In fact, this is exactly what happens in the famous *Higgs mechanism*! In this case, an $SU(2) \times U(1)$ symmetry is broken to a $U(1) \leq SU(2) \times U(1)$ symmetry by the vacuum state of the Higgs field. This is something you will see in the Part III Standard Model course.

In our case, we can imagine each of the subgroups of D_4 as what's leftover when we break some of the symmetry of the square. Explicitly, we have:

- The improper subgroup $D_4 \leq D_4$ corresponds to leaving the square alone.
- The subgroup of rotations, $\{e, r, r^2, r^3\} \leq D_4$, is the symmetry group of the square after it has been given a specific *orientation*, e.g. by drawing arrows on each of the edges of the square.
- The subgroups $\{e, r^2, a, ar^2\} \leq D_4$ and $\{e, r^2, ar, ar^3\} \leq D_4$ correspond to destroying some rotational symmetry and some reflectional symmetry. For example, the first subgroup corresponds to the symmetry group of a square with a horizontal (or vertical) line inserted through its centre, whilst the latter subgroup corresponds to the symmetry group of a square with one of its diagonals inserted.
- The order 2 subgroups correspond to adding even more features to the square. For example, we can add a direction to the horizontal lines or diagonal lines drawn in the previous point to generate the subgroups $\{e, a\}$, $\{e, ar\}$, $\{e, ar^2\}$, $\{e, ar^3\}$. We could add extra arrows to two sides of the square in the second point to generate the subgroup $\{e, r^2\}$.

(c) Returning to the question, we are now asked to determine the conjugacy classes of D_4. We begin by recalling the definition of *conjugacy* and *conjugacy classes*:

Definition: Let G be a group. We say that the elements $g, g' \in G$ are *conjugate* if there exists some element $h \in G$ such that $hgh^{-1} = g'$.

The relation 'is conjugate to' induces an equivalence relation on the group, and the equivalence classes are called *conjugacy classes*.

Geometrically speaking, conjugate elements effect the 'same transformation' on an object, just 'viewed from a different perspective'; we can make this precise as follows. Suppose that the group G acts on the object X, and suppose that you witness the transformation:

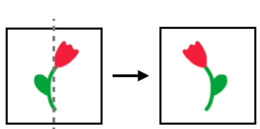
$$X \mapsto g \cdot X.$$

What would this transformation look like if you'd been standing on your head? Or on the other side of the page? Or with your head tilted clockwise by $\pi/2$? Provided that you have transformed yourself in a manner consistent with one of the group elements $h^{-1} \in G$ (if you tilt your head clockwise by $\pi/2$, then the object appears to rotate anticlockwise by $\pi/2$, so the inverse is needed here), then the transformation would have looked like:

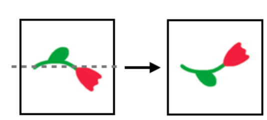
$$h \cdot X \mapsto h \cdot (g \cdot X) = (hg) \cdot X = (hgh^{-1}) \cdot (h \cdot X).$$

That is, from your perspective, it would look like the object (which initially looked like $h \cdot X$ to you) had been transformed by the group element hgh^{-1} instead of g. Hence, to emphasise the point again: *conjugate elements effect the same transformation, just viewed from a different perspective*.

As a concrete example, let's consider one of the symmetries of the square above. Consider what happens to the reflection in the vertical a when we rotate all our pictures by $\pi/2$ clockwise (corresponding to tilting your head by $\pi/2$ anticlockwise):



a: vertical reflection



$r^3ar^{-3} = ar^2$

We see that, from the new perspective, the vertical reflection a now looks like a horizontal reflection ar^2. We transformed the initial object by the rotation $r^{-1} = r^3$, and hence we have derived *geometrically* the conjugacy relation $r^3ar^{-3} = ar^2$ (which you can verify using the group algebra!).

In particular, we can just write down the conjugacy classes:

- $\{e\}$ is a conjugacy class, because the identity always just leaves things alone, regardless of the perspective we view the transformation from.
- No amount of changing perspective will make a reflection look like a rotation. So the rotations and reflections must lie in separate conjugacy classes. Furthermore, no amount of changing perspective will make a reflection in a diagonal look like a reflection in the vertical or horizontal; hence they must also lie in separate conjugacy classes.
- Now, the horizontal and vertical reflections are clearly conjugate via a $\pi/2$ rotation. It follows that $\{a, ar^2\}$ is a complete conjugacy class. Similarly, the two diagonal reflections are clearly conjugate via a $\pi/2$ rotation, and hence it similarly follows that $\{ar, ar^3\}$ is a complete conjugacy class.
- Finally, consider the rotations. A π rotation looks like a π rotation whether we look at it back-to-front, or from a different orientation. Hence $\{r^2\}$ forms its own conjugacy classes. However, a $\pi/2$ rotation and a $3\pi/2$ rotation are related by a reflection (i.e. looking at one from the other side of the paper), hence $\{r, r^3\}$ forms another conjugacy class.

Hence we have seen (without performing any calculations!) that the conjugacy classes are precisely: $\{e\}$, $\{r^2\}$, $\{r, r^3\}$, $\{a, ar^2\}$ and $\{ar, ar^3\}$. There are five in total.

(d) We can use the information on conjugacy classes to determine the normal subgroups. Recall that a subgroup H is normal if and only if $gHg^{-1} = H$ for all $g \in G$; using our new knowledge of the meaning of conjugation, this should be interpreted as the subgroup 'looking the same from all perspectives'. In particular, this immediately implies that a subgroup is normal *if and only if* it is the union of some of the conjugacy classes.[2]

Let's go through each of the subgroups in turn then and see whether they are normal:

- The subgroup $\{e\}$ is normal because it is the union $\{e\}$.
- The subgroup $\{e, r^2\}$ is normal because it is the union $\{e\} \cup \{r^2\}$. However, none of the other subgroups of order two are normal.
- The subgroup $\{e, r, r^2, r^3\}$ is normal because it is the union $\{e\} \cup \{r, r^3\} \cup \{r^2\}$. Similarly, the other subgroups of order four are normal: $\{e, r^2, a, ar^2\}$, and $\{e, r^2, ar, ar^3\}$.
- The subgroup D_4 of D_4 is normal because it is the union of all the conjugacy classes.

Hence the normal subgroups are: $\{e\}$, $\{e, r^2\}$, $\{e, r, r^2, r^3\}$, $\{e, r^2, ar^2, ar^3\}$, $\{e, r^2, ar, ar^3\}$, D_4. There are a total of 6 normal subgroups.

(e) Finally, we are asked whether D_4 is the non-trivial direct product of its subgroups (note it is trivially the direct product $D_4 = D_4 \times \{e\}$). To answer this, note that the order of a direct product group $H \times K$, where H and K are finite groups, is given by $|H| \cdot |K|$, hence if $D_4 = H \times K$ with H and K non-trivial, we must have $|H| = 2$ and $|K| = 4$ (without loss of generality).

Up to isomorphism, there is only one group of order 2, namely $\mathbb{Z}_2$. There are also only two groups of order 4, namely $\mathbb{Z}_4$ and $\mathbb{Z}_2 \times \mathbb{Z}_2$. Notice that all of these groups are Abelian however, so if D_4 can be expressed non-trivially as $D_4 = H \times K$, then the groups H and K in the direct product must both be Abelian. But the direct product of Abelian groups is Abelian, and D_4 is non-Abelian, hence we have a contradiction. It follows D_4 cannot be expressed in this way.

[2] This is sometimes given as the definition of normal subgroups.

2. $O(n)$ consists of $n \times n$ real matrices M satisfying $M^T M = I$, whereas $U(n)$ consists of $n \times n$ complex matrices U satisfying $U^\dagger U = I$. Check that $O(n)$ is a group. Check similarly that $U(n)$ is a group. Verify that the subset of all real matrices in $U(n)$ forms the group $O(n)$ and, similarly, that the subset of all real matrices in $SU(n)$ forms the group $SO(n)$. By considering the action of $U(n)$ on $\mathbb{C}^n$ and identifying $\mathbb{C}^n$ with $\mathbb{R}^{2n}$, show that $U(n)$ is a subgroup of $O(2n)$.

Solution: For the first part of the question, we are asked to demonstrate that both $O(n)$ and $U(n)$ are groups. To do so, we must check each of the group axioms in turn:

(G0) <u>Closure.</u> Suppose that $M, N \in O(n)$. Then we have $(MN)^T(MN) = N^T M^T M N = N^T N = I$, so $MN \in O(n)$. Thus the set of orthogonal matrices is closed under matrix multiplication.

Similarly, suppose that $U, V \in U(n)$. Then we have $(UV)^\dagger UV = V^\dagger U^\dagger UV = V^\dagger V = I$, so $UV \in U(n)$. Thus the set of unitary matrices is closed under matrix multiplication.

(G1) <u>Associativity.</u> We assert that matrix multiplication is associative (there is an easy proof using index notation[3]).

(G2) <u>Identity.</u> Let I denote the $n \times n$ identity matrix. Noting that $I^T I = I$, we see that $I \in O(n)$. Furthermore, for any matrix $M \in O(n)$ we have $IM = MI = M$, so that I is an identity for matrix multiplication on the set of orthogonal matrices.

Similarly, we have $I^\dagger I = I$, so that $I \in U(n)$. Again, for any matrix $U \in U(n)$, we have $IU = UI = U$, so I is also an identity for matrix multiplication on the set of unitary matrices.

(G3) <u>Inverses.</u> Suppose that $M \in O(n)$. By definition, $M^T M = I$, so the matrix M certainly has a left-inverse, given by M^T. Since M, M^T are square matrices, this is sufficient for the existence of a right-inverse, equal to the left inverse.[4] Therefore M is invertible, with $M^{-1} = M^T$. Furthermore, we have $M^{-1} \in O(n)$, since:

$$(M^{-1})^T M^{-1} = (M^T)^{-1} M^{-1} = (M^{-1})^{-1} M^{-1} = M M^{-1} = I,$$

using the standard facts $(A^{-1})^T = (A^T)^{-1}$ and $(A^{-1})^{-1} = A$, for a square matrix A.[5]

Similar considerations apply to the unitary group; just replace the transposes in the above argument with Hermitian conjugates, and exactly the same proof applies.

---

[3] In the most general possible setting, the proof goes as follows: let A, B, C be $p \times q$, $q \times r$ and $r \times s$ matrices respectively. Then by definition of matrix multiplication we have that the (i,j)th entry of the product $(AB)C$ is given by:

$$[(AB)C]_{ij} = \sum_{\alpha=1}^{r} (AB)_{i\alpha} C_{\alpha j} = \sum_{\alpha=1}^{r} \sum_{\beta=1}^{q} A_{i\beta} B_{\beta\alpha} C_{\alpha j} = \sum_{\beta=1}^{q} \sum_{\alpha=1}^{r} A_{i\beta} B_{\beta\alpha} C_{\alpha j} = \sum_{\beta=1}^{q} A_{i\beta} (BC)_{\beta j} = [A(BC)]_{ij}.$$

Thus all components of $(AB)C$ are equal to all components of $A(BC)$, and it follows that $(AB)C = A(BC)$.

[4] A simple proof is as follows. Suppose that A, B are square satisfying $AB = I$. Note that $\mathrm{Ker}(B) = \{\mathbf{0}\}$, else there exists a vector $\mathbf{v} \neq \mathbf{0}$ such that $B\mathbf{v} = \mathbf{0}$, giving $AB\mathbf{v} = \mathbf{0} \neq \mathbf{v} = I\mathbf{v}$. Thus by the rank-nullity theorem, $\mathrm{nullity}(B) = 0, \mathrm{rank}(B) = n$. Now note that:

$$(I - BA)B = B - (BA)B = B - B(AB) = B - BI = 0,$$

using distributivity and associativity of matrix multiplication. It follows that $(I - BA)B = 0$; since B has full rank, this immediately implies that $\mathrm{Ker}(I - BA)$ is the whole space, so that $I - BA$ is the zero matrix. Hence we have $BA = I$, as required.

[5] These facts have elementary proofs. To prove that the inverse of the inverse is the original matrix, note that by definition we have $(A^{-1})^{-1} A^{-1} = I$, so multiplying on the right by A we get $(A^{-1})^{-1} = A$. To prove that the inverse of the transpose is the transpose of the inverse, note that:

$$A^T (A^{-1})^T = (A^{-1} A)^T = I^T = I,$$

so that $(A^T)^{-1} = (A^{-1})^T$. A similar argument applies to the Hermitian conjugate of the inverse.

Next, we are asked to show that $O(n)$ is the subgroup of real matrices in $U(n)$, and $SO(n)$ is the subgroup of real matrices in $SU(n)$.

In the first case, suppose that $A \in U(n)$ is a real matrix. Then $I = A^\dagger A = A^T A$, so $A \in O(n)$. Conversely, suppose that $A \in O(n)$. Then $A^T = A^\dagger$ since A is a real matrix, so it follows that $A^\dagger A = A^T A = I$, and thus A is a purely real matrix in $U(n)$. It follows that the subset of real matrices in $U(n)$ is precisely given by $O(n)$. Since $O(n)$ is a group in its own right, it follows that $O(n)$ is indeed the ***subgroup*** of real matrices in $U(n)$.

A similar argument applies to $SO(n)$ and $SU(n)$, though we should really first check that both the special orthogonal and special unitary groups are indeed groups. One way of quickly doing this is to prove that $SO(n)$ is a subgroup of $O(n)$, and $SU(n)$ is a subgroup of $U(n)$, using the ***subgroup test***:

Proposition: (**The subgroup test**) Let $H \subseteq G$ be a subset of the group G. We have $H \leq G$ if and only if:

(i) H is non-empty, $H \neq \emptyset$;

(ii) for all $g, h \in H$, we have $gh^{-1} \in H$.

Proof: Suppose $H \leq G$. Then $e \in H$, so H is non-empty. For any $h \in H$, we have $h^{-1} \in H$ since H is closed is taking under inverses. Then by closure of the subgroup H, for all $g, h \in H$ we have $gh^{-1} \in H$.

Conversely, suppose (i) and (ii). Since H is non-empty, there exists some $g \in H$, and hence by (ii) we have $gg^{-1} = e \in H$. Thus H contains the identity. Now applying (ii) to the identity and a generic element $g \in H$, we have $eg^{-1} = g^{-1} \in H$, so H is closed under taking inverses. Finally, suppose that $g, h \in H$ are any two elements in H. Then $h^{-1} \in H$, since H is closed under taking inverses, and it follows by (ii) that $g(h^{-1})^{-1} = gh \in H$. Thus H is closed. $\square$

Let's apply this in the case of $SO(n)$. We have $SO(n) \subseteq O(n)$, since $SO(n)$ consists of orthogonal matrices with unit determinant. Furthermore, $SO(n)$ is non-empty since the identity matrix is in $SO(n)$, $I \in SO(n)$. To finish applying the subgroup test, we suppose that $M, N \in SO(n)$. Then $MN^{-1} \in O(n)$ by closure of the orthogonal group, and further:

$$\det(MN^{-1}) = \det(M)\det(N^{-1}) = \frac{\det(M)}{\det(N)} = 1,$$

using a standard property of the determinant. Thus $MN^{-1} \in SO(n)$. It follows $SO(n)$ is a subgroup of $O(n)$ by the subgroup test; in particular $SO(n)$ is a ***group***. The same argument applies to $SU(n)$ and $U(n)$.

We now show that $SO(n)$ is the subgroup of real matrices in $SU(n)$. Suppose that $A \in SO(n)$. Then $A^\dagger A = A^T A = I$ since A is real, so that A is unitary, and further $\det(A) = 1$, so that A is in fact special unitary. Conversely, suppose that $A \in SU(n)$ is a real matrix. Then $I = A^\dagger A = A^T A$, so A is orthogonal, and further $\det(A) = 1$, so that A is in fact special orthogonal. It follows that the set of real matrices in $SU(n)$ is precisely $SO(n)$. Since $SO(n)$ is a group in its own right, it follows that $SO(n)$ is indeed the ***subgroup*** of real matrices in $SU(n)$.

For the final part, we are asked to show that $U(n)$ is a subgroup of $O(2n)$, $U(n) \leq O(2n)$. To do so, we construct an injective group homomorphism $\theta : U(n) \to O(2n)$ by considering the action of $U(n)$ on $\mathbb{C}^n$ and $O(2n)$ on $\mathbb{R}^{2n}$, as suggested by the question.

First, note that any unitary matrix $U \in U(n)$ can be written in the form:

$$U = \mathrm{Re}(U) + i\,\mathrm{Im}(U),$$

where $\mathrm{Re}(U)$ and $\mathrm{Im}(U)$ are unique real $n \times n$ matrices.

The condition that U is unitary implies that $\mathrm{Re}(U), \mathrm{Im}(U)$ satisfy the non-trivial relation

$$I = (\mathrm{Re}(U)+i\,\mathrm{Im}(U))^\dagger(\mathrm{Re}(U)+i\,\mathrm{Im}(U)) = \mathrm{Re}(U)^T\,\mathrm{Re}(U)+\mathrm{Im}(U)^T\,\mathrm{Im}(U)+i(\mathrm{Re}(U)^T\,\mathrm{Im}(U)-\mathrm{Im}(U)^T\,\mathrm{Re}(U)),$$

which can be separated into real and imaginary parts to give:

$$\mathrm{Re}(U)^T\,\mathrm{Re}(U) + \mathrm{Im}(U)^T\,\mathrm{Im}(U) = I, \qquad \mathrm{Re}(U)^T\,\mathrm{Im}(U) - \mathrm{Im}(U)^T\,\mathrm{Re}(U) = 0. \tag{$*$}$$

Similarly, any complex vector $\mathbf{z} \in \mathbb{C}^n$ can be written in the form $\mathbf{z} = \mathbf{x} + i\mathbf{y}$ where $\mathbf{x}$ and $\mathbf{y}$ are real vectors. This means that the action of $U(n)$ on $\mathbb{C}^n$, given by $\mathbf{z} \mapsto \mathbf{z}' = U\mathbf{z}$, can be written in terms of real quantities via:

$$\mathbf{x}' + i\mathbf{y}' = (\mathrm{Re}(U) + i\,\mathrm{Im}(U))(\mathbf{x} + i\mathbf{y}) = \mathrm{Re}(U)\mathbf{x} - \mathrm{Im}(U)\mathbf{y} + i(\mathrm{Re}(U)\mathbf{y} + \mathrm{Im}(U)\mathbf{x}).$$

In particular, we can rewrite this equation in terms of $2n$-dimensional quantities as:

$$\begin{pmatrix}\mathbf{x}'\\ \mathbf{y}'\end{pmatrix} = \begin{pmatrix}\mathrm{Re}(U) & -\mathrm{Im}(U)\\ \mathrm{Im}(U) & \mathrm{Re}(U)\end{pmatrix}\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}.$$

Thus, we have identified the action of $U(n)$ with the action of some $2n\times 2n$ matrix on $\mathbb{R}^{2n}$. We hope to use this identification to define a map $\theta : U(n) \to O(2n)$ via:

$$\theta(U) = \begin{pmatrix}\mathrm{Re}(U) & -\mathrm{Im}(U)\\ \mathrm{Im}(U) & \mathrm{Re}(U)\end{pmatrix}.$$

It remains to show all required properties of this map.

- **The map is well-defined.** We need to check that the codomain of θ is indeed $O(2n)$. To do so, we must check that $\theta(U)$ is an orthogonal matrix, $\theta(U)^T\theta(U) = I$.

  This can be verified by a direct calculation. Note that:

$$\begin{aligned}\theta(U)^T\theta(U) &= \begin{pmatrix}\mathrm{Re}(U)^T & \mathrm{Im}(U)^T\\ -\mathrm{Im}(U)^T & \mathrm{Re}(U)^T\end{pmatrix}\begin{pmatrix}\mathrm{Re}(U) & -\mathrm{Im}(U)\\ \mathrm{Im}(U) & \mathrm{Re}(U)\end{pmatrix}\\ &= \begin{pmatrix}\mathrm{Re}(U)^T\,\mathrm{Re}(U) + \mathrm{Im}(U)^T\,\mathrm{Im}(U) & -\mathrm{Re}(U)^T\,\mathrm{Im}(U) + \mathrm{Im}(U)^T\,\mathrm{Re}(U)\\ -\mathrm{Im}(U)^T\,\mathrm{Re}(U) + \mathrm{Re}(U)^T\,\mathrm{Im}(U) & \mathrm{Im}(U)^T\,\mathrm{Im}(U) + \mathrm{Re}(U)^T\,\mathrm{Re}(U)\end{pmatrix}\\ &= \begin{pmatrix}I & 0\\ 0 & I\end{pmatrix},\end{aligned}$$

  using the relations from ($*$) in the final step. Therefore $\theta(U) \in O(2n)$ as required.

- **The map is a group homomorphism.** To verify this property, it is useful to consider the product UV where $U, V \in U(n)$. We note that this product can be written in terms of the real and imaginary parts of the two unitary matrices via:

$$UV = (\mathrm{Re}(U)+i\,\mathrm{Im}(U))(\mathrm{Re}(V)+i\,\mathrm{Im}(V)) = \mathrm{Re}(U)\,\mathrm{Re}(V)-\mathrm{Im}(U)\,\mathrm{Im}(V)+i(\mathrm{Im}(U)\,\mathrm{Re}(V)+\mathrm{Re}(U)\,\mathrm{Im}(V)),$$

  and hence $\mathrm{Re}(UV) = \mathrm{Re}(U)\,\mathrm{Re}(V) - \mathrm{Im}(U)\,\mathrm{Im}(V)$ and $\mathrm{Im}(UV) = \mathrm{Im}(U)\,\mathrm{Re}(V) + \mathrm{Re}(U)\,\mathrm{Im}(V)$. Applying this to the homomorphism property then, we have:

$$\begin{aligned}\theta(UV) &= \begin{pmatrix}\mathrm{Re}(U)\,\mathrm{Re}(V) - \mathrm{Im}(U)\,\mathrm{Im}(V) & -\mathrm{Im}(U)\,\mathrm{Re}(V) - \mathrm{Re}(U)\,\mathrm{Im}(V)\\ \mathrm{Im}(U)\,\mathrm{Re}(V) + \mathrm{Re}(U)\,\mathrm{Im}(V) & \mathrm{Re}(U)\,\mathrm{Re}(V) - \mathrm{Im}(U)\,\mathrm{Im}(V)\end{pmatrix}\\ &= \begin{pmatrix}\mathrm{Re}(U) & -\mathrm{Im}(U)\\ \mathrm{Im}(U) & \mathrm{Re}(U)\end{pmatrix}\begin{pmatrix}\mathrm{Re}(V) & -\mathrm{Im}(V)\\ \mathrm{Im}(V) & \mathrm{Re}(V)\end{pmatrix}\\ &= \theta(U)\theta(V),\end{aligned}$$

  as required.

- **The map is injective.** This is straightforward to see: if $\theta(U) = \theta(V)$, we must have:

$$\begin{pmatrix} \mathrm{Re}(U) & -\mathrm{Im}(U) \\ \mathrm{Im}(U) & \mathrm{Re}(U) \end{pmatrix} = \begin{pmatrix} \mathrm{Re}(V) & -\mathrm{Im}(V) \\ \mathrm{Im}(V) & \mathrm{Re}(V) \end{pmatrix},$$

  which implies $\mathrm{Re}(U) = \mathrm{Re}(V), \mathrm{Im}(U) = \mathrm{Im}(V)$, and thus $U = V$.

Therefore the map we have constructed above, $\theta : U(n) \to O(2n)$ is an injective group homomorphism, embedding $U(n)$ into $O(2n)$ as a subgroup as required.

❊ **Comments:** In fact, it is possible to show a stronger result: $U(n)$ is a subgroup of $SO(2n)$, rather than just $O(2n)$. To prove this, all we need to show is that the determinant of $\theta(U)$ is 1 (where $\theta : U(n) \to O(2n)$ is the injective group homomorphism we defined above). It is rather difficult to compute the determinant of $\theta(U)$ directly.[6] Instead, we can use an argument based on continuity and connectedness.

Note that θ is clearly a continuous function of U, and recall that the determinant is also a continuous function; therefore, the composition $\det \circ\, \theta : U(n) \to \mathbb{R}$ is a continuous function. But since the determinant of any orthogonal matrix is ± 1, this function must in fact restrict to $\det \circ\, \theta : U(n) \to \{-1, 1\}$.

Notice now that $\{-1, 1\}$ is disconnected, with connected components $\{-1\}$ and $\{+1\}$. But continuous functions preserve connectedness of spaces, so if we can show that $U(n)$ is connected, then the map must further restrict to either $\det \circ\, \theta : U(n) \to \{-1\}$ or $\det \circ\, \theta : U(n) \to \{+1\}$. In fact, the proof of connectedness of $U(n)$ is straightforward; we give it in the following Lemma.

Lemma: $U(n)$ is path-connected (and hence connected).

Proof: Begin by recalling that the eigenvalues of a unitary matrix $U \in U(n)$ are all phases; to see this, note that $U\mathbf{v} = \lambda \mathbf{v}$ for $\mathbf{v} \neq \mathbf{0}$ implies $|\lambda|^2 ||\mathbf{v}||^2 = (\lambda^* \mathbf{v}^\dagger)(\lambda \mathbf{v}) = \mathbf{v}^\dagger U^\dagger U \mathbf{v} = \mathbf{v}^\dagger \mathbf{v} = ||\mathbf{v}||^2$, hence $|\lambda| = 1$. Furthermore, from elementary linear algebra, we recall that any unitary matrix can be diagonalised via a unitary transformation. In particular, for any fixed $U \in U(n)$, we can write:

$$U = P \mathrm{diag}\{e^{i\theta_1}, ..., e^{i\theta_n}\} P^\dagger.$$

Let's now define $U(t) = P\mathrm{diag}\{e^{it\theta_1}, ..., e^{it\theta_n}\}P^\dagger$. We note that: (i) $U(0) = I$; (ii) for any $t \in [0, 1]$, the matrix $U(t)$ is unitary, since:

$$U(t)^\dagger U(t) = P\mathrm{diag}\{e^{-it\theta_1}, ..., e^{-it\theta_n}\}P^\dagger P\mathrm{diag}\{e^{it\theta_1}, ..., e^{it\theta_n}\}P^\dagger = I.$$

It follows that $\{U(t) : 0 \leq t \leq 1\}$ defines a continuous path in $U(n)$ from I to U. To connect any two points $U_1, U_2 \in U(n)$ via a path then, we just concatenate two paths U_1 to I, then I to U_2; it follows that $U(n)$ is path-connected. □

Hence we can conclude that we must have $\det \circ\, \theta \equiv 1$ or $\det \circ\, \theta \equiv -1$ identically. To see which is correct, note:

$$\det(\theta(I)) = \det \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} = 1,$$

thus $\det(\theta(U)) = 1$ for all $U \in U(n)$, and we're done.

[6] Note the standard formula for a 2×2 matrix does *not* extend in general to block matrices. If we attempt to apply it here, we get a determinant of $\mathrm{Re}(U)^2 + \mathrm{Im}(U)^2$, which cannot be further simplified using $(*)$.

3. Show that for matrices $M \in O(n)$, the first column of M is an arbitrary unit vector, the second is a unit vector orthogonal to the first, ..., the kth column is a unit vector orthogonal to the span of the previous ones, etc. Deduce the dimension of $O(n)$. By similar reasoning, determine the dimension of $U(n)$.

➼ **Solution:** Write the matrix M in terms of its columns $\mathbf{m}_1, \mathbf{m}_2, ..., \mathbf{m}_n$ as:

$$M = \begin{pmatrix} \mathbf{m}_1 & \mathbf{m}_2 & \cdots & \mathbf{m}_n \end{pmatrix}.$$

The condition that M is an orthogonal matrix can then be written as:

$$I = M^T M = \begin{pmatrix} \mathbf{m}_1^T \\ \mathbf{m}_2^2 \\ \vdots \\ \mathbf{m}_n^T \end{pmatrix} \begin{pmatrix} \mathbf{m}_1 & \mathbf{m}_2 & \cdots & \mathbf{m}_n \end{pmatrix} = \begin{pmatrix} \mathbf{m}_1^T \mathbf{m}_1 & \mathbf{m}_1^T \mathbf{m}_2 & \cdots & \mathbf{m}_1^T \mathbf{m}_n \\ \mathbf{m}_2^T \mathbf{m}_1 & \mathbf{m}_2^T \mathbf{m}_2 & \cdots & \mathbf{m}_2^T \mathbf{m}_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{m}_n^T \mathbf{m}_1 & \mathbf{m}_n^T \mathbf{m}_2 & \cdots & \mathbf{m}_n^T \mathbf{m}_n \end{pmatrix}.$$

Note that the matrix on the right hand side is symmetric, since $\mathbf{v}^T\mathbf{w} = \mathbf{w}^T\mathbf{v}$ for all $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$. In particular, if we choose $\mathbf{m}_1, ..., \mathbf{m}_n$ such that the upper triangular region of the matrix on the right hand side agrees with the upper triangular region of the identity on the left hand side, we will satisfy all necessary constraints.

We work from left to right across the matrix, ignoring all entries below the main diagonal. The only remaining constraint from the first column is $\mathbf{m}_1^T\mathbf{m}_1 = 1$, so let us choose $\mathbf{m}_1$ to be an arbitrary unit vector. In the second column, the remaining constraints are $\mathbf{m}_1^T\mathbf{m}_2 = 0$ and $\mathbf{m}_2^T\mathbf{m}_2 = 1$, so if $\mathbf{m}_1$ is already chosen, we must choose $\mathbf{m}_2$ to be an arbitrary unit vector orthogonal to $\mathbf{m}_1^T$.

Continuing in this fashion, suppose that the first $k-1$ vectors have already been chosen. The constraints on the kth vector $\mathbf{m}_k$ come from the kth column, and are given by:

$$\mathbf{m}_1^T\mathbf{m}_k = 0, \qquad \mathbf{m}_2^T\mathbf{m}_k = 0, \qquad ..., \qquad \mathbf{m}_{k-1}^T\mathbf{m}_k = 0, \qquad \mathbf{m}_k^T\mathbf{m}_k = 0.$$

Thus we see that $\mathbf{m}_k$ may be chosen to be an arbitrary unit vector orthogonal to $\mathbf{m}_1, ..., \mathbf{m}_{k-1}$; this is equivalent to $\mathbf{m}_k$ being orthogonal to the span of $\mathbf{m}_1, ..., \mathbf{m}_{k-1}$. The required result thus follows by induction.

To deduce the dimension, we count the degrees of freedom we had available to us in constructing the matrix M. The first column has 1 constraint coming from the requirement that the first column is a unit vector. The second column has 2 constraints coming from the requirement that the second column is a unit vector orthogonal to the first. Continuing in this fashion, we see that the kth column has k constraints, coming from the requirement that the kth column is a unit vector, and that it is orthogonal to each of the $k-1$ previous columns. Subtracting the number of constraints from the n^2 available entries in the matrix, we see that the dimension of the orthogonal group is:

$$n^2 - \sum_{k=1}^{n} k = n^2 - \frac{1}{2}n(n+1) = \frac{1}{2}n(n-1).$$

We can perform a similar analysis with the unitary group. The only change is that the transposes become Hermitian conjugates, and a single *complex* condition should now be interpreted as two *real* conditions (on the real and imaginary part of a complex number).

Let's begin. Let U be a unitary matrix, and let $\mathbf{u}_1, ..., \mathbf{u}_n$ be its columns so that:

$$U = \left(\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\right).$$

By the same argument as above, we see that if we choose $\mathbf{u}_1$ first, the only constraint on this vector is $\mathbf{u}_1^\dagger \mathbf{u}_1 = 1$. Whilst this constrains a *complex vector*, it is in fact a *real* constraint, since $\mathrm{Re}(\mathbf{u}_1^\dagger \mathbf{u}_1) = \mathbf{u}_1^\dagger \mathbf{u}_1$. In particular, it counts as a single real constraint on the first column $\mathbf{u}_1$.

Once we have fixed $\mathbf{u}_1$, the constraints on $\mathbf{u}_2$ are given by $\mathbf{u}_1^\dagger \mathbf{u}_2 = 0$ and $\mathbf{u}_2^\dagger \mathbf{u}_2 = 1$. Since $\mathbf{u}_1^\dagger \mathbf{u}_2$ could be a complex number, the first constraint is indeed a fully-fledged complex constraint, thus counts as two real constraints. The second constraint simply counts as one real constraint as before. Thus there are 3 real constraints on the second column $\mathbf{u}_2$.

Moving up to the kth column, the constraints are $\mathbf{u}_1^\dagger \mathbf{u}_k = 0, ..., \mathbf{u}_{k-1}^\dagger \mathbf{u}_k = 0$ and $\mathbf{u}_k^\dagger \mathbf{u}_k = 1$. Hence there are $2(k-1)+1 = 2k-1$ real constraints on the kth column (the first $k-1$ constraints are complex constraints, so count as two real constraints as before). Note that the conditions $\mathbf{u}_1^\dagger \mathbf{u}_k = 0, ..., \mathbf{u}_{k-1}^\dagger \mathbf{u}_k = 0$ are equivalent to the kth column of U being orthogonal to the complex span of the first $k-1$ columns.

Let's count everything up. There are n^2 entries of the matrix U, which are described by $2n^2$ real numbers when we consider their real and imaginary parts. Therefore subtracting the total number of constraints, we see that the dimension of $U(n)$ is given by:

$$2n^2 - \sum_{k=1}^{n}(2k-1) = 2n^2 - 2\left(\sum_{k=1}^{n} k\right) + n = 2n^2 - n(n+1) + n = n^2.$$

4. (a) Show that a general element of $SU(2)$ can be written as:

$$U = \begin{pmatrix} \alpha & \beta \\ -\beta^* & \alpha^* \end{pmatrix},$$

where α, β are complex numbers satisfying $|\alpha|^2 + |\beta|^2 = 1$.

(b) Deduce that an alternative form for an $SU(2)$ matrix is:

$$U = a_0 I + i\mathbf{a} \cdot \boldsymbol{\sigma},$$

with $(a_0, \mathbf{a})$ real, $\boldsymbol{\sigma}$ the Pauli matrices, and $a_0^2 + |\mathbf{a}|^2 = 1$.

(c) Using the second form, calculate the product of two $SU(2)$ matrices.

➾ **Solution:** (a) Recall that a 2×2 complex matrix U is special unitary if and only if $U^\dagger U = I$ and $\det(U) = 1$. Since special unitary matrices are invertible, we can rewrite these conditions as:

$$U^\dagger = U^{-1}, \qquad \det(U) = 1. \tag{$*$}$$

Now let's write:

$$U = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$$

for complex numbers $\alpha, \beta, \gamma, \delta \in \mathbb{C}$. Using the standard formula for the inverse of a 2×2 matrix, together with the condition $\det(U) = 1$, we see that the equations ($*$) can be rewritten as:

$$\begin{pmatrix} \alpha^* & \gamma^* \\ \beta^* & \delta^* \end{pmatrix} = \begin{pmatrix} \delta & -\beta \\ -\gamma & \alpha \end{pmatrix}, \qquad \alpha\delta - \beta\gamma = 1.$$

Thus we see that the matrix U is special unitary if and only if $\gamma = -\beta^*$, $\delta = \alpha^*$, and $|\alpha|^2 + |\beta|^2 = 1$, which yields:

$$U = \begin{pmatrix} \alpha & \beta \\ -\beta^* & \alpha^* \end{pmatrix}, \qquad |\alpha|^2 + |\beta|^2 = 1,$$

as required.

(b) For the second part, take $\alpha = a_0 + ia_3$ and $\beta = a_2 + ia_1$. Then our above parametrisation takes the form:

$$U = \begin{pmatrix} a_0 + ia_3 & a_2 + ia_1 \\ -a_2 + ia_1 & a_0 - ia_3 \end{pmatrix} = a_0 I + ia_1 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + ia_2 \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} + ia_3 \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = a_0 I + i\mathbf{a} \cdot \boldsymbol{\sigma},$$

with $1 = |\alpha|^2 + |\beta|^2 = a_0^2 + |\mathbf{a}|^2$, as required.

(c) Finally, we are asked to compute the product of two $SU(2)$ matrices in this parametrisation. Take $U = a_0 I + i\mathbf{a} \cdot \boldsymbol{\sigma}$ and $V = b_0 I + i\mathbf{b} \cdot \boldsymbol{\sigma}$, with $a_0^2 + |\mathbf{a}|^2 = b_0^2 + |\mathbf{b}|^2 = 1$. The product is then given by:

$$UV = (a_0 I + i\mathbf{a} \cdot \boldsymbol{\sigma})(b_0 I + i\mathbf{b} \cdot \boldsymbol{\sigma}) = a_0 b_0 I + i(a_0 \mathbf{b} + b_0 \mathbf{a}) \cdot \boldsymbol{\sigma} - (\mathbf{a} \cdot \boldsymbol{\sigma})(\mathbf{b} \cdot \boldsymbol{\sigma}).$$

To manipulate this into the form of an $SU(2)$ matrix given in (b), we need to evaluate $(\mathbf{a} \cdot \boldsymbol{\sigma})(\mathbf{b} \cdot \boldsymbol{\sigma})$. Using the summation convention, and the identity proved in 7(a), we have:

$$(\mathbf{a} \cdot \boldsymbol{\sigma})(\mathbf{b} \cdot \boldsymbol{\sigma}) = a_i \sigma_i b_j \sigma_j = a_i b_j \left(\delta_{ij} I + i\epsilon_{ijk}\sigma_k\right) = (\mathbf{a} \cdot \mathbf{b}) I + i(\mathbf{a} \times \mathbf{b}) \cdot \boldsymbol{\sigma}.$$

Putting everything together then, we have:

$$UV = (a_0 b_0 - \mathbf{a} \cdot \mathbf{b}) I + i(a_0 \mathbf{b} + b_0 \mathbf{a} - \mathbf{a} \times \mathbf{b}) \cdot \boldsymbol{\sigma}.$$

5. A Lie group has group elements $g(\mathbf{x})$ depending on group parameters x^r, with $g(\mathbf{0}) = e$, the identity, and under group multiplication $g(\mathbf{x})g(\mathbf{y}) = g(\phi(\mathbf{x}, \mathbf{y}))$ for some $\phi^r(\mathbf{x}, \mathbf{y})$. Let $g(\mathbf{x})^{-1} = g(\bar{\mathbf{x}})$ where $\phi(\bar{\mathbf{x}}, \mathbf{x}) = 0$.

(a) Why must $\phi^r(\mathbf{x}, \mathbf{0}) = x^r$, $\phi^r(\mathbf{0}, \mathbf{y}) = y^r$?

(b) Assume $\phi^r(\mathbf{x}, \mathbf{y})$ is expanded near the origin according to:

$$\phi^a(\mathbf{x}, \mathbf{y}) = x^a + y^a + c^a{}_{bc} x^b y^c + O(x^2 y, xy^2).$$

Use this to find $\bar{\mathbf{x}}(\mathbf{x})$ for $\mathbf{x}$ small.

(c) Let $g(\mathbf{d}) = g(\mathbf{x})^{-1} g(\mathbf{y})^{-1} g(\mathbf{x}) g(\mathbf{y})$ and show that for $\mathbf{x}, \mathbf{y}$ small $d^a = f^a{}_{bc} x^b y^c$, where $f^a{}_{bc} = c^a{}_{bc} - c^a{}_{cb}$.

(d) Using an expansion to one higher order, show that the associativity condition $\phi(\phi(\mathbf{x}, \mathbf{y}), \mathbf{z}) = \phi(\mathbf{x}, \phi(\mathbf{y}, \mathbf{z}))$ leads to the Jacobi identity.

(e) Assume the Lie algebra has generators T_a satisfying $[T_a, T_b] = f^a{}_{bc} T_c$. For an element of the Lie algebra $a^a T_a$ there is an associated group element given by $g(\mathbf{a}) = \exp\left(a^a T_a\right)$. Use the Baker-Campbell-Hausdorff formula:

$$\exp(tA)\exp(tB) = \exp\left(tA + tB + \frac{t^2}{2}[A, B] + O(t^3)\right)$$

to obtain $\phi(\mathbf{x}, \mathbf{y})$ for small $\mathbf{x}, \mathbf{y}$ and verify that this is compatible with the general expansion of ϕ.

Solution: This question is all about the structure of *Lie groups*, so it's worth reminding you of the definition:

Definition: A *Lie group* is a smooth manifold G which is also a group, such that the group operations $\cdot : G \times G \to G$ and $(-)^{-1} : G \to G$ are smooth maps between manifolds. The dimension n of the manifold is called the *dimension of the Lie group*, and is written $n = \dim(G)$.

In any coordinate patch $U \subseteq G$ on the Lie group with corresponding chart $\psi : U \to V \subseteq \mathbb{R}^n$, $\psi = (x^1, ..., x^n)$, we can parametrise the group elements $g \in U$ in terms of the local coordinates x^i as $g(x^1, ..., x^n) = g(\mathbf{x})$. For two group elements $g(\mathbf{x}), g(\mathbf{y})$ in this patch U, if their product $g(\mathbf{x})g(\mathbf{y})$ is also in this patch can write:

$$g(\mathbf{x})g(\mathbf{y}) = g(\phi(\mathbf{x}, \mathbf{y}))$$

where $\phi : V \times V \to V$ is a smooth map of the local coordinates (this is a consequence of the group multiplication $\cdot : G \times G \to G$ being a smooth map between manifolds).

In this course, we will always assume that we are working with a *global* coordinate patch $U = G$, which means that multiplication is defined *everywhere* by $g(\mathbf{x})g(\mathbf{y}) = g(\phi(\mathbf{x}, \mathbf{y}))$ for some smooth map $\phi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$. This means that we can avoid a lot of the mathematical subtleties that come from the fact that coordinates on a manifold are typically only locally defined.

In summary then, a 'Lie group' in this course is really a specification of a group G, some global coordinates on the group, and the map $\phi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$.

(a) We now begin the question proper. The identity axiom for a group tells us that:

$$g(\phi(\mathbf{0},\mathbf{x})) = g(\mathbf{0})g(\mathbf{x}) = g(\mathbf{x}) = g(\mathbf{x})g(\mathbf{0}) = g(\phi(\mathbf{x},\mathbf{0})),$$

and hence, reading off the coordinates, we have $\phi^r(\mathbf{0},\mathbf{x}) = \phi^r(\mathbf{x},\mathbf{0}) = x^r$ as required.

(b) Recall that we define $\bar{\mathbf{x}}$ to be the coordinates of the group element $g(\bar{\mathbf{x}}) = g(\mathbf{x})^{-1}$ (so in particular $\bar{\mathbf{x}}$ can be viewed as a function of $\mathbf{x}$, say $\bar{\mathbf{x}}(\mathbf{x})$). By the inverses axiom for a group, we have:

$$g(\phi(\bar{\mathbf{x}},\mathbf{x})) = g(\bar{\mathbf{x}})g(\mathbf{x}) = g(\mathbf{0}) = g(\mathbf{x})g(\bar{\mathbf{x}}) = g(\phi(\mathbf{x},\bar{\mathbf{x}})),$$

which implies $\phi(\bar{\mathbf{x}},\mathbf{x}) = \phi(\mathbf{x},\bar{\mathbf{x}}) = \mathbf{0}$, as given in the question.

Now consider the given expansion:

$$\phi^a(\mathbf{x},\mathbf{y}) = x^a + y^a + c^a{}_{bc}x^b y^c + O(x^2y, xy^2).$$

Let us set $\mathbf{y} = \bar{\mathbf{x}}$, and expand $\bar{\mathbf{y}}$ in terms of $\mathbf{x}$ as $\bar{\mathbf{x}}^r = \alpha^r + \beta^r{}_s x^s + \gamma^r{}_{st}x^s x^t + O(x^3)$, where γ is symmetric on its downstairs indices. Then we have:

$$0 = x^a + \alpha^a + \beta^a{}_b x^b + \gamma^a{}_{bc}x^b x^c + c^a{}_{bc}x^b\left(\alpha^c + \beta^c{}_s x^s\right) + O(x^3, x^2\bar{x}, x\bar{x}^2).$$

Comparing coefficients of $\mathbf{x}$, we see that $\alpha = 0, \beta^r{}_s = -\delta^r{}_s, \gamma^r{}_{(st)} = c^r{}_{(st)}$. Hence we can write (any antisymmetric part in $c^r{}_{st}$ cancels out):

$$\bar{x}^r(\mathbf{x}) = -x^r + c^r{}_{st}x^s x^t + O(x^3).$$

(c) In this part of the question, we are asked to look at the group commutator $g(\mathbf{d}) = [g(\mathbf{x}), g(\mathbf{y})] = g(\mathbf{x})^{-1}g(\mathbf{y})^{-1}g(\mathbf{x})g(\mathbf{y})$. We would like to find $\mathbf{d}$ in terms of the coordinates $\mathbf{x}$, $\mathbf{y}$. This amounts to combining the coordinates of the group elements on the right hand side using the expansion of the ϕ function. We have:

$$g(\mathbf{x})^{-1}g(\mathbf{y})^{-1}g(\mathbf{x})g(\mathbf{y}) = g(\phi(\bar{\mathbf{x}},\bar{\mathbf{y}}))g(\phi(\mathbf{x},\mathbf{y})) = g(\phi(\phi(\bar{\mathbf{x}},\bar{\mathbf{y}}),\phi(\mathbf{x},\mathbf{y}))).$$

Hence the coordinates $\mathbf{d}$ are given by (using the formula for the inverse $\bar{\mathbf{x}}$ we worked out in part (b)):

$$\begin{aligned}
\mathbf{d} &= \phi(\phi(\bar{\mathbf{x}},\bar{\mathbf{y}}),\phi(\mathbf{x},\mathbf{y})) \\
&= \phi\left(\bar{x}^r + \bar{y}^r + c^r{}_{st}\bar{x}^s\bar{x}^t + O(\bar{x}^2\bar{y}, \bar{x}\bar{y}^2), x^r + y^r + c^r{}_{st}x^s y^t + O(x^2y, xy^2)\right) \\
&= \bar{x}^r + \bar{y}^r + x^r + y^r + c^r{}_{st}\bar{x}^s\bar{y}^t + c^r{}_{st}x^s y^t + c^r{}_{st}(\bar{x}^r + \bar{y}^r)(x^r + y^r) + O(x^2y, xy^2) \\
&= c^r{}_{st}x^s x^t + c^r{}_{st}y^s y^t + 2c^r{}_{st}x^s y^t - c^r{}_{st}(x^s + y^s)(x^t + y^t) + O(x^3, y^3, x^2y, xy^2) \\
&= \left(c^r{}_{st} - c^r{}_{ts}\right)x^s y^t + O(x^3, y^3, x^2y, xy^2),
\end{aligned}$$

hence to leading order we have $d^r = f^r{}_{st}x^s y^t$, where $f^r{}_{st} = c^r{}_{st} - c^r{}_{ts}$, as required.

(d) As stated in the question, the associativity axiom for a group can be stated in terms of the ϕ function as:

$$\phi(\phi(\mathbf{x},\mathbf{y}),\mathbf{z}) = \phi(\mathbf{x},\phi(\mathbf{y},\mathbf{z})). \tag{$*$}$$

We are asked to expand this to a sufficiently high order in order to prove the Jacobi identity; it turns out that we will need to consider a higher order expansion than the one we previously used for ϕ. We set:

$$\phi^r(\mathbf{x},\mathbf{y}) = x^r + y^r + c^r{}_{st}x^sy^t + d^r{}_{stu}x^sy^ty^u + e^r{}_{stu}x^sx^ty^u + O(xy^3, x^2y^2, x^3y).$$

Let's write $O(4)$ for terms that are fourth order in x, y, z to save on notation. Then the left hand side of ($*$) becomes:

$$\begin{aligned}
\phi^r(\phi(\mathbf{x},\mathbf{y}),\mathbf{z}) &= \phi^r(\mathbf{x},\mathbf{y}) + z^r + c^r{}_{st}\phi^s(\mathbf{x},\mathbf{y})z^r + d^r{}_{stu}\phi^s(\mathbf{x},\mathbf{y})z^tz^u + e^r{}_{stu}\phi^s(\mathbf{x},\mathbf{y})\phi^t(\mathbf{x},\mathbf{y})z^u + O(4)\\
&= x^r + y^r + z^r + c^r{}_{st}x^sy^t + d^r{}_{stu}x^sy^ty^u + e^r{}_{stu}x^sx^ty^u + c^r{}_{st}\left(x^s + y^s + c^s{}_{uv}x^uy^v\right)z^t\\
&\quad + d^r{}_{stu}\left(x^s + y^s\right)z^tz^u + e^r{}_{stu}\left(x^s + y^s\right)\left(x^t + y^t\right)z^u + O(4)\\
&= x^r + y^r + z^r + c^r{}_{st}\left(x^sy^t + x^sz^t + y^sz^t\right) + c^r{}_{st}c^s{}_{uv}x^uy^vz^t\\
&\quad + d^r{}_{stu}\left(x^sy^ty^u + x^sz^tz^u + y^sz^tz^u\right) + e^r{}_{stu}\left(x^sx^ty^u + x^sx^tz^u + x^sy^tz^u + y^sx^tz^u + y^sy^tz^u\right) + O(4).
\end{aligned}$$

On the right hand side of ($*$), we have:

$$\begin{aligned}
\phi^r(\mathbf{x},\phi(\mathbf{y},\mathbf{z})) &= x^r + \phi^r(\mathbf{y},\mathbf{z}) + c^r{}_{st}x^s\phi^t(\mathbf{y},\mathbf{z}) + d^r{}_{stu}x^s\phi^t(\mathbf{y},\mathbf{z})\phi^u(\mathbf{y},\mathbf{z}) + e^r{}_{stu}x^sx^t\phi^u(\mathbf{y},\mathbf{z}) + O(4)\\
&= x^r + y^r + z^r + c^r{}_{st}y^sz^t + d^r{}_{stu}y^sz^tz^u + e^r{}_{stu}y^sy^tz^u + c^r{}_{st}x^s\left(y^t + z^t + c^t{}_{uv}y^uz^v\right)\\
&\quad + d^r{}_{stu}x^s\left(y^t + z^t\right)\left(y^u + z^u\right) + e^r{}_{stu}x^sx^t\left(y^u + z^u\right) + O(4)\\
&= x^r + y^r + z^r + c^r{}_{st}\left(y^sz^t + x^sy^t + x^sz^t\right) + c^r{}_{st}c^t{}_{uv}x^sy^uz^v\\
&\quad + d^r{}_{stu}\left(y^sz^tz^u + x^sy^ty^u + x^sy^tz^u + x^sz^ty^u + x^sz^tz^u\right) + e^r{}_{stu}\left(y^sy^tz^u + x^sx^ty^u + x^sx^tz^u\right) + O(4).
\end{aligned}$$

Comparing the coefficients of the third order terms on both sides, we have:

$$c^r{}_{st}c^s{}_{uv} + e^r{}_{uvt} + e^r{}_{vut} = c^r{}_{us}c^s{}_{vt} + d^r{}_{uvt} + d^r{}_{utv},$$

and hence on rearrangement we have:

$$c^r{}_{st}c^s{}_{uv} - c^r{}_{us}c^s{}_{vt} = d^r{}_{uvt} + d^r{}_{utv} - e^r{}_{uvt} - e^r{}_{vut}.$$

Now, the type of term we're aiming for is $f^r{}_{st}f^s{}_{uv} = (c^r{}_{st} - c^r{}_{ts})(c^s{}_{uv} - c^s{}_{vu}) = c^r{}_{st}c^s{}_{uv} - c^r{}_{st}c^s{}_{vu} - c^r{}_{ts}c^s{}_{uv} + c^r{}_{ts}c^s{}_{vu}$. Hence consider the sum of six different equations of the above form:

$$\begin{aligned}
c^r{}_{st}c^s{}_{uv} - c^r{}_{us}c^s{}_{vt} &= d^r{}_{uvt} + d^r{}_{utv} - e^r{}_{uvt} - e^r{}_{vut},\\
-c^r{}_{st}c^s{}_{vu} + c^r{}_{vs}c^s{}_{ut} &= -d^r{}_{vut} - d^r{}_{vtu} + e^r{}_{vut} + e^r{}_{uvt},\\
c^r{}_{sv}c^s{}_{tu} - c^r{}_{ts}c^s{}_{uv} &= d^r{}_{tuv} + d^r{}_{tvu} - e^r{}_{tuv} - e^r{}_{utv},\\
-c^r{}_{su}c^s{}_{tv} + c^r{}_{ts}c^s{}_{vu} &= -d^r{}_{tvu} - d^r{}_{tuv} + e^r{}_{tvu} + e^r{}_{vtu},\\
c^r{}_{su}c^s{}_{vt} - c^r{}_{vs}c^s{}_{tu} &= d^r{}_{vtu} + d^r{}_{vut} - e^r{}_{vtu} - e^r{}_{tvu},\\
-c^r{}_{sv}c^s{}_{ut} + c^r{}_{us}c^s{}_{tv} &= -d^r{}_{utv} - d^r{}_{uvt} + e^r{}_{utv} + e^r{}_{tuv}.
\end{aligned}$$

Adding these six equations gives us $f^r{}_{st}f^s{}_{uv} + f^r{}_{su}f^s{}_{vt} + f^r{}_{sv}f^s{}_{tu} = 0$, which is the Jacobi identity as required.

(e) In the final part of the question, we are asked to think about the structure of the *Lie algebra* $\mathfrak{g}$ of a Lie group G. In the following question we will recall carefully the construction of the Lie algebra of a Lie group; the main idea is that the Lie algebra 'linearises' the Lie group G at the identity, giving the Lie algebra $\mathfrak{g}$ the simple structure of a vector space. For now, we assume that we have already constructed this vector space, with generators $\{T_r\}$.

If $\{T_r\}$ is a basis for the Lie algebra, then any element of the Lie algebra can be expressed as $a^r T_r$ for some coordinates a^r. Recall from lectures that *exponentiation* allows us to turn this 'infinitesimal transformation' into a 'finite transformation', hence $g(\mathbf{a}) = \exp(a^r T_r)$ corresponds to some element in the Lie group with coordinates a^r. We know from the Lie group structure that in general we can write:

$$g(\mathbf{a})g(\mathbf{b}) = g(\phi(\mathbf{a}, \mathbf{b})),$$

which translates into a statement about exponentials given by:

$$\exp(a^r T_r)\exp(b^r T_r) = \exp\left(\phi^r(\mathbf{a}, \mathbf{b})T_r\right).$$

The question suggests that we can compute $\phi(\mathbf{a}, \mathbf{b})$ by combining the exponentials on the left hand side using the *Baker-Campbell-Hausdorff identity* which we will verify to small order as part of Question 8. We have:

$$\exp(a^r T_r)\exp(b^r T_r) = \exp\left(a^r T_r + b^r T_r + \frac{1}{2}a^r b^s [T_r, T_s] + \cdots\right) = \exp\left(\left(a^r + b^r + \frac{1}{2}f^r{}_{ab}\, a^a b^b + \cdots\right)T_r\right),$$

which implies, to this order, that we have:

$$\phi^r(\mathbf{a}, \mathbf{b}) = a^r + b^r + \frac{1}{2}f^r{}_{ab}\, a^a b^b + \cdots,$$

which is compatible with the more general $c^a{}_{bc}$ expression if we set $c^a{}_{bc} = \frac{1}{2}f^a{}_{bc}$.

A natural question that one might ask when we get this result is: *why* did we find that $c^a{}_{bc} = \frac{1}{2}f^a{}_{bc}$? The idea is that working in these coordinates a^r on the Lie algebra, such that elements are of the form $a^r T_r$, gives a *privileged, simplified* choice of coordinates when we exponentiate to recover the Lie group (so that elements are of the form $\exp(a^r T_r)$); in particular, any symmetric part of $c^a{}_{bc}$ vanishes if we choose to parametrise elements of the Lie group using such coordinates on the Lie algebra, as we have just shown. At a high level, the reason this happens is that writing elements of the Lie algebra as $a^r T_r$ respects the vector space structure of the Lie algebra (i.e. in these coordinates, the point with coordinates $a^r + b^r$ is given by $(a^r + b^r)T_r = a^r T_r + b^r T_r$), but more general 'non-linear' parametrisations of the Lie algebra do not. To see some of this explicitly, we have the following worked example:

Extended worked example. As an example of how the Lie algebra induces a special choice of coordinates on the Lie group, consider the *affine group* $\mathrm{Aff}(1, \mathbb{R})$ in one dimension, given by the set of affine transformations $(a_1, a_2) : \mathbb{R} \to \mathbb{R}$, with $(a_1, a_2)x = a_1 x + a_2$ and $a_1 \neq 0$, under function composition. We can write the affine group as a matrix group, where the elements are of the form:

$$(a_1, a_2) = \begin{pmatrix} a_1 & a_2 \\ 0 & 1 \end{pmatrix}, \qquad \text{with the action on } \mathbb{R} \text{ given by:} \qquad \begin{pmatrix} a_1 & a_2 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x \\ 1 \end{pmatrix} = \begin{pmatrix} a_1 x + a_2 \\ 1 \end{pmatrix}.$$

The group multiplication is then simply matrix multiplication, the identity is the identity matrix, and inverses are given by matrix inverses. Explicitly the composition law is:

$$(a_1, a_2)(b_1, b_2) = \begin{pmatrix} a_1 & a_2 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} b_1 & b_2 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a_1 b_1 & a_1 b_2 + a_2 \\ 0 & 1 \end{pmatrix} = (a_1 b_1, a_1 b_2 + a_2).$$

Clearly, $\mathrm{Aff}(1, \mathbb{R})$ is a two-dimensional Lie group that is covered by some global coordinate chart $(a_1, a_2) \in \mathbb{R}^2 \backslash \{(0, y) : y \in \mathbb{R}\} = \mathbb{R}^2 \backslash \{y\text{-axis}\}$.

The composition of group elements is described in terms of their coordinates by the smooth function

$$(\phi_1, \phi_2) : \mathbb{R}^2\backslash\{y\text{-axis}\} \times \mathbb{R}^2\backslash\{y\text{-axis}\} \to \mathbb{R}^2\backslash\{y\text{-axis}\}, \qquad (\phi_1(\mathbf{a}, \mathbf{b}), \phi_2(\mathbf{a}, \mathbf{b})) = (a_1 b_1, a_1 b_2 + a_2).$$

Expanding around the identity, we have $(a_1, a_2) = (1 + \epsilon_1, \epsilon_2)$, $(b_1, b_2) = (1 + \eta_1, \eta_2)$ for infinitesimal $\epsilon_1, \epsilon_2, \eta_1, \eta_2$ giving:

$$(\phi_1(\mathbf{a}, \mathbf{b}), \phi_2(\mathbf{a}, \mathbf{b})) = (1 + \epsilon_1 + \eta_1 + \epsilon_1\eta_1 + \cdots, \eta_2 + \epsilon_2 + \epsilon_1\eta_2 + \cdots),$$

and hence we see to this order, the Lie group is described by the constants:

$$c^1{}_{11} = 1, \qquad c^1{}_{12} = 0, \qquad c^1{}_{21} = 0, \qquad c^1{}_{22} = 0,$$

$$c^2{}_{11} = 0, \qquad c^2{}_{12} = 1, \qquad c^2{}_{21} = 0, \qquad c^2{}_{22} = 0.$$

We note, in particular, that $c^r{}_{ab} \neq -c^r{}_{ba}$. The structure constants $f^r{}_{ab} = c^r{}_{ab} - c^r{}_{ba}$ are given by:

$$f^1{}_{11} = 0, \qquad f^1{}_{12} = 0, \qquad f^1{}_{21} = 0, \qquad f^1{}_{22} = 0,$$

$$f^2{}_{11} = 0, \qquad f^2{}_{12} = 1, \qquad f^2{}_{21} = -1, \qquad f^2{}_{22} = 0.$$

To construct the Lie algebra of the one-dimensional affine group, we look at curves passing through the identity $(1, 0)$. For a curve $(a_1(t), a_2(t))$ passing near the identity $(1, 0)$, we have that:

$$(a_1(t), a_2(t)) = \begin{pmatrix} a_1(t) & a_2(t) \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + t \begin{pmatrix} \dot{a}_1(0) & \dot{a}_2(0) \\ 0 & 0 \end{pmatrix} + O(t^2),$$

so that the Lie algebra of the Lie group is given by the set of matrices of the form:

$$\mathfrak{aff}(1, \mathbb{R}) = \left\{ \begin{pmatrix} x_1 & x_2 \\ 0 & 0 \end{pmatrix} : x_1, x_2 \in \mathbb{R} \right\} = \operatorname{span}_{\mathbb{R}} \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \right\}.$$

Using standard matrix exponentiation methods, one can show that:

$$\exp\left(\begin{pmatrix} x_1 & x_2 \\ 0 & 0 \end{pmatrix} \right) = \begin{pmatrix} e^{x_1} & \frac{x_2}{x_1}(e^{x_1} - 1) \\ 0 & 1 \end{pmatrix},$$

which is an element of the Lie group $\mathrm{Aff}(1, \mathbb{R})$ (specifically, an element of the part of $\mathrm{Aff}(1, \mathbb{R})$ connected to the identity). With this choice of coordinates on the Lie algebra, the Baker-Campbell-Hausdorff identity gives:

$$\exp\left(\begin{pmatrix} x_1 & x_2 \\ 0 & 0 \end{pmatrix} \right) \exp\left(\begin{pmatrix} y_1 & y_2 \\ 0 & 0 \end{pmatrix} \right) = \exp\left(\begin{pmatrix} x_1 & x_2 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} y_1 & y_2 \\ 0 & 0 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 0 & x_1 y_2 - y_1 x_2 \\ 0 & 0 \end{pmatrix} + \cdots \right),$$

which implies that the functions describing the Lie group structure $(\tilde{\phi}_1(\mathbf{x}, \mathbf{y}, \tilde{\phi}_2(\mathbf{x}, \mathbf{y}))$ are given in these coordinates (and to this order) by:

$$\tilde{\phi}_1(\mathbf{x}, \mathbf{y}) = x_1 + y_1 + \cdots, \quad \tilde{\phi}_2(\mathbf{x}, \mathbf{y}) = x_2 + y_2 + \frac{1}{2} x_1 y_2 - \frac{1}{2} y_1 x_2 + \cdots$$

and hence the corresponding $\tilde{c}^r_{ab}$ constants are equal to the structure constants $\frac{1}{2} f^r{}_{ab}$ in these coordinates, as anticipated.

Furthermore, to recover the general coordinates (a_1, a_2) on $\mathrm{Aff}(1, \mathbb{R})$ with $c^r{}_{ab} \neq \frac{1}{2} f^r{}_{ab}$ from some choice of coordinates on the Lie algebra, we would have to choose the parametrisation

$$x_1 = \log(a_1), \qquad x_2 = \frac{\log(a_1)}{a_1 - 1} a_2, \qquad \mathfrak{aff}(1, \mathbb{R}) = \left\{ \begin{pmatrix} \log(a_1) & a_2 \log(a_1)/(a_1 - 1) \\ 0 & 0 \end{pmatrix} : a_1 > 0, a_2 \in \mathbb{R} \right\}.$$

of the Lie algebra. This 'non-linear' parametrisation of the Lie algebra in terms of the a_1, a_2 coordinates does not respect the vector space structure of the Lie algebra, since the element with coordinates $a_1 + a'_1, a_2 + a'_2$ is *not* the same as the sum of the elements with coordinates a_1, a'_1 and a_2, a'_2.

This is the key take-away point: choosing canonical 'linear' coordinates on the Lie algebra such that elements are of the form $a^r T_r$, then using these coordinates a^r on the Lie group, gives rise to a simplified Lie group multiplication rule in these coordinates (with $c^a{}_{bc} = \frac{1}{2} f^a{}_{bc}$). Other 'non-linear' choices of coordinates on the Lie algebra give rise to all other possible coordinate expressions for the Lie group multiplication rule, for general $c^a{}_{bc}$.

6. (*Optional*) Using the same notation as the previous question, where $z^r = \phi^r(\mathbf{x}, \mathbf{y})$, obtain:

$$\frac{\partial z^r}{\partial y^s} = \lambda_s{}^a(\mathbf{y})\mu_a{}^r(\mathbf{z}), \qquad \mu_a{}^r(\mathbf{z}) = \frac{\partial}{\partial y^a}\phi^r(\mathbf{z}, \mathbf{y})\bigg|_{\mathbf{y}=\mathbf{0}}, \qquad \mu_a{}^r(\mathbf{z})\lambda_r{}^b(\mathbf{z}) = \delta_a{}^b.$$

Show that the equation for the structure constants $f^a{}_{bc}$ may also be expressed as:

$$\frac{\partial}{\partial y^r}\lambda_s{}^a(\mathbf{y}) - \frac{\partial}{\partial y^s}\lambda_r{}^a(\mathbf{y}) = -f^a{}_{bc}\lambda_r{}^b(\mathbf{y})\lambda_s{}^c(\mathbf{y}).$$

Solution: The first part of this question is about proving that the Lie algebra of a Lie group is well-defined (by showing that the left-invariant vector fields are indeed invariant). We will outline the construction in detail to give more context to the question (the required proofs are quite short).

As usual, let's assume that we are working with a Lie group G, whose elements can be labelled as $g(\mathbf{x})$ for some global coordinates $\mathbf{x} \in \mathbb{R}^n$. Suppose also that the multiplication rule for the group is specified by:

$$g(\mathbf{x})g(\mathbf{y}) = g(\phi(\mathbf{x}, \mathbf{y})), \qquad (*)$$

where $\phi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$. We can also suppose without loss of generality that the global coordinate chart is centred on the identity of the group, $g(\mathbf{0}) = e$.

We would like to define the *Lie algebra* of the Lie group to be a 'linearisation of the Lie group close to the identity $e \in G$'. To see how this works, suppose that $\boldsymbol{\theta}$ are small coordinates, so that $g(\boldsymbol{\theta})$ is a group element 'close to the identity'. Multiplying $g(\boldsymbol{\theta})$ on the left[7] by any element $g(\mathbf{x}) \in G$ heuristically induces a small change in the coordinates of the group element $g(\mathbf{x})$, by smoothness of the group operation:

$$g(\mathbf{x})g(\boldsymbol{\theta}) = g(\mathbf{x} + \mathbf{dx}).$$

Using the group multiplication law $(*)$ and comparing coordinates, this equation can be rewritten as:

$$x^r + dx^r = \phi^r(\mathbf{x}, \boldsymbol{\theta}) \qquad \Rightarrow \qquad x^r + dx^r = \phi^r(\mathbf{x}, \mathbf{0}) + \theta^a \frac{\partial \phi^r}{\partial \theta^a}(\mathbf{x}, \boldsymbol{\theta})\bigg|_{\boldsymbol{\theta}=\mathbf{0}} + O(\boldsymbol{\theta}^2),$$

using the fact that $\boldsymbol{\theta}$ is small to Taylor expand the right hand side to linear order. Now recall from Question 6(a) that $\phi^r(\mathbf{x}, \mathbf{0}) = x^r$, so to linear order this equation reduces simply to:

$$\boxed{dx^r = \theta^a \frac{\partial \phi^r}{\partial \theta^a}(\mathbf{x}, \boldsymbol{\theta})\bigg|_{\boldsymbol{\theta}=\mathbf{0}} = \theta^a \mu_a{}^r(\mathbf{x}), \qquad \text{where} \quad \mu_a{}^r(\mathbf{x}) := \frac{\partial \phi^r}{\partial \theta^a}(\mathbf{x}, \boldsymbol{\theta})\bigg|_{\boldsymbol{\theta}=\mathbf{0}}.}$$

This represents the change in the coordinates, dx^r, as a linear combination of the coordinates $\boldsymbol{\theta}$; in particular, we have reduced the action of group elements near the identity, $g(\boldsymbol{\theta})$, to something linear - exactly what we want for the Lie algebra, the desired 'linearisation' of the Lie group. This suggests defining the Lie algebra to be:

$$\mathfrak{g} \overset{?}{:=} \{\theta^a \mu_a(\mathbf{x}) : \boldsymbol{\theta} \in \mathbb{R}^n, \mu_a(\mathbf{x}) = (\mu_a^1(\mathbf{x}), ..., \mu_a^n(\mathbf{x}))\}.$$

This definition looks a bit suspect though - we have ended up with a definition that depends on some arbitrary point $\mathbf{x} \in \mathbb{R}^n$, labelling some arbitrary group element $g(\mathbf{x}) \in G$. Indeed, our definition of the Lie algebra will not remain invariant under a change of the coordinate $\mathbf{x} \in \mathbb{R}^n$ to some new coordinate $\mathbf{x}' \in \mathbb{R}^n$.

[7] The choice to multiply on the left is arbitrary and will lead to *left-invariant vector fields* rather than *right-invariant vector fields*; both can be used to define the Lie algebra.

There is a simple fix to this problem. We note that ${\mu_a}^r(\mathbf{x})$ can be thought of as the components of a vector $({\mu_a}^1(\mathbf{x}), ..., {\mu_a}^n(\mathbf{x}))$, which depends on the point $\mathbf{x} \in \mathbb{R}^n$. To upgrade this to a global definition, we introduce an associated *vector field*[8] T_a, which we define to be the differential operator:

$$T_a = {\mu_a}^r(\mathbf{x})\frac{\partial}{\partial x^r}.$$

In order to avoid the problem from our earlier definition of the Lie algebra, the key requiement for this vector field is that it should *not* depend on the coordinate $\mathbf{x} \in \mathbb{R}^n$. That is, for any two coordinates $\mathbf{y}, \mathbf{z} \in \mathbb{R}^n$, we should have:

$${\mu_a}^r(\mathbf{y})\frac{\partial}{\partial y^r} = {\mu_a}^r(\mathbf{z})\frac{\partial}{\partial z^r}.$$

We prove this as follows:

Proposition: The vector field T_a is indeed invariant under the choice of coordinate.

Proof: Suppose that $\mathbf{y}, \mathbf{z} \in \mathbb{R}^n$ are coordinates of group elements $g(\mathbf{y}), g(\mathbf{z}) \in G$ respectively. From above, we know that the group elements $g(\mathbf{y} + \mathbf{dy}) = g(\mathbf{y})g(\boldsymbol{\theta})$, $g(\mathbf{z} + \mathbf{dz}) = g(\mathbf{z})g(\boldsymbol{\theta})$ given by multiplying by an element close to the identity $g(\boldsymbol{\theta}) \in G$ can be described by the infinitesimal variations:

$$dy^r = \theta^a {\mu_a}^r(\mathbf{y}), \qquad dz^r = \theta^a {\mu_a}^r(\mathbf{z}).$$

Next, note that we can view ${\mu_a}^r(\mathbf{z})$ as the components of an *invertible* matrix. To see the invertibility, note that for fixed $\mathbf{z} \in \mathbb{R}^n$, the map $\phi(\mathbf{z}, \cdot) : \mathbb{R}^n \to \mathbb{R}^n$ is a smooth bijection (given any fixed $\mathbf{z}$, for each $\mathbf{y}$ there exists a unique $\phi(\mathbf{z}, \mathbf{y})$ satisfying $g(\mathbf{z})g(\mathbf{y}) = g(\phi(\mathbf{z}, \mathbf{y}))$, and vice-versa - this follows from the fact that we can always invert group elements), hence its Jacobian matrix is an invertible matrix; ${\mu_a}^r(\mathbf{z})$ is precisely this Jacobian matrix.

In particular, the first equation can be rewritten in the form $\theta^a = {\lambda_r}^a(\mathbf{y})dy^r$, using the inverse matrix. Substituting into the second equation, we have:

$$dz^r = {\lambda_s}^a(\mathbf{y}){\mu_a}^r(\mathbf{z})dy^s \qquad \Rightarrow \qquad \frac{\partial z^r}{\partial y^s} = {\lambda_s}^a(\mathbf{y}){\mu_a}^r(\mathbf{z}),$$

as $\mathbf{dy}, \mathbf{dz} \to \mathbf{0}$. Thus we have obtained the first result required in the question. From this result, it is simple to verify the claimed proposition:

$${\mu_a}^s(\mathbf{y})\frac{\partial}{\partial y^s} = {\mu_a}^s(\mathbf{y})\frac{\partial z^r}{\partial y^s}\frac{\partial}{\partial z^r} = {\mu_a}^s(\mathbf{y}){\lambda_s}^a(\mathbf{y}){\mu_a}^r(\mathbf{z})\frac{\partial}{\partial z^r} = {\mu_a}^r(\mathbf{z})\frac{\partial}{\partial z^r}. \qquad \square$$

It follows that we can make a coordinate-independent definition of the Lie algebra via:

Definition: The *Lie algebra* $\mathfrak{g}$ *of the Lie group* G is defined to be:

$$\mathfrak{g} := \mathrm{span}_{\mathbb{R}}\{T_a\}.$$

The vector fields T_a are called *left-invariant vector fields*. The name is appropriate because no matter which left-multiplication by a group element we choose to start from, $g(\mathbf{x})g(\boldsymbol{\theta})$ or $g(\mathbf{y})g(\boldsymbol{\theta})$, for $\mathbf{x} \neq \mathbf{y}$, we arrive at the same definitions of the vector fields T_a.

[8] See the Part III General Relativity course, or the Part III Differential Geometry course, for a much more careful and detailed description of vector fields on manifolds. What we have written down isn't a vector field in the strictest sense; it is the expression for a vector field in a given coordinate system - however, since we assume our coordinates are global, this distinction is immaterial in our case.

The second part of the question concerns the *Lie bracket structure* of a Lie algebra; let's begin by recalling the construction from lectures. Recall that we define the *Lie bracket* of two vector fields $X = X^i(\mathbf{x})\partial_i$, $Y = Y^i(\mathbf{x})\partial_i$ to be another vector field:[9]

$$[X,Y] = [X,Y]^i(\mathbf{x})\frac{\partial}{\partial x^i}, \qquad \text{with} \qquad [X,Y]^i(\mathbf{x}) = XY^i(\mathbf{x}) - YX^i(\mathbf{x}).$$

It is possible to show that the Lie bracket obeys some nice properties including linearity, antisymmetry and the *Jacobi identity*. The key feature that is important here though is that the Lie algebra $\mathfrak{g}$ is *closed* under the Lie bracket.

Proposition: The Lie algebra $\mathfrak{g}$ of a Lie group G is closed under the Lie bracket of vector fields.

Proof: By linearity of the Lie bracket, it is sufficient to check that $[T_a, T_b] = f^c{}_{ab} T_c$ for some constants $f^c{}_{ab}$, called the *structure constants* of the Lie algebra. Using the definition of the left-invariant vector fields T_a, T_b from above, and the definition of the Lie bracket, we can write:

$$\begin{aligned}
[T_a, T_b] &= [T_a, T_b]^i(\mathbf{x})\frac{\partial}{\partial x^i} \\
&= \left(T_a T_b{}^i(\mathbf{x}) - T_b T_a{}^i(\mathbf{x})\right)\frac{\partial}{\partial x^i} \\
&= \left(\mu_a{}^r(\mathbf{x})\frac{\partial}{\partial x^r}\mu_b{}^i(\mathbf{x}) - \mu_b{}^r(\mathbf{x})\frac{\partial}{\partial x^r}\mu_a{}^i(\mathbf{x})\right)\frac{\partial}{\partial x^i} \\
&= \left(\mu_a{}^r(\mathbf{x})\frac{\partial}{\partial x^r}\mu_b{}^j(\mathbf{x}) - \mu_b{}^r(\mathbf{x})\frac{\partial}{\partial x^r}\mu_a{}^j(\mathbf{x})\right)\lambda_j{}^c(\mathbf{x})\mu_c{}^i(\mathbf{x})\frac{\partial}{\partial x^i}, \\
&= \left(\mu_a{}^r(\mathbf{x})\frac{\partial}{\partial x^r}\mu_b{}^j(\mathbf{x}) - \mu_b{}^r(\mathbf{x})\frac{\partial}{\partial x^r}\mu_a{}^j(\mathbf{x})\right)\lambda_j{}^c(\mathbf{x})T_c,
\end{aligned}$$

where in the second from last line we inserted the identity matrix, $\delta_j{}^i = \lambda_j{}^c(\mathbf{x})\mu_c{}^i(\mathbf{x})$. It remains to show that the quantity:

$$f^c{}_{ab}(\mathbf{x}) = \left(\mu_a{}^r(\mathbf{x})\frac{\partial}{\partial x^r}\mu_b{}^j(\mathbf{x}) - \mu_b{}^r(\mathbf{x})\frac{\partial}{\partial x^r}\mu_a{}^j(\mathbf{x})\right)\lambda_j{}^c(\mathbf{x})$$

is independent of $\mathbf{x}$ (i.e. the structure constants are indeed constants!). To do so, we recall from the first part of the question that for any other coordinate $\mathbf{y}$ on the Lie group, we have the relation:

$$\frac{\partial x^r}{\partial y^s} = \lambda_s{}^a(\mathbf{y})\mu_a{}^r(\mathbf{x}) \qquad \Rightarrow \qquad \mu_a{}^r(\mathbf{x}) = \mu_a{}^s(\mathbf{y})\frac{\partial x^r}{\partial y^s}. \tag{$*$}$$

Substituting this relationship into the equation for the structure constants $f^a{}_{bc}$, we have:

$$\begin{aligned}
f^c{}_{ab}(\mathbf{x}) &= \left(\mu_a{}^s(\mathbf{y})\frac{\partial x^r}{\partial y^s}\frac{\partial}{\partial x^r}\left(\mu_b{}^k(\mathbf{y})\frac{\partial x^j}{\partial y^k}\right) - \mu_b{}^s(\mathbf{y})\frac{\partial x^r}{\partial y^s}\frac{\partial}{\partial x^r}\left(\mu_a{}^k(\mathbf{y})\frac{\partial x^j}{\partial y^k}\right)\right)\lambda_j{}^c(\mathbf{x}) \\
&= \left(\mu_a{}^s(\mathbf{y})\frac{\partial}{\partial y^r}\left(\mu_b{}^k(\mathbf{y})\frac{\partial x^j}{\partial y^k}\right) - \mu_b{}^s(\mathbf{y})\frac{\partial}{\partial y^r}\left(\mu_a{}^k(\mathbf{y})\frac{\partial x^j}{\partial y^k}\right)\right)\lambda_j{}^c(\mathbf{x}) \\
&= \left(\mu_a{}^s(\mathbf{y})\frac{\partial}{\partial y^r}\mu_b{}^k(\mathbf{y}) - \mu_b{}^s(\mathbf{y})\frac{\partial}{\partial y^r}\mu_a{}^k(\mathbf{y})\right)\lambda_j{}^c(\mathbf{x})\frac{\partial x^j}{\partial y^k},
\end{aligned}$$

using the chain rule in the second line, and the symmetry of mixed partial derivatives in the final line. To finish, note that the relation ($*$) also implies $\lambda_j{}^c(\mathbf{x})\partial x^j/\partial y^k = \lambda_k{}^c(\mathbf{y})$, which allows us to conclude that $f^c{}_{ab}(\mathbf{x}) = f^c{}_{ab}(\mathbf{y})$. □

[9] See Part III General Relativity, or Part III Differential Geometry, for more details.

In the second part of the question, we are asked to derive the following alternative definition for the structure constants of the Lie algebra:

$$\frac{\partial}{\partial y^r}\lambda_s{}^c(\mathbf{y}) - \frac{\partial}{\partial y^s}\lambda_r{}^c(\mathbf{y}) = -f^c{}_{ab}\lambda_r{}^a(\mathbf{y})\lambda_s{}^b(\mathbf{y}).$$

This can be achieved directly from a series of manipulations of our original formula for the structure constants. We have:

$$\begin{aligned} f^c{}_{ab}\lambda_r{}^a(\mathbf{y})\lambda_s{}^b(\mathbf{y}) &= \left(\mu_a{}^i(\mathbf{y})\frac{\partial}{\partial y^i}\mu_b{}^j(\mathbf{y}) - \mu_b{}^i(\mathbf{y})\frac{\partial}{\partial y^i}\mu_a{}^j(\mathbf{y})\right)\lambda_j{}^c(\mathbf{y})\lambda_r{}^a(\mathbf{y})\lambda_s{}^b(\mathbf{y}) \\ &= \left(\delta_r{}^i\lambda_s{}^b(\mathbf{y})\frac{\partial}{\partial y^i}\mu_b{}^j(\mathbf{y}) - \delta_s{}^i\lambda_r{}^a(\mathbf{y})\frac{\partial}{\partial y^i}\mu_a{}^j(\mathbf{y})\right)\lambda_j{}^c(\mathbf{y}) \\ &= \left(\lambda_s{}^b(\mathbf{y})\frac{\partial}{\partial y^r}\mu_b{}^j(\mathbf{y}) - \lambda_r{}^a(\mathbf{y})\frac{\partial}{\partial y^s}\mu_a{}^j(\mathbf{y})\right)\lambda_j{}^c(\mathbf{y}) \end{aligned}$$

using the fact that the λ-matrices are inverse to the μ-matrices in the second line. Now consider taking the λ-matrices inside the derivatives, using the product rule in reverse:

$$\begin{aligned} f^c{}_{ab}\lambda_r{}^a(\mathbf{y})\lambda_s{}^b(\mathbf{y}) &= \left(\frac{\partial}{\partial y^r}\left(\lambda_s{}^b(\mathbf{y})\mu_b{}^j(\mathbf{y})\right) - \mu_b{}^j(\mathbf{y})\frac{\partial}{\partial y^r}\lambda_s{}^b(\mathbf{y}) - \frac{\partial}{\partial y^s}\left(\lambda_r{}^a(\mathbf{y})\mu_a{}^j(\mathbf{y})\right) + \mu_a{}^j(\mathbf{y})\frac{\partial}{\partial y^s}\lambda_r{}^a(\mathbf{y})\right)\lambda_j{}^c(\mathbf{y}) \\ &= \left(\frac{\partial}{\partial y^r}\left(\delta_s{}^j\right) - \mu_b{}^j(\mathbf{y})\frac{\partial}{\partial y^r}\lambda_s{}^b(\mathbf{y}) - \frac{\partial}{\partial y^s}\left(\delta_r{}^j\right) + \mu_a{}^j(\mathbf{y})\frac{\partial}{\partial y^s}\lambda_r{}^a(\mathbf{y})\right)\lambda_j{}^c(\mathbf{y}) \\ &= \left(\mu_a{}^j(\mathbf{y})\frac{\partial}{\partial y^s}\lambda_r{}^a(\mathbf{y}) - \mu_b{}^j(\mathbf{y})\frac{\partial}{\partial y^r}\lambda_s{}^b(\mathbf{y})\right)\lambda_j{}^c(\mathbf{y}) \\ &= \frac{\partial}{\partial y^s}\lambda_r{}^c(\mathbf{y}) - \frac{\partial}{\partial y^r}\lambda_s{}^c(\mathbf{y}), \end{aligned}$$

again applying the fact that the λ-matrices are inverse to the μ-matrices, and using the fact that the identity matrix is constant so has zero derivative. The required result follows immediately.

❊ **Comments:** The equivalent equation we obtain here for the structure constants arises naturally if we introduce the notion of *left-invariant one-forms*. Let's see how this is done.

Recall that given a coordinate $\boldsymbol{\theta}$ close to zero, we can write $g(\mathbf{x})g(\boldsymbol{\theta}) = g(\mathbf{x} + \mathbf{dx})$ for $\mathbf{dx}$ an infinitesimal change in the coordinate $\mathbf{x}$; this is because the group multiplication is a smooth map. As above, it follows that:

$$x^r + dx^r = \phi^r(\mathbf{x}, \boldsymbol{\theta}) \qquad \Rightarrow \qquad dx^r = \theta^a \mu_a{}^r(\mathbf{x}).$$

Using the fact that $\mu_a{}^r(\mathbf{x})$ is invertible, we can write $\theta^a = \lambda_r{}^a(\mathbf{x})dx^r$. This suggests introducing *left-invariant one forms*, defined in any coordinate patch by:

$$\omega^a = \lambda_r{}^a(\mathbf{x})dx^r.$$

Taking the exterior derivative, we have:

$$d\omega^a = \left(\frac{\partial}{\partial x^s}\lambda_r{}^a(\mathbf{x})\right)dx^s \wedge dx^r = \frac{1}{2}\left(\frac{\partial}{\partial x^s}\lambda_r{}^a(\mathbf{x}) - \frac{\partial}{\partial x^r}\lambda_s{}^a(\mathbf{x})\right)dx^s \wedge dx^r.$$

Inserting appropriate μ and λ matrices, we can reduce this equation to:

$$d\omega^a = -\frac{1}{2}f^a{}_{bc}\omega^b \wedge \omega^c,$$

using the equation for the structure constants we developed above.

7. This question regards the Pauli matrices. Verify the following properties of the Pauli matrices $\boldsymbol{\sigma} := (\sigma_1, \sigma_2, \sigma_3)$:

(a) $\sigma_i \sigma_j = \delta_{ij} I + i\epsilon_{ijk}\sigma_k$;

(b) $\sigma_2 \boldsymbol{\sigma} \sigma_2 = -\boldsymbol{\sigma}^*$;

(c) $\exp(-i\theta\hat{\mathbf{n}} \cdot \boldsymbol{\sigma}/2) = I\cos(\theta/2) - i\hat{\mathbf{n}} \cdot \boldsymbol{\sigma}\sin(\theta/2)$;

Solution: (a) One method for this part of the question simply involves multiplying all the pairs of Pauli matrices out, e.g. computing $\sigma_1\sigma_2, \sigma_1\sigma_3$, etc. However, there is also a quick method based on the transformation properties of the Pauli matrices under rotations.

First let's notice that the set $\{I, \sigma_1, \sigma_2, \sigma_3\}$ forms a basis for the four-dimensional complex vector space consisting of 2×2 complex matrices. Hence, since $\sigma_i\sigma_j$ is a 2×2 complex matrix for all i, j, we must be able to write:

$$\sigma_i \sigma_j = A_{ij} I + iB_{ijk}\sigma_k \tag{$*$}$$

for some coefficients A_{ij}, B_{ijk}. We will introduce an action of rotations $R \in SO(3)$ on the Pauli matrices, and show that under this action the objects A_{ij}, B_{ijk} transform as *isotropic tensors*; from this fact we will be able to deduce the form of $\sigma_i\sigma_j$.

We begin by introducing the following transformation.[10] Under a rotation $R \in SO(3)$, the general 2×2 complex matrix $z_0 I + \mathbf{z} \cdot \boldsymbol{\sigma}$, with $z_0 \in \mathbb{C}$ and $\mathbf{z} \in \mathbb{C}^3$, can be defined to transform as:

$$z_0 I + \mathbf{z} \cdot \boldsymbol{\sigma} \mapsto z_0 I + (R\mathbf{z}) \cdot \boldsymbol{\sigma}; \qquad \text{or in index notation,} \qquad z_0 I + z_i\sigma_i \mapsto z_0 I + R_{ji} z_i \sigma_j.$$

In particular, this means that the Pauli matrices themselves can be thought of as transforming via $\sigma_i \mapsto R_{ji}\sigma_j$, with the identity matrix transformed as $I \mapsto I$.

We now apply this transformation law to both sides of the equation ($*$). We have:

$$\sigma_i\sigma_j \mapsto R_{ai}R_{bj}\sigma_a\sigma_b, \qquad \text{and} \qquad A_{ij}I + iB_{ijk}\sigma_k \mapsto A_{ij}I + iB_{ijk}R_{ak}\sigma_a.$$

Comparing both sides of the transformation law, we see that:

$$R_{ai}R_{bj}A_{ab}I + iR_{ai}R_{bj}B_{abk}\sigma_k = A_{ij}I + iB_{ijk}R_{ak}\sigma_a.$$

Since the matrices I, σ_i are all linearly independent, we can compare both sides to read off the transformation laws:

$$R_{ai}R_{bj}A_{ab} = A_{ij}, \qquad R_{ai}R_{bj}B_{abl} = B_{ijk}R_{lk}.$$

The second transformation law can be rewritten as $R_{ai}R_{bj}R_{ck}B_{abc} = B_{ijk}$ by multiplying through by the inverse of R. Thus we see that A_{ij} and B_{ijk} are two-index and three-index isotropic tensors under rotations $R \in SO(3)$. It follows that: $A_{ij} = a\delta_{ij}$ and $B_{ijk} = b\epsilon_{ijk}$. Hence we have shown:

$$\sigma_i\sigma_j = a\delta_{ij}I + ib\epsilon_{ijk}\sigma_k.$$

It remains to fix the values of a, b. We note that $\sigma_3\sigma_3 = \sigma_3^2 = I$, so we deduce that $a = 1$. We note also that:

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

and hence $ib\epsilon_{123}\sigma_3 = i\sigma_3$, implying $b = 1$. The identity follows.

[10] Compare this to the proofs of the relationships $SO(3) \cong SU(2)/\mathbb{Z}_2$ and $SO_+(1,3) \cong SL(2,\mathbb{C})/\mathbb{Z}_2$ later in the course. In the latter case, there is a correspondence between four-vectors and certain 2×2 complex matrices, given by $x^\mu \leftrightarrow x^\mu\sigma_\mu = x^0 I + \mathbf{x} \cdot \boldsymbol{\sigma}$, so that we can view our transformation $R \in SO(3)$ as fixing the time coordinate (corresponding to the identity matrix) and transforming the space coordinates (corresponding to the Pauli matrices).

(b) For once, it is slightly easier not to use index notation. We use the multiplication rules from above to find:

$$\sigma_2\sigma_1\sigma_2 = \sigma_2(i\sigma_3) = i(i\sigma_1) = -\sigma_1 = -\sigma_1^*$$
$$\sigma_2\sigma_2\sigma_2 = \sigma_2(I) = \sigma_2 = -\sigma_2^*$$
$$\sigma_2\sigma_3\sigma_2 = \sigma_2(-i\sigma_1) = -i(-i\sigma_3) = -\sigma_3 = -\sigma_3^*.$$

Hence $\sigma_2\boldsymbol{\sigma}\sigma_2 = -\boldsymbol{\sigma}^*$ as required.

(c) We are now asked to establish the identity: $\exp(-i\theta\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}/2) = I\cos(\theta/2) - i\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}\sin(\theta/2)$. The idea is to expand the exponential directly; hence, we will need to be able to evaluate powers of $\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}$. We note that for $r \in \{0, 1, 2...\}$, we have:

$$\begin{aligned}
(\hat{\mathbf{n}}\cdot\sigma)^{2r} &= (\hat{n}_i\sigma_i\hat{n}_j\sigma_j)^r \\
&= (\hat{n}_i\hat{n}_j(I\delta_{ij} + i\epsilon_{ijk}\sigma_k))^r && \text{(using part (i))} \\
&= I && \text{(since } \hat{n}_i\hat{n}_j\epsilon_{ijk} = 0, \text{ and } \hat{\mathbf{n}}^2 = 1)
\end{aligned}$$

This also implies $(\hat{\mathbf{n}}\cdot\boldsymbol{\sigma})^{2r+1} = \hat{\mathbf{n}}\cdot\boldsymbol{\sigma}$. Now turning our attention to the exponential $\exp(-i\theta\mathbf{n}\cdot\sigma/2)$, we can Taylor expand to find:

$$\begin{aligned}
\exp\left(-\frac{i\theta\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}}{2}\right) &= \sum_{k=0}^{\infty}\frac{(-i\theta\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}/2)^k}{k!} \\
&= \sum_{r=0}^{\infty}\frac{(-i\theta\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}/2)^{2r}}{(2r)!} + \sum_{r=0}^{\infty}\frac{(-i\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}/2)^{2r+1}}{(2r+1)!} \\
&= \left(\sum_{r=0}^{\infty}\frac{(-1)^r(\theta/2)^{2r}}{(2r)!}\right) I - i\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}\left(\sum_{r=0}^{\infty}\frac{(-1)^r(\theta/2)^{2r+1}}{(2r+1)!}\right) \\
&= \cos\left(\frac{\theta}{2}\right) I - i\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}\sin\left(\frac{\theta}{2}\right).
\end{aligned}$$

8. Verify the Baker-Campbell-Hausdorff formula

$$\exp(tA)\exp(tB) = \exp\left(tA + tB + \frac{t^2}{2}[A,B] + \frac{t^3}{12}\left([A,[A,B]] + [B,[B,A]]\right) + \cdots\right)$$

up to and including terms of order t^2 (i.e. omitting the order t^3 term).

Solution: Even though the question only asks us to verify the result to order $O(t^2)$, we will perform the calculation to order $O(t^3)$ for pedagogical purposes. Expanding the product $e^{tA}e^{tB}$, we have:

$$\begin{aligned}\exp(tA)\exp(tB) &= \left(I + tA + \frac{1}{2}t^2A^2 + \frac{1}{6}t^3A^3 + \cdots\right)\left(I + tB + \frac{1}{2}t^2B^2 + \frac{1}{6}t^3B^3 + \cdots\right) \\ &= I + t(A+B) + \frac{1}{2}t^2\left(A^2 + 2AB + B^2\right) + \frac{1}{6}t^3\left(A^3 + 3A^2B + 3AB^2 + B^3\right) + O(t^4). \qquad (*)\end{aligned}$$

Expanding the right hand side, we have:

$$\begin{aligned}&\exp\left(tA + tB + \frac{t^2}{2}[A,B] + \frac{t^3}{12}\left([A,[A,B]] + [B,[B,A]]\right) + \cdots\right) \\ &\qquad = I + \left(tA + tB + \frac{t^2}{2}[A,B] + \frac{t^3}{12}\left([A,[A,B]] + [B,[B,A]]\right) + \cdots\right) \\ &\qquad\quad + \frac{1}{2}\left(tA + tB + \frac{t^2}{2}[A,B] + \frac{t^3}{12}\left([A,[A,B]] + [B,[B,A]]\right) + \cdots\right)^2 \\ &\qquad\quad + \frac{1}{6}\left(tA + tB + \frac{t^2}{2}[A,B] + \frac{t^3}{12}\left([A,[A,B]] + [B,[B,A]]\right) + \cdots\right)^3 + O(t^4).\end{aligned}$$

We now collect all like terms on the right hand side. We have:

- The term of order $O(t^0)$ is simply I on the right hand side. This agrees with $(*)$.
- The term of order $O(t^1)$ is $tA + tB$, which again agrees with $(*)$.
- The term of order $O(t^2)$ is:
$$\frac{t^2}{2}[A,B] + \frac{t^2}{2}(A+B)^2 = \frac{t^2}{2}\left(AB - BA + A^2 + AB + BA + B^2\right) = \frac{t^2}{2}\left(A^2 + 2AB + B^2\right),$$
which agrees with $(*)$.
- Finally, the term of order $O(t^3)$ is:
$$\begin{aligned}&\frac{t^3}{12}\left([A,[A,B]] + [B,[B,A]]\right) + \frac{t^3}{4}\left((A+B)[A,B] + [A,B](A+B)\right) + \frac{t^3}{6}(A+B)^3 \\ &= \frac{t^3}{12}\Big(A^2B - 2ABA + BA^2 + B^2A - 2BAB + AB^2 \\ &\quad + 3A^2B - 3ABA + 3BAB - 3B^2A + 3ABA - 3BA^2 + 3AB^2 - 3BAB \\ &\quad + 2A^3 + 2A^2B + 2ABA + 2BA^2 + 2AB^2 + 2BAB + 2B^2A + 2B^3\Big) \\ &= \frac{t^3}{6}\left(A^3 + 3A^2B + 3AB^2 + B^3\right),\end{aligned}$$
which again agrees with $(*)$.

Hence we have verified the Baker-Campbell-Hausdorff identity to the given order.

9. Let $g(t) = \exp(it\sigma_1)$. By evaluating $g(t)$ as a matrix, show that $\{g(t) : 0 \leq t \leq 2\pi\}$ is a one-parameter subgroup of $SU(2)$. Describe geometrically how this subgroup sits inside the $SU(2)$ manifold.

➾ **Solution:** We can evaluate $g(t)$ using the identity in Question 7(c). Using the notation there, we see that $\theta = -2t$, $\hat{\mathbf{n}} = (1,0,0)^T$. Hence we have:

$$\exp(it\sigma_1) = \cos(t)I + i\sigma_1\sin(t) = \begin{pmatrix} \cos(t) & i\sin(t) \\ i\sin(t) & \cos(t) \end{pmatrix}.$$

The resulting matrix is clearly in $SU(2)$, since it has orthonormal columns (so is unitary) and has unit determinant.

It remains to show that $\{g(t) : 0 \leq t \leq 2\pi\}$ is a one-parameter subgroup of $SU(2)$. Recall from lectures that the key requirement is the multiplication rule $g(t)g(s) = g(t+s)$; we can verify this using standard trigonometric identities:

$$\begin{aligned} g(t)g(s) &= \begin{pmatrix} \cos(t) & i\sin(t) \\ i\sin(t) & \cos(t) \end{pmatrix}\begin{pmatrix} \cos(s) & i\sin(s) \\ i\sin(s) & \cos(s) \end{pmatrix} \\ &= \begin{pmatrix} \cos(t)\cos(s) - \sin(t)\sin(s) & i\cos(t)\sin(s) + i\cos(s)\sin(t) \\ i\sin(t)\cos(s) + i\sin(s)\cos(t) & \cos(t)\cos(s) - \sin(t)\sin(s) \end{pmatrix} \\ &= \begin{pmatrix} \cos(t+s) & i\sin(t+s) \\ i\sin(t+s) & \cos(t+s) \end{pmatrix} \\ &= g(t+s), \end{aligned}$$

as required. Alternatively, we can use the matrix form: $g(t)g(s) = \exp(it\sigma_1)\exp(is\sigma_1) = \exp(i(t+s)\sigma_1) = g(t+s)$, since $it\sigma_1$ and $is\sigma_1$ commute.

Verifying the multiplication rule is sufficient for $\{g(t) : t \in \mathbb{R}\}$ to be a one-parameter subgroup of $SU(2)$. To show that we can in fact write the set as $\{g(t) : 0 \leq t \leq 2\pi\}$, we need to show that some elements of $\{g(t) : t \in \mathbb{R}\}$ are the same; it is sufficient to check that $g(0) = g(2\pi)$. Examining the matrix form above, we have:

$$g(0) = \exp(0) = I = \cos(2\pi)I + i\sigma_1\sin(2\pi) = \exp(2\pi i\sigma_1) = g(2\pi),$$

and hence it follows that we can write $\{g(t) : t \in \mathbb{R}\} = \{g(t) : 0 \leq t \leq 2\pi\}$ as required (furthermore, you can easily verify that no two elements of this final set are equal to one another - apart from $g(0) = g(2\pi)$ - so there is no more redundancy in the given set).

Finally, we are asked to describe how the subgroup $\{g(t) : 0 \leq t \leq 2\pi\}$ sits inside the $SU(2)$ manifold. Recall that the $SU(2)$ manifold is the unit 3-sphere (i.e. the unit sphere sitting in four-dimensions); this follows from the general form:

$$U = a_0 I + i\mathbf{a}\cdot\boldsymbol{\sigma}, \qquad a_0^2 + |\mathbf{a}|^2 = 1$$

for an $SU(2)$ matrix U which we derived in Question 4. The matrix exponential we computed above shows that the elements of the one-parameter subgroup obey $a_0 = \cos(t)$, $\mathbf{a} = (\sin(t), 0, 0)$, so that the one-parameter subgroup is a circle on the surface of the 3-sphere.

We can also think of this as the intersection of the hyperplane $\{a_2 = 0, a_3 = 0\}$ with the 3-sphere; since these are coordinate axes, the hyperplanes pass through the centre of the 3-sphere, and thus in particular our circle is a *great circle* on the surface of the 3-sphere.

10. Let $\exp(iH) = U$. Show that if H is Hermitian, then U is unitary. Show that if H is also traceless then $\det(U) = 1$. How do these results relate to the theorem that the exponential map $X \mapsto \exp(X)$ sends $\mathfrak{g}$, the Lie algebra of G, to G?

Solution: In the first part of this question, we must show that $U = \exp(iH)$ is unitary whenever H is Hermitian. First, note that $U^\dagger = \exp((iH)^\dagger) = \exp(-iH)$ whenever H is Hermitian (one can verify this carefully by using the power series definition of the exponential for example). Therefore,

$$U^\dagger U = \exp(iH)\exp(-iH) = \exp(iH - iH) = \exp(0) = I,$$

using the fact that iH and $-iH$ commute, so that we can combine the exponential's arguments (e.g. by the Baker-Campbell-Hausdorff identity, or by an explicit check with the definition of the matrix exponential). The result follows.

We are next asked to show that if H is additionally traceless, we have $\det(U) = 1$. We use the following famous identity:

Proposition: If M is a diagonalisable matrix,[11] we have:

$$\det(\exp(M)) = \exp(\mathrm{Tr}(M)).$$

Proof: Let $M = PDP^{-1}$ where D is diagonal, with the eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$ of M along the diagonal. Then:

$$\det(\exp(PDP^{-1}))) = \det(P\exp(D)P^{-1}) = \det(P)\det(\exp(D))\det(P^{-1}) = \det(\exp(D)),$$

using the identity $\exp(PDP^{-1}) = P\exp(D)P^{-1}$, which can be proved directly from the series definition of the matrix exponential. Since D is diagonal, taking the exponential is trivial; we get:

$$\exp(D) = \exp\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} = \begin{pmatrix} e^{\lambda_1} & 0 & \cdots & 0 \\ 0 & e^{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{\lambda_n} \end{pmatrix} = \mathrm{diag}\{e^{\lambda_1}, ..., e^{\lambda_n}\}.$$

It follows that:

$$\det(\exp(D)) = e^{\lambda_1}...e^{\lambda_n} = e^{\lambda_1 + \cdots + \lambda_n} = e^{\mathrm{Tr}(M)}. \qquad \square$$

In particular, since H is Hermitian, it is diagonalisable (and hence iH is also diagonalisable). It follows that:

$$\det(U) = \det(\exp(iH)) = \exp\left(\mathrm{Tr}(iH)\right) = \exp\left(i\mathrm{Tr}(H)\right) = \exp(0) = 1,$$

using the tracelessness of H.

Finally, we are asked to comment on how these results relate to the theorem that $\exp$ is a map from the Lie algebra of a Lie group to the Lie group, $\exp : \mathfrak{g} \to G$. If we know already that the Lie algebras of $U(n)$ and $SU(n)$ are given respectively by:

$$\mathfrak{u}(n) = \{n \times n \text{ anti-Hermitian complex matrices}\},$$

$$\mathfrak{su}(n) = \{\text{traceless } n \times n \text{ anti-Hermitian complex matrices}\},$$

then this question merely validates the stated theorem: anti-Hermitian matrices exponentiate to unitary ones, and traceless anti-Hermitian matrices exponentiate to special unitary ones.

[11] In fact, this identity holds for all matrices (although we don't need this full generality in this question). There are numerous ways of seeing this, but one efficient way is to write $M = PJP^{-1}$ where J is in Jordan normal form. The problem then comes down to being able to compute the exponential of a Jordan normal form matrix, which is not too bad. As an alternative, observe that diagonalisable matrices are dense in the space of all matrices; therefore, $\det \circ \exp$ and $\exp \circ \mathrm{Tr}$ are continuous functions agreeing on a dense subspace, so agree on the entire space.

On the other hand, if we don't already know the Lie algebras, then the exponential map we have constructed in this question provides 'one direction of the proof' in the construction of the algebras. This is described in detail as follows:

Proposition: The Lie algebras of $U(n)$ and $SU(n)$ are given respectively by:

$$\mathfrak{u}(n) = \{n \times n \text{ anti-Hermitian complex matrices}\},$$

$$\mathfrak{su}(n) = \{\text{traceless } n \times n \text{ anti-Hermitian complex matrices}\}.$$

Proof: Let $g : (-\epsilon, \epsilon) \to U(n)$ be a curve in $U(n)$ passing through the identity at 0, $g(0) = I$. Then:

$$g(t)^\dagger g(t) = I$$

for all $t \in (-\epsilon, \epsilon)$. Taking the derivative of both sides with respect to t, and then setting $t = 0$, we see that:

$$\dot{g}(t)^\dagger g(t) + g(t)^\dagger \dot{g}(t) = 0 \qquad \Rightarrow \qquad \dot{g}(0)^\dagger = -\dot{g}(0).$$

In particular, the tangent to the curve $g(t)$ at the origin is an anti-Hermitian matrix. Thus *any element of the Lie algebra $\mathfrak{u}(n)$ is an anti-Hermitian matrix*.

Conversely, suppose that iH is an anti-Hermitian matrix (where H is Hermitian, say). Then itH is an anti-Hermitian matrix for all $t \in (-\epsilon, \epsilon)$ for any $\epsilon > 0$, so it follows that $\exp(itH) \in U(n)$ for all $t \in (-\epsilon, \epsilon)$, by the results we proved earlier in this question. Consequently, $\exp(itH)$ is a curve in $U(n)$, with derivative at $t = 0$ given by:

$$\frac{d}{dt}\exp(itH)\bigg|_{t=0} = (iH\exp(itH))\bigg|_{t=0} = iH.$$

It follows that *any anti-Hermitian matrix is an element of the Lie algebra $\mathfrak{u}(n)$*. Thus the Lie algebra $\mathfrak{u}(n)$ is precisely the set of $n \times n$ complex anti-Hermitian matrices.

The argument extends to $SU(n)$. In this case, a curve $g : (-\epsilon, \epsilon) \to SU(n)$ not only obeys $g(t)^\dagger g(t) = I$ for all $t \in (-\epsilon, \epsilon)$, giving the condition that $\dot{g}(0)$ is anti-Hermitian as before, but the curve g must additionally obey:

$$\det(g(t)) = 1$$

for all $t \in (-\epsilon, \epsilon)$. We can take the derivative of this equation with respect to t by employing *Jacobi's formula* for the derivative of the determinant of an invertible matrix:

$$0 = \frac{d}{dt}\det(g(t))\bigg|_{t=0} = \left(\det(g(t))\mathrm{Tr}\left(g(t)^{-1}\dot{g}(t)\right)\right)\bigg|_{t=0} = \mathrm{Tr}(\dot{g}(0)).$$

It follows that *any element of the Lie algebra $\mathfrak{su}(n)$ is a traceless anti-Hermitian matrix*.

Conversely, suppose that iH is a traceless anti-Hermitian matrix. Then just as above, we can use the exponential $\exp(itH)$ to provide a curve into $SU(n)$ with tangent iH at the origin. Thus *any traceless anti-Hermitian matrix is an element of the Lie algebra $\mathfrak{su}(n)$*. It follows that the Lie algebra $\mathfrak{su}(n)$ is precisely the set of $n \times n$ complex, traceless, anti-Hermitian matrices. $\square$

In particular, this question fills in a possible gap from lectures in the construction of the Lie algebras $\mathfrak{u}(n)$ and $\mathfrak{su}(n)$.

11. A useful basis for the Lie algebra of $GL(n)$ consists of the n^2 matrices T^{ij} ($1 \leq i, j \leq n$) where $(T^{ij})_{\alpha\beta} = \delta_{i\alpha}\delta_{j\beta}$.[12] Find the structure constants in this basis.

Solution: We can compute the structure constants directly using some index notation (the only difficult here is the slightly odd notation; instead of labelling the Lie algebra generators with a single index, T_a, like we did in Questions 5 and 6, we instead use two indices). The (α, β)th component of the matrix $[T^{ij}, T^{kl}]$ is given by (the summation convention is implied throughout):

$$\begin{aligned}
[T^{ij}, T^{kl}]_{\alpha\beta} &= (T^{ij}T^{kl})_{\alpha\beta} - (T^{kl}T^{ij})_{\alpha\beta} \\
&= T^{ij}_{\alpha\gamma}T^{kl}_{\gamma\beta} - T^{kl}_{\alpha\gamma}T^{ij}_{\gamma\beta} \\
&= \delta_{i\alpha}\delta_{j\gamma}\delta_{k\gamma}\delta_{l\beta} - \delta_{k\alpha}\delta_{l\gamma}\delta_{i\gamma}\delta_{j\beta} \\
&= \delta_{i\alpha}\delta_{jk}\delta_{l\beta} - \delta_{k\alpha}\delta_{il}\delta_{j\beta} \\
&= (\delta_{m\alpha}\delta_{n\beta}) \cdot (\delta_{im}\delta_{jk}\delta_{ln} - \delta_{km}\delta_{il}\delta_{jn}) \\
&= (\delta_{im}\delta_{jk}\delta_{ln} - \delta_{km}\delta_{il}\delta_{jn})\, T^{mn}_{\alpha\beta}.
\end{aligned}$$

Thus the structure constants are given by:

$$f^{mn}{}_{ij,kl} = \delta_{im}\delta_{jk}\delta_{ln} - \delta_{km}\delta_{il}\delta_{jn}.$$

[12] Note that we're not bothering to keep track of upstairs and downstairs indices in this question; it's not important here.

Part III: Symmetries, Fields and Particles
Examples Sheet 2 Solutions

Please send all comments and corrections to mjb318@cam.ac.uk and jmm232@cam.ac.uk.

1. Show that a representation of the dihedral group, $D : D_4 \to GL(2, \mathbb{R})$, can be constructed using the matrices:

$$D(r) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \qquad D(s) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}.$$

(a) Is this a faithful representation of D_4? Is it a reducible representation of D_4?

(b) Consider the subgroup $K_4 = \{e, r^2, s, sr^2\}$ (Klein's *Vierergruppe*) and the corresponding matrices used above. Show that these four matrices constitute a reducible representation of K_4, and identify the invariant subspaces.

Solution: Let us begin by recalling the definition of a group representation:

> **Definition:** A *(linear) group representation* is a triple (G, ρ, V) where G is a group, V is a vector space, and $\rho : G \to GL(V)$ is a group homomorphism. Here, $GL(V)$ denotes the *general linear group* of the vector space V, which is the group of all automorphisms of V (i.e. bijective linear maps $V \to V$). We often abuse language here for brevity. Instead of quoting the triple (G, ρ, V) and giving equal importance to each ingredient of the representation, we often simply refer to the map ρ as the *representation*. Alternatively, if the group G and the map $\rho : G \to GL(V)$ are implied, then we often simply refer to V as the *representation*.[1]
>
> When V is a finite-dimensional vector space over the field $\mathbb{F}$, then choosing a basis for V induces a canonical isomorphism $GL(V) \cong GL(\dim(V), \mathbb{F})$ between $GL(V)$ and the set of invertible matrices over the field $\mathbb{F}$. Thus we can identify each $\rho(g) \in GL(V)$ with some matrix $\tilde{\rho}(g) \in GL(\dim(V), \mathbb{F})$ (but this identification will depend on the basis). Conversely, given a homomorphism $\tilde{\rho} : G \to GL(\dim(V), \mathbb{F})$, we can construct a representation $\rho : G \to GL(V)$ through $\rho(g)(\mathbf{v}) := \tilde{\rho}(g)\mathbf{v}$, where the right hand side simply denotes matrix multiplication with $\mathbf{v}$ written in the appropriate basis.

In particular, to show that $D : D_4 \to GL(2, \mathbb{R})$ is a representation of D_4, all we need to do is to show $D(g_1 g_2) = D(g_1)D(g_2)$ for all $g_1, g_2 \in D_4$.

But hang on a second! The question has only told us the values of D at two of the group elements $r, s \in D_4$. How are we supposed to check $D(sr) = D(s)D(r)$, for example, if we haven't been told the value of $D(sr)$? Since s, r generate the dihedral group, the idea is to *define* $D(s^i r^j)$ using the desired homomorphism property of D. Let us define:

$$D(s^i r^j) := D(s)^i D(r)^j$$

for $0 \le i \le 1, 0 \le j \le 3$. Note we are careful to provide a definition for each $g \in D_4$ only once, using the restrictions $0 \le i \le 1, 0 \le j \le 3$.

[1] This seems confusing at first, but causes little issue in practice. Complaints should be sent to the algebra group in DPMMS.

We claim that this constitutes a valid representation of D_4.

Proposition: The map $D : D_4 \to GL(2, \mathbb{R})$ defined by $D(s^i r^j) = D(s)^i D(r)^j$ with $0 \leq i \leq 1$ and $0 \leq j \leq 3$ defines a representation of D_4.

Proof: Certainly this map is well-defined, since $D(s), D(r)$ are invertible real matrices, so $D(s)^i D(r)^j \in GL(2, \mathbb{R})$ for $0 \leq i \leq 1, 0 \leq j \leq 3$. Furthermore, the restrictions $0 \leq i \leq 1$ and $0 \leq j \leq 3$ ensure that we have defined the image of each element of D_4 only once.

Now suppose that $s^i r^j, s^a r^b \in D_4$ for $0 \leq i, a \leq 1$ and $0 \leq j, b \leq 3$. Then we have:

$$D(s^i r^j) D(s^a r^b) = D(s)^i D(r)^j D(s)^a D(r)^b \tag{$*$}$$

by definition. On the other hand, $s^i r^j s^a r^b = s^i s^a r^{(-1)^a j} r^b = s^{i+a \pmod 2} r^{b+(-1)^a j \pmod 4}$, using the group algebra. Hence we must compare ($*$) to:

$$D(s^i r^j s^a r^b) = D(s)^{i+a \pmod 2} D(r)^{b+(-1)^a j \pmod n}. \tag{$\dagger$}$$

To make the comparison, we note the following relations between the matrices $D(s), D(r)$:

- Trivially, we have $D(s)^2 = I$.
- We have:
$$D(r)^2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \qquad \Rightarrow \qquad D(r)^4 = I.$$
- Finally, note that (using $D(r)$ orthogonal, since it has orthonormal columns):
$$D(r)D(s) = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}, \qquad D(s)D(r)^{-1} = D(s)D(r)^T = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}.$$
  Hence we have $D(r)D(s) = D(s)D(r)^{-1}$.

It follows that we can rewrite the right hand side of ($*$) as:

$$D(s)^i D(r)^j D(s)^a D(r)^b = D(s)^i D(s)^a D(r)^{(-1)^a j} D(r)^b = D(s)^{i+a \pmod 2} D(r)^{b+(-1)^a j \pmod 4},$$

so ($*$) and ($\dagger$) agree, and the homomorphism property is satisfied. $\square$

(a) Now that we have established that $D : D_4 \to GL(2, \mathbb{R})$ is indeed a representation, in this part of the question we establish some of its properties. Recall:

Definition: Let $\rho : G \to GL(V)$ be a representation of the group G. We say that ρ is *faithful* if ρ is injective.

This is a useful property for a representation, because it means every group element gets mapped to its own linear map - thus no information about the group is lost when passing to the linear maps which represent the group elements. Contrast this with the *trivial* representation $\rho : G \to GL(V)$ given by $\rho(g) = \mathrm{id}_V$, where $\mathrm{id}_V : V \to V$ is the identity; in this case, all information about the group is completely washed away, because all group elements are mapped to the same linear map (unless we are in the special case where $G = \{e\}$ is the trivial group - then we lose no information).

We can prove that the given representation $D : D_4 \to GL(2, \mathbb{R})$ is faithful as follows. Suppose that $D(s^i r^j) = D(s^a r^b)$ for $0 \le i, a \le 1$ and $0 \le j, b \le 3$. Then by definition we have:

$$D(s)^i D(r)^j = D(s)^a D(r)^b \qquad \Rightarrow \qquad D(s)^{i-a \;(\mathrm{mod}\ 2)} = D(r)^{b-j \;(\mathrm{mod}\ 4)}.$$

The only distinct powers of $D(s)$ are I and $D(s)$ itself. For $D(r)$, the distinct powers are:

$$I, \qquad D(r), \qquad D(r)^2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad D(r)^3 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Therefore a power of $D(s)$ and a power of $D(r)$ can only coincide if they are both the identity. It follows that $i \equiv a \pmod 2$ and $b \equiv j \pmod 4$. Hence $s^i r^j = s^a r^b$, and the representation is indeed faithful.

Next, we are asked whether the representation of D_4 is reducible. First, let's recall what this means:

Definition: Let (G, ρ, V) be a representation. A *subrepresentation* of (G, ρ, V) is a representation $(G, \tilde{\rho}, W)$ where $W \le V$ is a vector subspace of V and $\tilde{\rho}(g) = \rho(g)|_W$ for all $g \in G$. A subrepresentation is called *trivial* if $W = \{0\}$, and *improper* if $W = V$. A representation is *irreducible* if it has no proper, non-trivial subrepresentations; otherwise, it is *reducible*.

Therefore, to discover whether a representation is reducible or not, we must ask whether it has any subrepresentations. These can be quickly identified by finding *invariant subspaces*, as described in the following proposition:

Proposition: Let (G, ρ, V) be a representation, and let $W \le V$ be a vector subspace of V. Then $(G, \tilde{\rho}, W)$ is a subrepresentation of (G, ρ, V) if and only if W is an *invariant subspace* of ρ, i.e. for all $\mathbf{w} \in W$ and for all $g \in G$, we have:

$$\rho(g)\mathbf{w} \in W.$$

Proof: Suppose that $W \le V$ is invariant. Then for each $g \in G$, we have that $\tilde{\rho}(g) := \rho(g)|_W : W \to W$ is an invertible linear map, since $\rho(g)\mathbf{w} \in W$ for all $\mathbf{w} \in W$. Thus $\tilde{\rho} : G \to GL(W)$ given by $\tilde{\rho}(g) := \rho(g)|_W$ is well-defined. It is obviously a homomorphism since ρ is a homomorphism. Hence, we have a subrepresentation.

Conversely, suppose that $(G, \tilde{\rho}, W)$ is a subrepresentation of (G, ρ, V). Then we clearly require the restriction of ρ to W to be a map from W to W, i.e. we require W to be invariant. $\square$

Now, we can argue that the given representation of D_4 is irreducible for the following simple geometric reason. The only proper, non-trivial subspaces of the representation space $\mathbb{R}^2$ are lines through the origin. But clearly no lines through the origin are fixed by a $\pi/2$ rotation, i.e. $D(r)$, and hence there are no invariant subspaces of the representation. Thus it is irreducible.

(b) First, note that the restriction $D|_{K_4} : K_4 \to GL(2,\mathbb{R})$ is clearly a representation, since it is the restriction of a homomorphism to a subgroup (so in particular is still a homomorphism from that subgroup). This time however, the representation is clearly reducible, since the matrices in the image of $D|_{K_4}$ correspond to a rotation by π, and reflections in the horizontal and vertical. Therefore, the horizontal and vertical are both invariant subspaces; we have:

$$\mathbb{R}^2 = \langle \begin{pmatrix} 1 \\ 0 \end{pmatrix} \rangle \oplus \langle \begin{pmatrix} 0 \\ 1 \end{pmatrix} \rangle,$$

as a decomposition of $\mathbb{R}^2$ into irreducibles, where $\langle \cdot \rangle$ denotes the span (both invariant subspaces are one-dimensional so cannot be reduced further).

❉ **Comments:** In the first part of this question, we argued that we could construct a representation from the matrices $D(r), D(s)$ using the homomorphism property for a representation. The main step was checking that the group relations $r^4 = e$, $s^2 = e$ and $rs = sr^{-1}$ held when translated to the representing matrices.

This idea can be extended to any group specified by a group presentation (i.e. in terms of generators and relations); we describe this as follows, but the proof is omitted because it is difficult and beyond the scope of the course:

Proposition: Let $G = \langle S|R\rangle$ be a group, specified by a group presentation, with generating set S and set of relations $R = \{r_\alpha = e\}$ (by appropriate rearrangement, any relation can be written in the form $r_\alpha = e$). Furthermore, since S is a generating set, given any relation $r_\alpha = e$, we can write $r_\alpha = s_{\alpha,1}...s_{\alpha,n_\alpha}$ where $s_{\alpha,i} \in S$ and $n_\alpha \in \mathbb{Z}$.

Suppose we are given any other group H, and some function $f : S \to H$. Then there exists a unique group homomorphism $\theta : G \to H$ extending f (i.e. such that $\theta|_S = f$) if and only if $f(s_{\alpha,1})...f(s_{\alpha,n_\alpha}) = e$ for all relations $e = r_\alpha = s_{\alpha,1}...s_{\alpha,n_\alpha}$ in R.

Proof: Suppose that $\theta : G \to H$ is such a homomorphism. Then for every relation $e = r_\alpha = s_{\alpha,1}...s_{\alpha,n_\alpha}$ in R, we must have:

$$e = \theta(e) = \theta(r_\alpha) = \theta(s_{\alpha,1})...\theta(s_{\alpha,n_\alpha}).$$

But $\theta|_S = f$, since θ extends f. Thus we must have $e = f(s_{\alpha,1})...f(s_{\alpha,n_\alpha})$ as required.

The converse is tricky, and beyond the scope of the course. □

2. The adjoint representation of the Lie group $SU(2)$ is defined to be the map $\mathrm{Ad} : SU(2) \to GL(\mathfrak{su}(2))$ given by:

$$\mathrm{Ad}_A(X) = AXA^\dagger \qquad (*)$$

for all $A \in SU(2), X \in \mathfrak{su}(2)$.

(a) Show that Ad is indeed a group representation. This will require checking: (i) for each $A \in SU(2)$, we have that Ad_A is an automorphism of $\mathfrak{su}(2)$; (ii) given $A, B \in SU(2)$, we have $\mathrm{Ad}_{AB} = \mathrm{Ad}_A \circ \mathrm{Ad}_B$.

(b) By writing $A = I + Y + O(Y^2)$ in $(*)$, construct the associated adjoint representation $\mathrm{ad} : \mathfrak{su}(2) \to \mathfrak{gl}(\mathfrak{su}(2))$, where $\mathfrak{gl}(\mathfrak{su}(2))$ is the space of linear maps $\mathfrak{su}(2) \to \mathfrak{su}(2)$ of the Lie algebra $\mathfrak{su}(2)$. Verify that your proposed representation of $\mathfrak{su}(2)$ indeed constitutes a Lie algebra representation.

Solution: (a) We do as the question advises us, and check (i) and (ii) separately.

(i) **The map is well-defined.** We need to check that for each $A \in SU(2)$, Ad_A is an automorphism of $\mathfrak{su}(2)$ (recall that $\mathfrak{su}(2)$ is the set of 2×2 complex anti-Hermitian traceless matrices). First, note that for all $A \in SU(2)$ and $X \in \mathfrak{su}(2)$, we have:

$$\mathrm{Ad}_A(X)^\dagger = (AXA^\dagger)^\dagger = AX^\dagger A^\dagger = -AXA^\dagger = -\mathrm{Ad}_A(X),$$

hence $\mathrm{Ad}_A(X)$ is anti-Hermitian. Also, we have:

$$\mathrm{Tr}(\mathrm{Ad}_A(X)) = \mathrm{Tr}(AXA^\dagger) = \mathrm{Tr}(A^\dagger AX) = \mathrm{Tr}(X) = 0,$$

by the cyclicity of the trace. It follows that $\mathrm{Ad}_A(X) \in \mathfrak{su}(2)$; therefore Ad_A is always a map $\mathfrak{su}(2) \to \mathfrak{su}(2)$.

Next, note that Ad_A is linear, since:

$$\mathrm{Ad}_A(\lambda X + \mu Y) = A(\lambda X + \mu Y)A^\dagger = \lambda AXA^\dagger + \mu AYA^\dagger = \lambda \mathrm{Ad}_A(X) + \mu \mathrm{Ad}_A(Y)$$

for all $\lambda, \mu \in \mathbb{R}$ and $X, Y \in \mathfrak{su}(2)$. Finally, note that Ad_A is a bijection, since it has inverse $\mathrm{Ad}_{A^\dagger}$:

$$\mathrm{Ad}_A \circ \mathrm{Ad}_{A^\dagger}(X) = \mathrm{Ad}_A(A^\dagger XA) = AA^\dagger XAA^\dagger = X$$

for all $X \in \mathfrak{su}(2)$, and similarly $\mathrm{Ad}_{A^\dagger} \circ \mathrm{Ad}_A = \mathrm{id}_{\mathfrak{su}(2)}$.

(ii) **The map is a homomorphism.** For all $A, B \in SU(2)$ and $X \in \mathfrak{su}(2)$, we have:

$$\mathrm{Ad}_A \circ \mathrm{Ad}_B(X) = \mathrm{Ad}_A(BXB^\dagger) = ABXB^\dagger A^\dagger = (AB)X(AB)^\dagger = \mathrm{Ad}_{AB}(X),$$

hence $\mathrm{Ad}_A \circ \mathrm{Ad}_B = \mathrm{Ad}_{AB}$ as required.

Therefore, the adjoint representation of $SU(2)$ on $\mathfrak{su}(2)$ is indeed a group representation, as required.

(b) Setting $A = I + Y + O(Y^2)$, we see that the 'linearisation' of the group representation in (a) is:

$$\mathrm{Ad}_A(X) = AXA^\dagger = (I + Y + O(Y^2))X(I + Y + O(Y^2))^\dagger = X + YX + XY^\dagger + O(Y^2).$$

Furthermore, the requirement that $AA^\dagger = I$ implies that $I = (I+Y+O(Y^2))(I+Y+O(Y^2))^\dagger = I+Y+Y^\dagger+O(Y^2)$, so that $Y = -Y^\dagger$. In particular, we have:

$$\mathrm{Ad}_A(X) = X + YX - XY + O(Y^2) = X + [Y, X] + O(Y^2),$$

which suggests that the Lie algebra representation corresponding to the adjoint representation of the Lie group should be $\mathrm{ad} : \mathfrak{su}(2) \to \mathfrak{gl}(\mathfrak{su}(2))$ given by:

$$\mathrm{ad}_Y(X) = [Y, X],$$

where ad_Y is the image of $Y \in \mathfrak{su}(2)$ under ad.

It remains to verify that this indeed constitutes a Lie algebra representation. Recall from lectures:

> **Definition:** A *Lie algebra representation* is a triple $(\mathfrak{g}, \rho, V)$ comprising a Lie algebra $\mathfrak{g}$, a vector space V, and a linear map $\rho : \mathfrak{g} \to \mathfrak{gl}(V)$ (where $\mathfrak{gl}(V)$ is the set of all linear maps $V \to V$) satisfying $\rho([X, Y]) = [\rho(X), \rho(Y)]$ for all $X, Y \in \mathfrak{g}$. Similar abuse of language applies in this case as to the case of a Lie group representation (e.g. if $\rho, \mathfrak{g}$ are implied, we often refer to V as the representation). Similarly when V is finite-dimensional over the field $\mathbb{F}$, we may pick a basis to identify $\mathfrak{gl}(V) \cong \mathfrak{gl}(\dim(V), \mathbb{F})$, where $\mathfrak{gl}(\dim(V), \mathbb{F})$ is the set of $\dim(V) \times \dim(V)$ matrices over the field $\mathbb{F}$.

Let us check that ad_Y satisfies the definition:

(i) **The map is well-defined.** We begin by checking that $\mathrm{ad} : \mathfrak{su}(2) \to \mathfrak{gl}(\mathfrak{su}(2))$ is a well-defined map. For each $X, Y \in \mathfrak{g}$, we have that:

$$\mathrm{ad}_Y(X) = [Y, X] \in \mathfrak{su}(2),$$

since $\mathfrak{su}(2)$ is closed under the Lie bracket. Therefore, for each $Y \in \mathfrak{g}$, we have that ad_Y is a map $\mathfrak{su}(2) \to \mathfrak{su}(2)$. It is also a linear map since:

$$\mathrm{ad}_Y(\lambda_1 X_1 + \lambda_2 X_2) = [Y, \lambda_1 X_1 + \lambda_2 X_2] = \lambda_1 [Y, X_1] + \lambda_2 [Y, X_2] + \lambda_1 \mathrm{ad}_Y(X_1) + \lambda_2 \mathrm{ad}_Y(X_2)$$

by linearity of the Lie bracket. Therefore for each $Y \in \mathfrak{g}$, we have $\mathrm{ad}_Y \in \mathfrak{gl}(\mathfrak{su}(2))$.

(ii) **The map is linear.** We need to check that $\mathrm{ad} : \mathfrak{su}(2) \to \mathfrak{gl}(\mathfrak{su}(2))$ is linear. This follows from the linearity of the Lie bracket, since:

$$\mathrm{ad}_{\lambda_1 Y_1 + \lambda_2 Y_2}(X) = [\lambda_1 Y_1 + \lambda_2 Y_2, X] = \lambda_1 [Y_1, X] + \lambda_2 [Y_2, X] = \lambda_1 \mathrm{ad}_{Y_1}(X) + \lambda_2 \mathrm{ad}_{Y_2}(X).$$

(iii) **The map preserves the Lie bracket.** This follows from the Jacobi identity. Given $Y, Z \in \mathfrak{g}$ and $X \in \mathfrak{su}(2)$, we have:

$$\begin{aligned} \mathrm{ad}_{[Y,Z]}(X) &= [[Y, Z], X] = -[[Z, X], Y] - [[X, Y], Z] = [Y, [Z, X]] - [Z, [Y, X]] \\ &= \mathrm{ad}_Y \circ \mathrm{ad}_Z(X) - \mathrm{ad}_Z \circ \mathrm{ad}_Y(X) = [\mathrm{ad}_Y, \mathrm{ad}_Z](X). \end{aligned}$$

Hence, $\mathrm{ad} : \mathfrak{su}(2) \to \mathfrak{gl}(\mathfrak{su}(2))$ is indeed a Lie algebra representation.

❊ **Comments:** In this question, we construct a Lie algebra representation from a Lie group representation; this is always possible, as you saw in lectures. Suppose that $D : G \to GL(n, \mathbb{F})$ is a Lie group representation, where G is some matrix Lie group; then, we define the Lie algebra representation $d : \mathfrak{g} \to \mathfrak{gl}(n, \mathbb{F})$ via:

$$d(X) = \frac{d}{dt} D\left(e^{tX}\right)\bigg|_{t=0}, \tag{$*$}$$

where e^{tX} is the matrix exponential.

3. (a) If d_1 and d_2 are representations of a Lie algebra $\mathfrak{g}$, show that $d_1 \oplus d_2$ is too. Via the exponential map, show that $D_1 \oplus D_2$, defined by $(D_1 \oplus D_2)(\exp(X)) = \exp((d_1 \oplus d_2)(X)), X \in \mathfrak{g}$, is a representation of G. You may assume that each D_i, defined by $D_i(\exp(X)) = \exp(d_i(X))$ for $X \in \mathfrak{g}$, constitutes a well-defined representation of the Lie group G.

(b) Prove that the tensor product $d_1 \otimes d_2$ is a representation of $\mathfrak{g}$. Exponentiate to show how $D_1 \otimes D_2$ is a representation of G.

Solution: (a) Suppose that $d_1 : \mathfrak{g} \to \mathfrak{gl}(V)$, $d_2 : \mathfrak{g} \to \mathfrak{gl}(W)$ are representations on vector spaces V, W respectively. Recall that we define $d_1 \oplus d_2 : \mathfrak{g} \to \mathfrak{gl}(V \oplus W)$ by:

$$(d_1 \oplus d_2)(X)(v \oplus w) = d_1(X)v \oplus d_2(X)w,$$

for all $v \in V$ and $w \in W$. To check this is a representation, we need to check three things:

- **The map is well-defined.** In particular, we need to check that $(d_1 \oplus d_2)(X) \in \mathfrak{gl}(V \oplus W)$ for all $X \in \mathfrak{g}$; in other words, we need to verify that $(d_1 \oplus d_2)(X)$ is a linear map on $V \oplus W$ for each $X \in \mathfrak{g}$. This is a straightforward check:

$$\begin{aligned}(d_1 \oplus d_2)(X)(\lambda(v \oplus w) + \mu(v' \oplus w')) &= (d_1 \oplus d_2)(X)((\lambda v + \mu v') \oplus (\lambda w + \mu w'))\\ &= d_1(X)(\lambda v + \mu v') \oplus d_2(X)(\lambda w + \mu w')\\ &= (\lambda d_1(X)v + \mu d_1(X)v') \oplus (\lambda d_2(X)w + \mu d_2(X)w')\\ &= \lambda(d_1(X)v \oplus d_2(X)w) + \mu(d_1(X)v' \oplus d_2(X)w')\\ &= \lambda(d_1 \oplus d_2)(X)(v \oplus w) + \mu(d_1 \oplus d_2)(X)(v' \oplus w').\end{aligned}$$

  Hence $(d_1 \oplus d_2)(X)$ is indeed a linear map on $V \oplus W$ for each $X \in \mathfrak{g}$.

- **The map is linear.** Next, we need to check that $d_1 \oplus d_2 : \mathfrak{g} \to \mathfrak{gl}(V \oplus W)$ is a linear map from the Lie algebra to the set of linear maps on $V \oplus W$. For each $v \oplus w \in V \oplus W$, and for all $X, Y \in \mathfrak{g}$, and all scalars λ, μ we have:

$$\begin{aligned}(d_1 \oplus d_2)(\lambda X + \mu Y)(v \oplus w) &= d_1(\lambda X + \mu Y)v \oplus d_2(\lambda X + \mu Y)w\\ &= (\lambda d_1(X)v + \mu d_1(Y)v) \oplus (\lambda d_2(X)w + \mu d_2(Y)w)\\ &= \lambda(d_1(X)v \oplus d_2(X)w) + \mu(d_1(Y)v \oplus d_2(Y)w)\\ &= \lambda(d_1 \oplus d_2)(X)v \oplus w + \mu(d_1 \oplus d_2)(Y)v \oplus w,\end{aligned}$$

  hence we have shown that $(d_1 \oplus d_2)(\lambda X + \mu Y) = \lambda(d_1 \oplus d_2)(X) + \mu(d_1 \oplus d_2)(Y)$ for all $X, Y \in \mathfrak{g}$ and all scalars λ, μ.

- **The map preserves the Lie bracket.** Finally, we need to show that for any $X, Y \in \mathfrak{g}$, we have $(d_1 \oplus d_2)([X,Y]) = [(d_1 \oplus d_2)(X), (d_1 \oplus d_2)(Y)]$. We have for all $v \oplus w \in V \oplus W$:

$$\begin{aligned}(d_1 \oplus d_2)([X,Y])v \oplus w &= d_1([X,Y])v \oplus d_2([X,Y])w \\ &= [d_1(X), d_1(Y)]v \oplus [d_2(X), d_2(Y)]w \\ &= (d_1(X)d_1(Y)v - d_1(Y)d_1(X)v) \oplus (d_2(X)d_2(Y)w - d_2(Y)d_2(X)w) \\ &= d_1(X)d_1(Y)v \oplus d_2(X)d_2(Y)w - d_1(Y)d_1(X)v \oplus d_2(Y)d_2(X)w \\ &= \big((d_1 \oplus d_2)(X)(d_1 \oplus d_2)(Y) - (d_1 \oplus d_2)(Y)(d_1 \oplus d_2)(X)\big)v \oplus w \\ &= [(d_1 \oplus d_2)(X), (d_1 \oplus d_2)(Y)]v \oplus w.\end{aligned}$$

  Since this holds for all $v \oplus w \in V \oplus W$, the result follows.

Hence $d_1 \oplus d_2$ indeed constitutes a representation of $\mathfrak{g}$ on $V \oplus W$, as required.

Aside: exponentiation of Lie algebra representations

Next, we are asked to discuss the *exponentiation* of this Lie algebra representation. Before doing the question, it is useful to recall some key facts about matrix exponentiation.

For a matrix Lie algebra and a matrix Lie group, recall that the *matrix exponential* provides a map:

$$\exp : \mathfrak{g} \to G.$$

Now, whilst every element of the Lie algebra $X \in \mathfrak{g}$ maps to some element of the Lie group $\exp(X) \in \mathfrak{g}$ under the matrix exponential, there are two key deficiencies of this map:

(1) **The map might not be a surjection.** In particular, the matrix exponential is a continuous operation and the exponential of the zero matrix is the identity, $\exp(0) = I \in G$. It follows that the matrix exponential can map *only* into the connected component of G containing the identity. For example,

$$\exp : \mathfrak{o}(n) \to O(n)$$

will in fact restrict to a map $\exp : \mathfrak{o}(n) \to SO(n)$, so that *reflections in the Lie group $O(n)$ are inaccessible by exponentiation*.

(2) **The map might not be an injection.** We saw on Sheet 1, Question 9, for example that:

$$\exp(0) = \exp(2\pi i \sigma_1),$$

where σ_1 is the first Pauli matrix. Whilst 0 and $2\pi i \sigma_1$ are *distinct* elements of the Lie algebra $\mathfrak{su}(2)$, they both map to the *same* element of the Lie group $SU(2)$ under the matrix exponential $\exp : \mathfrak{su}(2) \to SU(2)$.

These problems can cause issues when we try to use the matrix exponential to convert a Lie algebra representation into a Lie group representation. Suppose that we are given a Lie algebra representation $d : \mathfrak{g} \to \mathfrak{gl}(V)$. The 'naïve' approach in constructing a Lie group representation from this Lie algebra representation is simply to say: 'for each $g \in G$, let $X \in \mathfrak{g}$ be an element of the Lie algebra such that $g = \exp(X)$, and set:

$$D(g) = \exp(d(X)).\text{'} \qquad (*)$$

This fails on two fronts, related to points (1) and (2) above.

(1) Since $\exp : \mathfrak{g} \to G$ might not be surjective, there may not exist some $X \in \mathfrak{g}$ such that $g = \exp(X)$ - in particular, our definition $(*)$ might not work for all elements of the Lie group G. To fix this, we can demand that G is a Lie group such that $\exp : \mathfrak{g} \to G$ is surjective.[2]

(2) Another problem arises because the exponential map $\exp : \mathfrak{g} \to G$ is not necessarily *injective* - there might be two different Lie algebra elements $X_1, X_2 \in \mathfrak{g}$ such that $g = \exp(X_1)$ *and* $g = \exp(X_2)$. Then our definition $(*)$ requires:

$$D(g) = \exp(d(X_1)) \quad \text{and} \quad D(g) = \exp(d(X_2)) \qquad \Rightarrow \qquad \exp(d(X_1)) = \exp(d(X_2)).$$

There is no guarantee that this equation holds! It follows that $(*)$ might be *ill-defined*; we could get two different answers for $D(g)$ depending on the representative of the Lie algebra $X \in \mathfrak{g}$ we chose such that $g = \exp(X)$. This problem is a lot more difficult to fix in general.

However, all is not lost - there is a general theorem which gives a sufficient condition ensuring our naïve definition $(*)$ works:

Theorem: Suppose that G is a *simply-connected* Lie group, and suppose that $\exp : \mathfrak{g} \to G$ is a surjection.[3] Then given any representation $d : \mathfrak{g} \to \mathfrak{gl}(V)$, we can construct a well-defined representation $D : G \to GL(V)$ via:

$$D(g) = \exp(d(X)),$$

where $X \in \mathfrak{g}$ is any Lie algebra element such that $g = \exp(X)$.

Proof: The proof is very difficult and beyond the scope of the course. $\square$

With this aside out of the way, we can return to our study of the direct sum of representations. Let us suppose that the representations $d_1 : \mathfrak{g} \to \mathfrak{gl}(V)$, $d_2 : \mathfrak{g} \to \mathfrak{gl}(W)$ given in the question are such that the maps $D_1 : G \to GL(V)$, $D_2 : G \to GL(W)$, given by:

$$D_1(\exp(X)) = \exp(d_1(X)), \qquad D_2(\exp(X)) = \exp(d_2(X)), \tag{$\dagger$}$$

constitute well-defined representations of the Lie group G (e.g. by requiring G to be simply-connected and $\exp : \mathfrak{g} \to G$ to be a surjection). Consider the map $D_1 \oplus D_2 : G \to GL(V \oplus W)$ defined by:

$$(D_1 \oplus D_2)(\exp(X))v \oplus w = D_1(\exp(X))v \oplus D_2(\exp(X))w = \exp(d_1(X))v \oplus \exp(d_2(X))w, \tag{$\dagger\dagger$}$$

for all $v \oplus w \in V \oplus W$. To show that this is a Lie group representation, we must check two properties:

- **The map is well-defined.** First, note that since the definition ($\dagger$) gives two well-defined definitions of a representation, we may assume that we are working in the case that any element $g \in G$ can be written in the form $g = \exp(X)$ for some $X \in \mathfrak{g}$. In particular, the definition ($\dagger\dagger$) can be applied to any element in the Lie group.

[2] A sufficient condition is that G is *connected* and *compact*.

[3] In fact, one can eliminate the need for the surjectivity here - we only really need that G is simply-connected. In general, it is possible to show that if G is connected, then any element $g \in G$ can be written as $g = \exp(X_1)...\exp(X_n)$ for some Lie algebra elements $X_1, ..., X_n$, even if g is not in the image of the exponential. The representation can then be defined by $D(g) = \exp(d(X))$ for g in the image of the exponential map, and $D(g) = D(\exp(X_1))...D(\exp(X_n))$ for general $g = \exp(X_1)...\exp(X_n)$.

Next, suppose that $\exp(X) = \exp(Y)$ for $X, Y \in \mathfrak{g}$ (indeed, X, Y can be distinct because the exponential map can fail to be injective). Then we have:

$$\begin{aligned}(D_1 \oplus D_2)(\exp(X)) &= D_1(\exp(X)) \oplus D_2(\exp(X)) \\ &= D_1(\exp(Y)) \oplus D_2(\exp(Y)) && (D_1, D_2 \text{ well-defined}) \\ &= (D_1 \oplus D_2)(\exp(Y)).\end{aligned}$$

Thus $D_1 \oplus D_2$ maps $\exp(X), \exp(Y)$ to the same function when they are equal.

Finally, we must check that the codomain is correct; that is, we must check that $(D_1 \oplus D_2)(g)$ is a linear map on $V \oplus W$ for all $g \in G$. We have:

$$\begin{aligned}(D_1 \oplus D_2)(g)(\lambda v \oplus w + \mu v' \oplus w') &= (D_1 \oplus D_2)(g)((\lambda v + \mu v') \oplus (\lambda w + \mu w')) \\ &= D_1(g)(\lambda v + \mu v') \oplus D_2(g)(\lambda w + \mu w') \\ &= (\lambda D_1(g)v + \mu D_1(g)v') \oplus (\lambda D_2(g)w + \mu D_2(g)w') \quad (D_1, D_2 \text{ well-defined}) \\ &= \lambda D_1(g)v \oplus D_2(g)w + \mu D_1(g)v' \oplus D_2(g)w' \\ &= \lambda(D_1 \oplus D_2)(g)v \oplus w + \mu(D_1 \oplus D_2)(g)v' \oplus w'.\end{aligned}$$

It follows that $(D_1 \oplus D_2)(g)$ is indeed a linear map for each $g \in G$.

- **The map is a group homomorphism.** Now we must check that $(D_1 \oplus D_2)(gh) = (D_1 \oplus D_2)(g)(D_1 \oplus D_2)(h)$ for all $g, h \in G$. We have:

$$\begin{aligned}(D_1 \oplus D_2)(gh) &= D_1(gh) \oplus D_2(gh) && \text{(definition)} \\ &= D_1(g)D_1(h) \oplus D_2(g)D_2(h) && (D_1, D_2 \text{ are representations}) \\ &= (D_1(g) \oplus D_2(g))(D_1(h) \oplus D_2(h)) \\ &= (D_1 \oplus D_2)(g)(D_1 \oplus D_2)(h),\end{aligned}$$

hence it is a homomorphism as required.

Hence, on the assumption that D_1, D_2 give well-defined representations of G on V, W respectively, it follows that $D_1 \oplus D_2$ gives a well-defined representation of G on $V \oplus W$.

(b) Things are quite similar in the case of tensor product representations. Suppose that $d_1 : \mathfrak{g} \to \mathfrak{gl}(V)$, $d_2 : \mathfrak{g} \to \mathfrak{gl}(W)$ are representations on vector spaces V, W respectively. Recall that we define $d_1 \otimes d_2 : \mathfrak{g} \to \mathfrak{gl}(V \otimes W)$ by:

$$(d_1 \otimes d_2)(X)(v \otimes w) = d_1(X)v \otimes w + v \otimes d_2(X)w,$$

for all elements of the form $v \otimes w \in V \otimes W$, and then extend by linearity to all of $V \otimes W$. To check this is a representation, we need to check three things:

- **The map is well-defined.** In particular, we need to check that $(d_1 \otimes d_2)(X) \in \mathfrak{gl}(V \otimes W)$ for all $X \in \mathfrak{g}$; in other words, we need to verify that $(d_1 \otimes d_2)(X)$ is a linear map on $V \otimes W$ for each $X \in \mathfrak{g}$. This is trivial, since we define $(d_1 \otimes d_2)(X)$ as a linear extension from its action on elements of the form $v \otimes w$ in the tensor product, and all elements in the tensor product can be written as linear combinations of elements of this form.

- **The map is linear.** Next, we need to check that $d_1 \otimes d_2 : \mathfrak{g} \to \mathfrak{gl}(V \otimes W)$ is a linear map from the Lie algebra to the set of linear maps on $V \otimes W$. For each $v \otimes w \in V \oplus W$, and for all $X, Y \in \mathfrak{g}$, and for all scalars λ, μ, we have:

$$\begin{aligned}(d_1 \otimes d_2)(\lambda X + \mu Y)(v \otimes w) &= d_1(\lambda X + \mu Y)v \otimes w + v \otimes d_2(\lambda X + \mu Y)w \\ &= (\lambda d_1(X)v + \mu d_1(Y)v) \otimes w + v \otimes (\lambda d_2(X)w + \mu d_2(Y)w) \\ &= \lambda(d_1(X)v \otimes w + v \otimes d_2(X)w) + \mu(d_1(Y)v \otimes w + v \otimes d_2(Y)w) \\ &= \lambda(d_1 \otimes d_2)(X)v \otimes w + \mu(d_1 \otimes d_2)(Y)v \otimes w.\end{aligned}$$

  Hence the linear maps $(d_1 \otimes d_2)(\lambda X + \mu Y)$ and $\lambda(d_1 \otimes d_2)(X) + \mu(d_1 \otimes d_2)(Y)$ agree on all elements of the form $v \otimes w$, and hence by linear extension, agree on all $V \otimes W$. Thus we have

$$(d_1 \otimes d_2)(\lambda X + \mu Y) = \lambda(d_1 \otimes d_2)(X) + \mu(d_1 \otimes d_2)(Y)$$

  for all $X, Y \in \mathfrak{g}$ and all scalars λ, μ as required.

- **The map preserves the Lie bracket.** Finally, we need to show that for any $X, Y \in \mathfrak{g}$, we have $(d_1 \otimes d_2)([X, Y]) = [(d_1 \otimes d_2)(X), (d_1 \otimes d_2)(Y)]$. We have for all elements of the form $v \otimes w \in V \otimes W$:

$$\begin{aligned}(d_1 \otimes d_2)([X, Y])v \otimes w &= d_1([X, Y])v \otimes w + v \otimes d_2([X, Y])w \\ &= [d_1(X), d_1(Y)]v \otimes w + v \otimes [d_2(X), d_2(Y)]w \\ &= (d_1(X)d_1(Y)v - d_1(Y)d_1(X)v) \otimes w + v \otimes (d_2(X)d_2(Y)w - d_2(Y)d_2(X)w) \\ &= d_1(X)d_1(Y)v \otimes w + v \otimes d_2(X)d_2(Y)w - d_1(Y)d_1(X)v \otimes w - v \otimes d_2(Y)d_2(X)w \\ &= (d_1 \otimes d_2)(X)(d_1 \otimes d_2)(Y)v \otimes w - (d_1 \otimes d_2)(Y)(d_1 \otimes d_2)(X)v \otimes w \\ &= [(d_1 \otimes d_2)(X), (d_1 \otimes d_2)(Y)]v \otimes w.\end{aligned}$$

  Since this holds for all elements of the form $v \otimes w \in V \otimes W$, by linear extension to all of $V \otimes W$ we have:

$$(d_1 \otimes d_2)([X, Y]) = [(d_1 \otimes d_2)(X), (d_1 \otimes d_2)(Y)].$$

Hence $d_1 \otimes d_2$ indeed constitutes a representation of $\mathfrak{g}$ on $V \otimes W$, as required.

To finish off, we need to show that $(D_1 \otimes D_2)(\exp(X)) = \exp((d_1 \otimes d_2)(X))$ constitutes a representation of the Lie group G, provided the representations $D_1(\exp(X)) = \exp(d_1(X))$ and $D_2(\exp(X)) = \exp(d_2(X))$ are well-defined (e.g. by assuming that the exponential is surjective and the group G is simply-connected). We begin by noting that we can write this representation as:

$$\begin{aligned}(D_1 \otimes D_2)(\exp(X)) &= \exp((d_1 \otimes d_2)(X)) && \text{(definition)}\\ &= \exp(d_1(X) \otimes I + I \otimes d_2(X)) && \text{(definition of } d_1 \otimes d_2\text{)}\\ &= \exp(d_1(X) \otimes I)\exp(I \otimes d_2(X)) && (d_1(X) \otimes I, I \otimes d_2(X) \text{ commute})\\ &= (D_1(\exp(X)) \otimes I)(I \otimes D_2(\exp(X))),\end{aligned}$$

where $A \otimes B$ acts on the tensor product space by $(A \otimes B)(v \otimes w) = Av \otimes Bw$ for all elements of the form $v \otimes w \in V \otimes W$, together with the appropriate extension by linearity.

We can use this form to check the two important properties:

- **The map is well-defined.** Since $D_i(\exp(X)) = \exp(d_i(X))$ is well-defined for $i = 1, 2$, we must be able to write any $g \in G$ in the form $g = \exp(X)$ for some $X \in \mathfrak{g}$, so the definition $(D_1 \otimes D_2)(\exp(X)) = \exp((d_1 \otimes d_2)(X))$ makes sense everywhere on the domain G.

  Next, suppose that $\exp(X) = \exp(Y)$ for $X, Y \in \mathfrak{g}$. Then we have:

  $$\begin{aligned}(D_1 \otimes D_2)(\exp(X)) &= (D_1(\exp(X)) \otimes I)(I \otimes D_2(\exp(X))) && \text{(above work)}\\ &= (D_1(\exp(Y)) \otimes I)(I \otimes D_2(\exp(Y))) && (D_1, D_2 \text{ well-defined})\\ &= (D_1 \otimes D_2)(\exp(Y)) && \text{(above work)}\end{aligned}$$

  Thus $D_1 \otimes D_2$ maps $\exp(X), \exp(Y)$ to the same function when they are equal.

  It remains to show that $D_1 \otimes D_2$ has the correct codomain, i.e. $(D_1 \otimes D_2)(\exp(X))$ is a linear map on $V \otimes W$. This is clear from the form $(D_1 \otimes D_2)(\exp(X)) = (D_1(\exp(X)) \otimes I)(I \otimes D_2(\exp(X)))$.

- **The map is a group homomorphism.** It remains to check that $(D_1 \otimes D_2)(gh) = (D_1 \otimes D_2)(g)(D_1 \otimes D_2)(h)$ for all $g, h \in G$. We have:

  $$\begin{aligned}(D_1 \otimes D_2)(gh) &= (D_1(gh) \otimes I)(I \otimes D_2(gh))\\ &= (D_1(g)D_1(h) \otimes I)(I \otimes D_2(g)D_2(h)) && (D_1, D_2 \text{ are representations})\\ &= (D_1(g) \otimes I)(D_1(h) \otimes I)(I \otimes D_2(g))(I \otimes D_2(h))\\ &= (D_1(g) \otimes I)(I \otimes D_2(g))(D_1(h) \otimes I)(I \otimes D_2(h)) && (D_1(g) \otimes I, I \otimes D_2(g) \text{ commute})\\ &= (D_1 \otimes D_2)(g)(D_1 \otimes D_2)(h),\end{aligned}$$

  and so we're done.

Hence, on the assumption that D_1, D_2 give well-defined representations of G on V, W respectively, it follows that $D_1 \otimes D_2$ gives a well-defined representation of G on $V \otimes W$.

4. Let D be a finite-dimensional representation of G acting on V, and let $(\cdot, \cdot)$ be an inner product on V invariant under G, i.e.

$$(D(g)u, D(g)v) = (u, v), \qquad \text{for all } \; u, v \in V, g \in G.$$

We say that D is a *unitary* representation in this case.

(a) Let W be an invariant subspace of V. Show that $W_\perp$, the orthogonal complement of W in V, is also invariant.

(b) Deduce that D is completely reducible.

Solution: (a) Recall that we can decompose the vector space V via $V = W \oplus W_\perp$, where $W_\perp$ is the orthogonal complement to the invariant subspace W of V. Let $w \in W$ and let $w_\perp \in W_\perp$. Then for any $g \in G$, we have:

$$\begin{aligned} 0 &= (D(g^{-1})w, w_\perp) && (\text{since } W \text{ is invariant, } D(g^{-1})w \in W) \\ &= (D(g)^{-1}w, w_\perp) && (\text{representations are homomorphisms}) \\ &= (D(g)^\dagger w, w_\perp) && (\text{the representation is unitary}) \\ &= (w, D(g)w_\perp). \end{aligned}$$

We conclude that $D(g)w_\perp \in W_\perp$. It follows that $W_\perp$ is an invariant subspace of V, as required. □

(b) If D is irreducible, then we're done. Otherwise there exists a (proper, non-trivial) invariant subspace $W \leq V$, with $0 < \dim(W) < \dim(V)$. We showed in (a) that the orthogonal complement $W_\perp$ of W in V is also an invariant subspace of V; furthermore, counting dimensions we must have $0 < \dim(W_\perp) < \dim(V)$ so that $W_\perp$ is proper and non-trivial.

In particular, we see that we can write:

$$V = W_1 \oplus W_2,$$

where $W_1 = W, W_2 = W_\perp$ are proper, non-trivial invariant subspaces of V. If W_1, W_2 are irreducible, then we're done. Otherwise, repeat the procedure applied to the reducible subspaces amongst W_1, W_2. Eventually we must reach irreducibles because the dimension of the subspaces decreases each time, and we assume that we are working with a finite-dimensional representation. It follows that D is completely reducible, as required. □

Comments: In fact, we can use the result in this question to deduce a more general result, that all finite-dimensional representations of a finite group are completely reducible. The proof is as follows.

Maschke's Theorem: All finite-dimensional representations of a finite group are completely reducible.

Proof: Let $D : G \to GL(V)$ be a finite-dimensional representation of the finite group G on the vector space V. Since V is assumed to be finite-dimensional, it is isomorphic to $\mathbb{F}^r$ for some r (where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$ depending on whether we are using a real or complex vector space). In particular, the vector space V can be given some arbitrary inner product structure $(\cdot, \cdot) : V \times V \to \mathbb{F}$, by identifying V with $\mathbb{F}^r$. Without loss of generality, let's take $\mathbb{F} = \mathbb{C}$ (get rid of all conjugates if you want to reduce to $\mathbb{R}$). The idea of the rest of the proof is to use this inner product to define a new inner product which makes D unitary with respect to the new inner product.

Let's define the new inner product via:

$$[v, w] := \frac{1}{|G|} \sum_{g \in G} (D(g)v, D(g)w).$$

This is well-defined because G is a finite group, so its order, $|G|$, is finite. We must also check the standard axioms for an inner product:

(i) The new inner product is sesquilinear, since:

$$[v, \lambda w + \mu u] = \frac{1}{|G|}\sum_{g\in G}(D(g)v, \lambda D(g)w + \mu D(g)u)$$

$$= \frac{1}{|G|}\sum_{g\in G}\left(\lambda(D(g)v, D(g)w) + \mu(D(g)v, D(g)u)\right)$$

$$= \lambda[v,w] + \mu[v,u],$$

and similarly:

$$[\lambda v + \mu u, w] = \lambda^*[v,w] + \mu^*[u,w].$$

(ii) For all $v \in V$, we have:

$$[v,v] = \frac{1}{|G|}\sum_{g\in G}(D(g)v, D(g)v) = \frac{1}{|G|}\sum_{g\in G}||D(g)v||^2 \geq 0,$$

since $(\cdot,\cdot)$ induces a norm as usual. Furthermore, we see equality holds if and only if $D(g)v = 0$ for all $g \in G$. But each $D(g)$ is an invertible matrix, so this holds if and only if $v = 0$.

(iii) For all $v, w \in V$, we have:

$$[v,w] = \frac{1}{|G|}\sum_{g\in G}(D(g)v, D(g)w) = \frac{1}{|G|}\sum_{g\in G}(D(g)w, D(g)v)^* = \left(\frac{1}{|G|}\sum_{g\in G}(D(g)w, D(g)v)\right)^* = [w,v]^*$$

using conjugate symmetry of the inner product $(\cdot,\cdot)$.

We now check that this inner product is unitary. For all $h \in G$ and $v, w \in V$, we have:

$$[D(h)v, D(h)w] = \frac{1}{|G|}\sum_{g\in G}(D(g)D(h)v, D(g)D(h)w) = \frac{1}{|G|}\sum_{g\in G}(D(gh)v, D(gh)w)$$

$$= \frac{1}{|G|}\sum_{kh^{-1}\in G}(D(k)v, D(k)w).$$

Now, for any fixed h, kh^{-1} ranges over the whole group and takes the value of each element exactly once as k varies. Thus we have:

$$\frac{1}{|G|}\sum_{kh^{-1}\in G}(D(k)v, D(k)w) = \frac{1}{|G|}\sum_{k\in G}(D(k)v, D(k)w) = [v,w],$$

and the inner product is unitary as required. It follows from the result of this question that the representation V of G is completely reducible. □

In fact, this proof can be naturally extended to groups which have a 'finite volume', namely *compact* (Lie) groups. Here, the sum which defines the inner product in the above proof is replaced by some *integral* over the group elements (this integral is called the *Haar integral* on the group).

5. (Note that this question uses physics conventions for the generators t_i, such that they are Hermitian.) Three 3×3 matrices $\mathbf{t} := (t_1, t_2, t_3)$ are defined by $(t_i)_{jk} = -i\epsilon_{ijk}$.

(a) Prove $[t_i, t_j] = i\epsilon_{ijk} t_k$.

(b) Prove $(\mathbf{n} \cdot \mathbf{t})^3 = |\mathbf{n}|^2 \mathbf{n} \cdot \mathbf{t}$, where $\mathbf{n}$ is any vector (not necessarily a unit vector).

(c) What are the possible eigenvalues of $\mathbf{n} \cdot \mathbf{t}$ if $\mathbf{n}$ is a unit vector?

(d) We may represent a rotation by an angle θ about an axis that points along the unit vector $\hat{\mathbf{n}}$ by the member of $SO(3)$ $R_{ij}(\hat{\mathbf{n}}, \theta) = \exp(-i\theta \hat{\mathbf{n}} \cdot \mathbf{t})_{ij}$. By convention, $\hat{\mathbf{n}}$ points in any direction any $0 \leq \theta \leq \pi$. Evaluate R_{ij} explicitly by summing the Taylor series of the exponential, and show that

$$R_{ij}(\hat{\mathbf{n}}, \theta) = \hat{n}_i \hat{n}_j + (\delta_{ij} - \hat{n}_i \hat{n}_j) \cos(\theta) - \epsilon_{ijk} \hat{n}_k \sin(\theta).$$

(e) Verify the formula: $e^{-i\theta \hat{\mathbf{n}} \cdot \boldsymbol{\sigma}/2} \sigma_j e^{i\theta \hat{\mathbf{n}} \cdot \boldsymbol{\sigma}/2} = R_{ij}(\hat{\mathbf{n}}, \theta) \sigma_i$.

(f) Given an n-dimensional representation $D : G \to GL(n, \mathbb{C})$ of a group G, we can define its *conjugate representation* $\overline{D} : G \to GL(n, \mathbb{C})$ by complex conjugation: $\overline{D}(g) = D(g)^*$ for all $g \in G$. If D and $\overline{D}$ are inequivalent, then we say D is a *complex representation*. If D and $\overline{D}$ are equivalent, then there exists some invertible $n \times n$ matrix S such that $\overline{D}(g) = SD(g)S^{-1}$ for all $g \in G$. In this case, if $S^T = S$, then D is said to be a *real representation*; otherwise, $S^T = -S$ and D is said to be *pseudoreal*. (These are the only two possibilities for equivalent, finite-dimensional representations.)

The set of matrices $\exp(-i\theta \hat{\mathbf{n}} \cdot \boldsymbol{\sigma}/2)$ constitutes the defining representation of $G = SU(2)$. Show that this representation is pseudoreal and that the conjugate representation has the same weights as the original.

Solution: (a) The matrices $\mathbf{t} = (t_1, t_2, t_3)$ span the fundamental representation of the Lie algebra $\mathfrak{so}(3)$ (compare this to the Pauli matrices $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ which span the fundamental representation of the Lie algebra $\mathfrak{su}(2)$). First, we are asked to establish some commutation relations, namely $[t_i, t_j] = i\epsilon_{ijk} t_k$. First, let us note that:

$$([t_i, t_j])_{ml} = (t_i t_j)_{ml} - (t_j t_i)_{ml}.$$

Recalling that the components of t_i are given by $(t_i)_{jk} = -i\epsilon_{ijk}$, we can evaluate each of the product $t_i t_j$ using the formula for the contraction of two epsilon symbols:

$$(t_i t_j)_{ml} = -\epsilon_{imp} \epsilon_{jpl} = -(\delta_{il} \delta_{mj} - \delta_{ij} \delta_{ml}).$$

Hence we're left with:

$$\begin{aligned}
([t_i, t_j])_{ml} &= -(\delta_{il} \delta_{mj} - \delta_{ij} \delta_{ml}) + (\delta_{im} \delta_{jl} - \delta_{ij} \delta_{ml}) \\
&= \delta_{im} \delta_{jl} - \delta_{il} \delta_{mj} \\
&= \epsilon_{ijk} \epsilon_{kml} && \text{(formula for contraction of two } \epsilon \text{ symbols)} \\
&= i\epsilon_{ijk} (t_k)_{ml} && \text{(definition of } t_k\text{)}
\end{aligned}$$

Hence $[t_i, t_j] = i\epsilon_{ijk} t_k$, as required.

(b) Next we are asked to show that $(\hat{\mathbf{n}}\cdot\mathbf{t})^3 = |\hat{\mathbf{n}}|^2\hat{\mathbf{n}}\cdot\mathbf{t}$, where $\mathbf{n}$ is *not* a unit vector in this part of the question. This is analogous to the result we had for the Pauli matrices, $(\hat{\mathbf{n}}\cdot\boldsymbol{\sigma})^2 = I$ for a unit vector $\hat{\mathbf{n}}$, which we proved on Sheet 1, Question 7(c).

First, let us consider $(\mathbf{n}\cdot\mathbf{t})^2$:

$$\begin{aligned}(\mathbf{n}\cdot\mathbf{t})^2_{ml} &= n_i n_j (t_i t_j)_{ml} \\ &= -n_i n_j(\delta_{il}\delta_{mj} - \delta_{ij}\delta_{ml}) && \text{(using product } t_i t_j \text{ derived in (a))} \\ &= -(n_l n_m - |\mathbf{n}|^2\delta_{ml}).\end{aligned}$$

Consider multiplying this by $\mathbf{n}\cdot\mathbf{t}$ on the right:

$$\begin{aligned}(\mathbf{n}\cdot\mathbf{t})^3_{mq} &= (\mathbf{n}\cdot\mathbf{t})^2_{ml}(\mathbf{n}\cdot\mathbf{t})_{lq} \\ &= i(n_l n_m - |\mathbf{n}|^2\delta_{ml})n_i\epsilon_{ilq} && \text{(from above, and } (t_i)_{lq} = -i\epsilon_{ilq}) \\ &= -i|\mathbf{n}|^2 n_i \epsilon_{imq} && \text{(since } n_i n_l \epsilon_{ilq} = 0) \\ &= |\mathbf{n}|^2(\mathbf{n}\cdot\mathbf{t})_{mq}.\end{aligned}$$

We deduce that $(\mathbf{n}\cdot\mathbf{t})^3 = |\mathbf{n}|^2\mathbf{n}\cdot\mathbf{t}$.

(c) Next, we are asked to give the possible values of the eigenvalues $\hat{\mathbf{n}}\cdot\mathbf{t}$ for a unit vector $\hat{\mathbf{n}}$. From (b), we know that if $\hat{\mathbf{n}}$ is a unit vector, we have $(\hat{\mathbf{n}}\cdot\mathbf{t})^3 = \hat{\mathbf{n}}\cdot\mathbf{t}$. If we now suppose that $\mathbf{v}$ is an eigenvector with eigenvalue λ, then we see that:

$$\lambda\mathbf{v} = (\hat{\mathbf{n}}\cdot\mathbf{t})\mathbf{v} = (\hat{\mathbf{n}}\cdot\mathbf{t})^3\mathbf{v} = \lambda^3\mathbf{v}$$

Thus the eigenvalues of $\hat{\mathbf{n}}\cdot\mathbf{t}$ satisfy $\lambda^3 = \lambda$; thus the possible values of the eigenvalues are $-1, 0$ and $+1$.

(d) Similar to Sheet 1, Question 7(c), we are asked to evaluate a matrix exponential $\exp(-i\theta\hat{\mathbf{n}}\cdot\mathbf{t})$. We will use the same trick here as we did for Sheet 1, Question 7(c), based on writing all powers of $\hat{\mathbf{n}}\cdot\mathbf{t}$ in terms of $\hat{\mathbf{n}}\cdot\mathbf{t}$. We first note that for $r \in \{2, 3, ...\}$:

$$(\hat{\mathbf{n}}\cdot\mathbf{t})^{2r} = (\hat{\mathbf{n}}\cdot\mathbf{t})^{2r-3}(\hat{\mathbf{n}}\cdot\mathbf{t})^3 = (\hat{\mathbf{n}}\cdot\mathbf{t})^{2r-3}(\hat{\mathbf{n}}\cdot\mathbf{t}) = (\hat{\mathbf{n}}\cdot\mathbf{t})^{2r-2},$$

using the result $(\hat{\mathbf{n}}\cdot\mathbf{t})^3 = \hat{\mathbf{n}}\cdot\mathbf{t}$ from part (b). Hence we can evaluate all even powers of $\hat{\mathbf{n}}\cdot\mathbf{t}$; for $r \in \{1, 2, ...\}$ we have:

$$\begin{aligned}(\hat{\mathbf{n}}\cdot\mathbf{t})^{2r}_{ij} &= (\hat{\mathbf{n}}\cdot\mathbf{t})^2_{ij} \\ &= -(\hat{n}_i\hat{n}_j - \delta_{ij}) && \text{(using part (b)).}\end{aligned}$$

We also note that the zeroth power is given by $(\hat{\mathbf{n}}\cdot\mathbf{t})^0_{ij} = \delta_{ij}$, the identity matrix.

Similarly, we can evaluate all odd powers of $\hat{\mathbf{n}}\cdot\mathbf{t}$; for $r \in \{0, 1, 2, 3...\}$ we have:

$$\begin{aligned}(\hat{\mathbf{n}}\cdot\mathbf{t})^{2r+1}_{ij} &= (\hat{\mathbf{n}}\cdot\mathbf{t})_{ij} \\ &= -i\hat{n}_k\epsilon_{kij} && \text{(since } (t_k)_{ij} = -i\epsilon_{kij}).\end{aligned}$$

We now have everything in place to evaluate the exponential:

$$
\begin{aligned}
\exp\left(-i\theta \hat{\mathbf{n}}\cdot\mathbf{t}\right)_{ij} &= \sum_{k=0}^{\infty}\frac{(-i\theta\hat{\mathbf{n}}\cdot\mathbf{t})^k_{ij}}{k!} \\
&= \sum_{r=0}^{\infty}\frac{(-i\theta)^{2r}}{(2r)!}(\hat{\mathbf{n}}\cdot\mathbf{t})^{2r}_{ij} + \sum_{r=0}^{\infty}\frac{(-i\theta)^{2r+1}}{(2r+1)!}(\hat{\mathbf{n}}\cdot\mathbf{t})^{2r+1}_{ij} \\
&= \delta_{ij} + (\delta_{ij} - \hat{n}_i\hat{n}_j)\sum_{r=1}^{\infty}\frac{(-i\theta)^{2r}}{(2r)!} - i\hat{n}_k\epsilon_{kij}\sum_{r=0}^{\infty}\frac{(-i\theta)^{2r+1}}{(2r+1)!} \\
&= \hat{n}_i\hat{n}_j + (\delta_{ij} - \hat{n}_i\hat{n}_j)\sum_{r=0}^{\infty}\frac{(-1)^r(\theta)^{2r}}{(2r)!} - \hat{n}_k\epsilon_{kij}\sum_{r=0}^{\infty}\frac{(-1)^{2r+1}(\theta)^{2r+1}}{(2r+1)!} \\
&= \hat{n}_i\hat{n}_j + (\delta_{ij} - \hat{n}_i\hat{n}_j)\cos(\theta) - \hat{n}_k\epsilon_{ijk}\sin(\theta).
\end{aligned}
$$

which is the required result.

(e) We evaluate $e^{-i\theta\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}/2}\sigma_j e^{i\theta\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}/2}$ and compare to the expression $R_{ij}(\hat{\mathbf{n}},\theta)\sigma_i$. Using Sheet 1, Question 7(c), we have:

$$
\begin{aligned}
e^{-i\theta\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}/2}\sigma_j e^{i\theta\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}/2} &= \left(\cos\left(\frac{\theta}{2}\right)I - i\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}\sin\left(\frac{\theta}{2}\right)\right)\sigma_j\left(\cos\left(\frac{\theta}{2}\right)I + i\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}\sin\left(\frac{\theta}{2}\right)\right) \\
&= \cos^2\left(\frac{\theta}{2}\right)\sigma_j - i\sin\left(\frac{\theta}{2}\right)\cos\left(\frac{\theta}{2}\right)\hat{n}_i[\sigma_i,\sigma_j] + \sin^2\left(\frac{\theta}{2}\right)(\hat{\mathbf{n}}\cdot\boldsymbol{\sigma})\sigma_j(\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}).
\end{aligned}
$$

Recall that $[\sigma_i,\sigma_j] = 2i\epsilon_{ijk}\sigma_k$, allowing us to simplify one of the terms. The final term has a product of three Pauli matrices, which can be simplified using the anticommutator:

$$
(\hat{\mathbf{n}}\cdot\boldsymbol{\sigma})\sigma_j(\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}) = -(\hat{\mathbf{n}}\cdot\boldsymbol{\sigma})^2\sigma_j + (\hat{\mathbf{n}}\cdot\boldsymbol{\sigma})\hat{n}_i\{\sigma_j,\sigma_i\} = -\sigma_j + 2\hat{n}_j(\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}).
$$

Substituting into the expansion of $e^{-i\theta\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}/2}\sigma_j e^{i\theta\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}/2}$, we have:

$$
e^{-i\theta\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}/2}\sigma_j e^{i\theta\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}/2} = \left(\cos^2\left(\frac{\theta}{2}\right) - \sin^2\left(\frac{\theta}{2}\right)\right)\sigma_j + 2\sin\left(\frac{\theta}{2}\right)\cos\left(\frac{\theta}{2}\right)\hat{n}_i\epsilon_{ijk}\sigma_k + 2\hat{n}_j(\hat{\mathbf{n}}\cdot\boldsymbol{\sigma})\sin^2\left(\frac{\theta}{2}\right)
$$

It is useful to simplify this using the double-angle formulae for the trigonometric functions. We have: $\cos^2(\theta/2) - \sin^2(\theta/2) = \cos(\theta)$, $2\sin(\theta/2)\cos(\theta/2) = \sin(\theta)$ and $\sin^2(\theta/2) = \frac{1}{2}(1-\cos(\theta))$. Hence we can write:

$$
\begin{aligned}
e^{-i\theta\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}/2}\sigma_j e^{i\theta\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}/2} &= \cos(\theta)\sigma_j + \sin(\theta)\hat{n}_i\epsilon_{ijk}\sigma_k + \hat{n}_i\hat{n}_j\sigma_i(1-\cos(\theta)) \\
&= \hat{n}_i\hat{n}_j\sigma_i + (\delta_{ij} - \hat{n}_i\hat{n}_j)\cos(\theta)\sigma_i - \hat{n}_k\epsilon_{ijk}\sin(\theta)\sigma_i \\
&= R_{ij}(\theta,\hat{\mathbf{n}})\sigma_i,
\end{aligned}
$$

using the result of part (d). Note that in the second from last equality, we performed some relabelling of dummy indices, $\hat{n}_i\epsilon_{ijk}\sigma_k = -\hat{n}_k\epsilon_{ijk}\sigma_i$.

(f) In the last part of this question, we must show that the fundamental representation $\exp(-i\theta\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}/2)$ of $SU(2)$ is a *pseudoreal* representation. First of all, let's explore what this means, and why we care.

Recall that given an n-dimensional representation $D : G \to GL(n,\mathbb{C})$ of the group G, we can define the *conjugate representation* via:

$$\bar{D} : G \to GL(n,\mathbb{C}), \qquad \bar{D}(g) = D(g)^*.$$

That is, all representation matrices of $\bar{D}$ are simply given by the complex conjugate of the corresponding representation matrices of D. A natural question we might ask is: when are these two representations *equivalent*? We recall that two matrix representations are equivalent if they are related by some change of basis matrix, that is, there exists an invertible $n \times n$ matrix S such that:

$$\bar{D}(g) = SD(g)S^{-1}.$$

In physical applications, we are particularly interested in *unitary* representations, so let's suppose that $D(g)$ is a unitary representation[4] (from which it follows that $\bar{D}(g)$ is a unitary representation, since $\bar{D}(g)\bar{D}(g)^\dagger = D(g)^*(D(g)^*)^\dagger = (D(g)D(g)^\dagger)^* = I^* = I$). Then we have:

$$SD(g)S^{-1} = \bar{D}(g) = D(g)^* = D(g)^{-T}.$$

We can use this equation to deduce something about S; multiplying on the left by S^{-T} and on the right by S^T, we have:

$$S^{-T}SD(g)S^{-1}S^T = S^{-T}D(g)^{-T}S^T = (SD(g)S^{-1})^{-T} = (D(g)^{-T})^{-T} = D(g).$$

This implies that $S^{-T}SD(g) = D(g)S^{-T}S$ for all $g \in G$, and so by Schur's Lemma (see Sheet 3), we have $S^{-T}S = aI$, i.e. $S = aS^T$ for some constant $a \in \mathbb{C}$. Note that this implies:

$$S = aS^T = a(aS^T)^T = a^2 S \qquad \Rightarrow \qquad a = \pm 1.$$

Thus the change of basis matrix S obeys *either* $S = S^T$ or $S = -S^T$.

This suggests that we can distinguish between three cases: D, $\bar{D}$ are not equivalent, D, $\bar{D}$ are equivalent through a symmetric matrix and finally D, $\bar{D}$ are equivalent through an antisymmetric matrix. We formalise this in the following definition:

Definition: Let $D : G \to GL(n,\mathbb{C})$ be an n-dimensional unitary representation of the group G on the vector space $\mathbb{C}^n$. We make the following definitions:

- We say that this is a *real representation* if (i) D is equivalent to its conjugate representation $\bar{D}$ *and* (ii) the change of basis matrix S obeys $S = S^T$.
- We say that this is a *pseudoreal representation* if (i) D is equivalent to its conjugate representation $\bar{D}$ *and* (ii) the change of basis matrix S obeys $S = -S^T$.
- The representation is called a *complex representation* if it is not equivalent to its conjugate representation.

In the case that we have a real or pseudoreal representation, the change of basis matrix S can be scaled such that it is also a *unitary* matrix. To see this, recall the defining equation $SD(g)S^{-1} = D(g)^{-T}$, which implies that:

$$S = D(g)^T SD(g), \qquad S^\dagger = D(g)^\dagger S^\dagger D(g)^*.$$

Multiplying these expressions, we have:

$$S^\dagger S = D(g)^\dagger S^\dagger D(g)^* D(g)^T SD(g) = D(g)^\dagger S^\dagger (D(g)D(g)^\dagger)^T SD(g) = D(g)^\dagger S^\dagger SD(g)$$

$$\Rightarrow \qquad D(g)S^\dagger S = S^\dagger SD(g),$$

for all $g \in G$. Hence by Schur's Lemma, we have $S^\dagger S = \lambda I$, for some $\lambda \in \mathbb{C}$.

[4] In fact, all finite-dimensional representations of finite or compact groups are equivalent to unitary representations, so this isn't much of a restriction.

We further note that:

$$S = \lambda S^{-\dagger} = \lambda(\lambda S^{-\dagger})^{-\dagger} = \frac{\lambda}{\lambda^*}S \qquad \Rightarrow \qquad \lambda = \lambda^*,$$

so $\lambda \in \mathbb{R}$ is a *real* constant. It follows that $S \mapsto S' = S/\sqrt{\lambda}$ gives a unitary change of basis matrix, since $(S')^\dagger S' = (S/\sqrt{\lambda})^\dagger(S/\sqrt{\lambda}) = S^\dagger S/\lambda = I$. This allows us to show the interesting result:

Theorem: Let $D : G \to GL(n, \mathbb{C})$ be a real representation. Then there exists a basis in which all of the representation matrices $D(g)$ have purely real entries.

Proof: Since D is real, there exists a unitary, symmetric matrix S such that $SD(g)S^\dagger = D(g)^*$ for all $g \in G$. Since S is unitary, all its eigenvalues are of the form $e^{i\theta}$, which implies that S can be written in the form $S = e^{iT}$ for some matrix T. The further condition that S is symmetric implies that T is also symmetric. Let $R = e^{iT/2}$ so that $S = R^2$, and R is also a unitary, symmetric matrix (so in particular still a change of basis matrix). Then:

$$R^2 D(g) R^{-2} = D(g)^*.$$

We can use this to show that $RD(g)R^{-1}$ is real for all $g \in G$. Since R is symmetric and unitary, we have $R^* = R^\dagger = R^{-1}$; hence we have:

$$(RD(g)R^{-1})^* = R^{-1}D(g)^*R = R^{-1}R^2D(g)R^{-2}R = RD(g)R^{-1} \;\Rightarrow\; RD(g)R^{-1} \in \mathrm{Mat}_{n\times n}(\mathbb{C}). \quad \square$$

More generally (but strictly beyond the scope of this course), one can show that: a *real representation* is a representation where all representation matrices can be chosen to be real matrices, a *pseudoreal representation* is a representation which is equivalent to its conjugate representation *but* where there is no basis in which all the representation matrices are real matrices, a *complex representation* is a representation which is inequivalent to its conjugate representation *and* where there is no basis in which all the representation matrices are real matrices.

The actual question itself is fairly trivial. We note that the conjugate representation of the fundamental representation of $SU(2)$ is given by:

$$\exp\left(-\frac{i\theta\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}}{2}\right)^* = \exp\left(\frac{i\theta\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}^*}{2}\right).$$

Now by Sheet 1, Question 7(b), we have $\boldsymbol{\sigma}^* = -\sigma_2\boldsymbol{\sigma}\sigma_2$. Hence we have:

$$\exp\left(\frac{i\theta\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}^*}{2}\right) = \exp\left(-\frac{i\theta\hat{\mathbf{n}}\cdot(\sigma_2\boldsymbol{\sigma}\sigma_2)}{2}\right).$$

Now, note that when we expand the exponential, we will get lots of products of the form $(\sigma_2\sigma_{i_1}\sigma_2)...(\sigma_2\sigma_{i_n}\sigma_2)$. But $\sigma_2^2 = I$, hence we can simply draw the σ_2's out of these products to leave: $\sigma_2(\sigma_{i_1}...\sigma_{i_n})\sigma_2$. Hence the exponential can be simplified to:

$$\exp\left(-\frac{i\theta\hat{\mathbf{n}}\cdot(\sigma_2\boldsymbol{\sigma}\sigma_2)}{2}\right) = \sigma_2\exp\left(-\frac{i\theta\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}}{2}\right)\sigma_2$$

Hence we have shown that the conjugate representation and fundamental representation of $SU(2)$ are conjugate via the matrix σ_2 (which is self-inverse):

$$\exp\left(-\frac{i\theta\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}}{2}\right)^* = \sigma_2\exp\left(-\frac{i\theta\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}}{2}\right)\sigma_2.$$

Noting that $\sigma_2^T = -\sigma_2$, we see that the fundamental representation of $SU(2)$ is indeed *pseudoreal*.

We can also note that the weights of the two representations are the same; this follows immediately from the first part of Question 8.

6. This question regards the explicit map of $SO(3) \cong SU(2)/\mathbb{Z}_2$.

(a) Show that $\mathrm{Tr}(\sigma_i\sigma_j) = 2\delta_{ij}$. Why does this imply that any 2×2 complex matrix A can be expressed as:

$$A = \frac{1}{2}\mathrm{Tr}(A)I + \frac{1}{2}\mathrm{Tr}(\boldsymbol{\sigma}A)\cdot\boldsymbol{\sigma}?$$

(b) Define a one-to-one correspondence between real three vectors and Hermitian traceless 2×2 complex matrices by $\mathbf{x} \leftrightarrow \mathbf{x}\cdot\boldsymbol{\sigma}$. Show that $\det(\mathbf{x}\cdot\boldsymbol{\sigma}) = -\mathbf{x}^2$.

(c) Next, we define a linear transformation $\mathbf{x} \mapsto \mathbf{x}'$ by $\mathbf{x}'\cdot\boldsymbol{\sigma} = A\mathbf{x}\cdot\boldsymbol{\sigma}A^\dagger$, for $A \in SU(2)$. Deduce that $(\mathbf{x}')^2 = \mathbf{x}^2$, and so $x_i' = R_{ij}x_j$ for some $R \in SO(3)$. Finally show that:

$$R_{ij} = \frac{1}{2}\mathrm{Tr}(\sigma_i A\sigma_j A^\dagger).$$

(d) Show that $\sigma_j\sigma_i\sigma_j = -\sigma_i$ implies $\sigma_j A^\dagger\sigma_j = 2\mathrm{Tr}(A^\dagger)I - A^\dagger$ to obtain the equations $\sigma_i R_{ij}\sigma_j = 2\mathrm{Tr}(A^\dagger)A - I$ and $R_{jj} = |\mathrm{Tr}(A)|^2 - 1$.

(e) Why must $\mathrm{Tr}(A) \in \mathbb{R}$? Solve for $\mathrm{Tr}(A)$ and then A to show that:

$$A = \pm\frac{I + \sigma_i R_{ij}\sigma_j}{2\sqrt{1+R_{jj}}}.$$

Solution: (a) For the first part, we use the identity from Sheet 1, Question 7(a):

$$\mathrm{Tr}(\sigma_i\sigma_j) = \mathrm{Tr}(\delta_{ij}I + i\epsilon_{ijk}\sigma_k) = 2\delta_{ij},$$

since the trace distributes over addition, the Pauli matrices are traceless, and $\mathrm{Tr}(I) = 2$ for the 2×2 identity matrix. It will also be useful to notice that:

$$\mathrm{Tr}(I\sigma_j) = 0, \qquad \mathrm{Tr}(\sigma_i I) = 0, \qquad \text{and} \qquad \mathrm{Tr}(I^2) = 2.$$

Defining σ_μ to be the components of the vector $(I, \boldsymbol{\sigma})$, these identities can be summarised as:

$$\frac{1}{2}\mathrm{Tr}(\sigma_\mu\sigma_\nu) = \delta_{\mu\nu},$$

where $\delta_{\mu\nu}$ is the Kronecker delta. This property suggests that the Pauli matrices, together with the identity matrix, are 'orthonormal' with respect to the product $(A, B) = \frac{1}{2}\mathrm{Tr}(A^\dagger B)$ defined on the set of 2×2 complex matrices. In fact, this is an *inner product*, justifying the use of the word orthonormal here:

Proposition: The map $(\cdot,\cdot) : \mathrm{Mat}_{2\times 2}(\mathbb{C}) \times \mathrm{Mat}_{2\times 2}(\mathbb{C}) \to \mathbb{C}$ given by:

$$(A, B) = \frac{1}{2}\mathrm{Tr}(A^\dagger B)$$

is an inner product. We call this inner product the *Frobenius inner product* on the vector space of 2×2 complex matrices. Furthermore the set $\{I, \sigma_1, \sigma_2, \sigma_3\}$ forms an orthonormal basis for $\mathrm{Mat}_{2\times 2}(\mathbb{C})$ with respect to this inner product.

Proof: We begin by noting that the set $\{I, \sigma_1, \sigma_2, \sigma_3\}$, consisting of the Pauli matrices together with the identity, constitutes a basis for the space of 2×2 complex matrices (viewed as a four-dimensional complex vector space). To see this, note that:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \left(\frac{a+d}{2}\right)I + \left(\frac{a-d}{2}\right)\sigma_3 + \left(\frac{b+c}{2}\right)\sigma_1 + i\left(\frac{b-c}{2}\right)\sigma_2,$$

hence the Pauli matrices together with the identity span.

Furthermore, the left hand side is zero if and only if the coefficients on the right hand side all vanish. Thus the Pauli matrices together with the identity are linearly independent, and we conclude that $\{I, \sigma_1, \sigma_2, \sigma_3\}$ is a basis for $\mathrm{Mat}_{2\times 2}(\mathbb{C})$.

In particular, we can write any matrix $A \in \mathrm{Mat}_{2\times 2}(\mathbb{C})$ in the form $A = x^\mu \sigma_\mu$. Given $A = x^\mu \sigma_\mu$ and $B = y^\mu \sigma_\mu$, we have:

$$(A, B) = \frac{1}{2} x^{\mu *} y^\nu \mathrm{Tr}(\sigma_\mu \sigma_\nu) = x \cdot y,$$

using the result from above on the trace of $\sigma_\mu \sigma_\nu$. The right hand side is the standard inner product on $\mathbb{C}$. Thus $(\cdot, \cdot)$ is also an inner product as required, and the Pauli matrices together with the identity clearly constitute an orthonormal basis with respect to this inner product. $\square$

This result immediately implies:

$$A = (I, A)I + (\boldsymbol{\sigma}, A) \cdot \boldsymbol{\sigma} = \frac{1}{2}\mathrm{Tr}(A)I + \frac{1}{2}\mathrm{Tr}(\boldsymbol{\sigma} A)$$

as required, simply expanding A with respect to this orthonormal basis.

(b) To make things concrete, we shall write $\mathcal{H}$ for the set of Hermitian, traceless 2×2 complex matrices. This set $\mathcal{H}$ is clearly a vector space, with a basis given by $\{\sigma_1, \sigma_2, \sigma_3\}$.

The given correspondence is a map $\phi : \mathbb{R}^3 \to \mathcal{H}$ given by $\phi(\mathbf{x}) = \mathbf{x} \cdot \boldsymbol{\sigma}$. This is clearly a linear bijection, so it is an isomorphism of vector spaces. Furthermore, it is an isomorphism of *inner product spaces*, with respect to the Frobenius inner product restricted to the subspace $\mathcal{H} \leq \mathrm{Mat}_{2\times 2}(\mathbb{C})$, since:

$$(\mathbf{x}, \mathbf{y})_{\mathbb{R}^3} = \mathbf{x} \cdot \mathbf{y} = \frac{1}{2}\mathrm{Tr}((\mathbf{x} \cdot \boldsymbol{\sigma})(\mathbf{y} \cdot \boldsymbol{\sigma})) = (\mathbf{x} \cdot \boldsymbol{\sigma}, \mathbf{y} \cdot \boldsymbol{\sigma})_{\mathcal{H}} = (\phi(\mathbf{x}), \phi(\mathbf{y}))_{\mathcal{H}}.$$

We use this relation to prove the key results in this question.

In part (b), we begin by showing that $\det(\mathbf{x} \cdot \boldsymbol{\sigma}) = -\mathbf{x}^2$. Setting $\mathbf{x} = \mathbf{y}$ in the above relation, we have:

$$\mathbf{x}^2 = \frac{1}{2}\mathrm{Tr}((\mathbf{x} \cdot \boldsymbol{\sigma})^2). \qquad (*)$$

Let λ, μ be the eigenvalues of a 2×2 complex matrix A. Then A^2 has eigenvalues λ^2, μ^2 (with the same eigenvectors), so we have:

$$\det(A) = \lambda\mu = \frac{1}{2}(\lambda + \mu)^2 - \frac{1}{2}\lambda^2 - \frac{1}{2}\mu^2 = \frac{1}{2}\mathrm{Tr}(A)^2 - \frac{1}{2}\mathrm{Tr}(A^2).$$

Since $\mathbf{x} \cdot \boldsymbol{\sigma}$ is traceless, applying this result to $(*)$ yields:

$$\mathbf{x}^2 = \frac{1}{2}\mathrm{Tr}((\mathbf{x} \cdot \boldsymbol{\sigma})^2) = -\det(\mathbf{x} \cdot \boldsymbol{\sigma}),$$

as required. The key takeaway from this point is that the induced norms from the inner products on $\mathbb{R}^3$ and $\mathcal{H}$ obey:

$$||\mathbf{x}||^2_{\mathbb{R}^3} = ||\phi(\mathbf{x})||^2_{\mathcal{H}} = -\det(\mathbf{x} \cdot \boldsymbol{\sigma})$$

in terms of the map $\phi(\mathbf{x}) = \mathbf{x} \cdot \boldsymbol{\sigma}$.

(c) In this part, we show that for each $A \in SU(2)$, there exists $R \in SO(3)$ such that $A(\mathbf{x} \cdot \boldsymbol{\sigma})A^\dagger = (R\mathbf{x}) \cdot \boldsymbol{\sigma}$ for all $\mathbf{x} \in \mathbb{R}^3$.

Fix some $A \in SU(2)$. We can define a linear map $\mathrm{Ad}(A) : \mathcal{H} \to \mathcal{H}$ via:

$$\mathrm{Ad}(A)(\mathbf{x} \cdot \boldsymbol{\sigma}) = A(\mathbf{x} \cdot \boldsymbol{\sigma})A^\dagger.$$

This map induces a corresponding linear map $\phi^{-1} \circ \mathrm{Ad}(A) \circ \phi : \mathbb{R}^3 \to \mathbb{R}^3$ via the isomorphism $\phi : \mathbb{R}^3 \to \mathcal{H}$ (the map is linear because it is the composition of linear maps). Furthermore, the induced map $\phi^{-1} \circ \mathrm{Ad}(A) \circ \phi$ preserves the inner product on $\mathbb{R}^3$:

$$\begin{aligned}
(\phi^{-1} \circ \mathrm{Ad}(A) \circ \phi(\mathbf{x}), \phi^{-1} \circ \mathrm{Ad}(A) \circ \phi(\mathbf{y}))_{\mathbb{R}^3} &= (\mathrm{Ad}(A) \circ \phi(\mathbf{x}), \mathrm{Ad}(A) \circ \phi(\mathbf{y}))_{\mathcal{H}} && (\phi \text{ preserves inner products}) \\
&= (A\phi(\mathbf{x})A^\dagger, A\phi(\mathbf{y})A^\dagger)_{\mathcal{H}} \\
&= \frac{1}{2}\mathrm{Tr}(A\phi(\mathbf{x})A^\dagger A\phi(\mathbf{y})A^\dagger) \\
&= \frac{1}{2}\mathrm{Tr}(\phi(\mathbf{x})\phi(\mathbf{y})) && (\text{cyclicity of trace, } A^\dagger A = I) \\
&= (\phi(\mathbf{x}), \phi(\mathbf{y}))_{\mathcal{H}} \\
&= (\mathbf{x}, \mathbf{y})_{\mathbb{R}^3} && (\phi \text{ preserves inner products})
\end{aligned}$$

Alternatively, we can observe that $\phi^{-1} \circ \mathrm{Ad}(A) \circ \phi$ preserves the norm on $\mathbb{R}^3$, since:

$$\begin{aligned}
||\phi^{-1} \circ \mathrm{Ad}(A) \circ \phi(\mathbf{x})||^2 &= -\det(\mathrm{Ad}(A) \circ \phi(\mathbf{x})) && (\text{above determinant result}) \\
&= -\det(A\phi(\mathbf{x})A^\dagger) \\
&= -\det(\phi(\mathbf{x})) && (\det(A) = \det(A)^\dagger = 1) \\
&= ||\mathbf{x}||^2 && (\text{above determinant result})
\end{aligned}$$

Either way, we see that the linear map $\phi^{-1} \circ \mathrm{Ad}(A) \circ \phi$ is an *orthogonal* map on $\mathbb{R}^3$ (it preserves inner products, or alternatively, preserves norms). It follows that there exists some unique matrix $R(A) \in O(3)$ (dependent on A, with respect to the standard basis, say) such that:

$$\phi^{-1} \circ \mathrm{Ad}(A) \circ \phi(\mathbf{x}) = R(A)\mathbf{x}.$$

Applying ϕ to both sides, and recalling the definitions of ϕ, $\mathrm{Ad}(A)$, we have:

$$A(\mathbf{x} \cdot \boldsymbol{\sigma})A^\dagger = (R(A)\mathbf{x}) \cdot \boldsymbol{\sigma}.$$

It remains to show that $R(A) \in SO(3)$. This can be shown using a continuity argument similar to that given in the comments of Examples Sheet 1, Question 2. The left hand side is clearly a continuous function of $A \in SU(2)$, and $SU(2)$ is a *path-connected group* (i.e. it has no sudden jumps), hence $R(A)$ must lie in a path-connected component of $O(3)$ too. Since $I(\mathbf{x} \cdot \boldsymbol{\sigma})I^\dagger = \mathbf{x} \cdot \boldsymbol{\sigma} = (R(I)\mathbf{x}) \cdot \boldsymbol{\sigma}$ implies that $R(I) = I$, it follows that $R(A) \in SO(3)$.

We have now established that given any $A \in SU(2)$, there exists some unique $R \in SO(3)$ such that:

$$A(\mathbf{x}\cdot\boldsymbol{\sigma})A^\dagger = (R(A)\mathbf{x})\cdot\boldsymbol{\sigma}$$

for all $\mathbf{x} \in \mathbb{R}^3$. The right hand side is already written with respect to the orthonormal basis $\{\sigma_1, \sigma_2, \sigma_3\}$ of $\mathcal{H}$. Expanding the left hand side in terms of the orthonormal basis, we can compare coefficients:

$$A(\mathbf{x}\cdot\boldsymbol{\sigma})A^\dagger = (\sigma_i, A(\mathbf{x}\cdot\boldsymbol{\sigma}A^\dagger)\sigma_i = \frac{1}{2}\mathrm{Tr}(\sigma_i A(\mathbf{x}\cdot\boldsymbol{\sigma})A^\dagger)\sigma_i.$$

Comparing coefficients, we have:

$$(R(A)\mathbf{x})_i = R(A)_{ij}x_j = \frac{1}{2}\mathrm{Tr}(\sigma_i A(\mathbf{x}\cdot\boldsymbol{\sigma})A^\dagger) = \frac{1}{2}x_j\,\mathrm{Tr}\left(\sigma_i A\sigma_j A^\dagger\right).$$

Since this must hold for all $\mathbf{x} \in \mathbb{R}^3$, we have:

$$R(A)_{ij} = \frac{1}{2}\mathrm{Tr}\left(\sigma_i A\sigma_j A^\dagger\right),$$

as required.

(d) In the final two parts of the question, we construct the inverse map which takes us from $R(A)$ to a coset $\{\pm A\}$ of $SU(2)/\mathbb{Z}_2$. Fix $R \in SO(3)$. Assuming the map we constructed in the first part is surjective (proving this is possible, but tedious), there exists some $A(R) \in SU(2)$ such that:

$$A(R)(\mathbf{x}\cdot\boldsymbol{\sigma})A(R)^\dagger = (R\mathbf{x})\cdot\boldsymbol{\sigma}$$

for all $\mathbf{x} \in \mathbb{R}^3$. Since this must hold for all $\mathbf{x} \in \mathbb{R}^3$, we have in index notation:

$$A(R)\sigma_j A(R)^\dagger = R_{ij}\sigma_i.$$

We would like to derive a formula that says $A(R) = \cdots$, with no free indices on the right hand side. Thus let's contract with σ_j on the right hand side:

$$A(R)\sigma_j A(R)^\dagger\sigma_j = \sigma_i R_{ij}\sigma_j. \tag{†}$$

Next, we need to remove $\sigma_j A(R)^\dagger\sigma_j$ from the left hand side. It is at this point that the question's hints become useful. Using the result of part (a), we can expand $A^\dagger$ as (suppressing the R dependence everywhere):

$$A^\dagger = \frac{1}{2}\mathrm{Tr}\left(A^\dagger\right)I + \frac{1}{2}\mathrm{Tr}\left(\sigma_i A^\dagger\right)\sigma_i,$$

which gives:

$$\begin{aligned}
\sigma_j A^\dagger\sigma_j &= \frac{1}{2}\sigma_j\left(\mathrm{Tr}\left(A^\dagger\right)I + \mathrm{Tr}\left(\sigma_i A^\dagger\right)\sigma_i\right)\sigma_j \\
&= \frac{3}{2}\mathrm{Tr}\left(A^\dagger\right)I - \frac{1}{2}\mathrm{Tr}\left(\sigma_i A^\dagger\right)\sigma_i && \text{(using } \sigma_j\sigma_i\sigma_j = -\sigma_i\text{)} \\
&= 2\mathrm{Tr}\left(A^\dagger\right)I - \frac{1}{2}\mathrm{Tr}\left(A^\dagger\right)I - \frac{1}{2}\mathrm{Tr}\left(\sigma_i A^\dagger\right)\sigma_i \\
&= 2\mathrm{Tr}\left(A^\dagger\right)I - A^\dagger,
\end{aligned}$$

as required.

Inserting this into the equation (†), we have:

$$A(R)\left(2\mathrm{Tr}\left(A(R)^\dagger\right)I - A(R)^\dagger\right) = \sigma_i R_{ij}\sigma_j.$$

Simplifying the left hand side using $A(R)A(R)^\dagger = I$, we have:

$$2\mathrm{Tr}\left(A(R)^\dagger\right)A(R) - I = \sigma_i R_{ij}\sigma_j. \tag{††}$$

This is one of the equations the question wanted us to obtain.

It remains to express $\mathrm{Tr}(A(R)^\dagger)$ in terms of R; we can then solve for $A(R)$ and we will be done. Taking the trace of the equation we have just derived, we obtain (noting that $\mathrm{Tr}(A^\dagger) = \mathrm{Tr}(A)^*$):

$$2\left(|\mathrm{Tr}(A(R))|^2 - 1\right) = R_{ij}\mathrm{Tr}(\sigma_i\sigma_j) = R_{ij}\mathrm{Tr}(\delta_{ij}I + i\epsilon_{ijk}\sigma_k) = 2R_{jj}.$$

Hence we have derived $R_{jj} = |\mathrm{Tr}(A(R))|^2 - 1$, as required.

(e) We can argue that $\mathrm{Tr}(A(R))$ is real using the fact that $A(R) \in SU(2)$. Since $A(R)$ is special unitary, its eigenvalues must be phases, and must multiply together to give 1. Thus the eigenvalues are of the form $e^{\pm i\theta}$, and $\mathrm{Tr}(A(R)) = e^{i\theta} + e^{-i\theta} = 2\cos(\theta) \in \mathbb{R}$. It follows that:

$$R_{jj} = \mathrm{Tr}(A(R))^2 - 1 \qquad \Rightarrow \qquad \mathrm{Tr}(A(R)) = \pm\sqrt{1 + R_{jj}}.$$

Substituting into (††) (note that $\mathrm{Tr}(A(R)^\dagger) = \mathrm{Tr}(A(R))^* = \mathrm{Tr}(A(R))$ in this case) and rearranging, we obtain the final result:

$$A(R) = \pm\frac{I + \sigma_i R_{ij}\sigma_j}{2\sqrt{1 + R_{jj}}},$$

as required.

7. Finding the explicit map of $SO_+(1,3) \cong SL(2,\mathbb{C})/\mathbb{Z}_2$ follows a similar calculation to the one finding the map of $SO(3) \cong SU(2)/\mathbb{Z}_2$ in Question 6.

(a) Defining $\sigma_\mu = (I, \boldsymbol{\sigma}), \overline{\sigma}_\mu = (I, -\boldsymbol{\sigma})$, argue that any 2×2 complex matrix A may be written as $A = \frac{1}{2}\mathrm{Tr}(\overline{\sigma}^\mu A)\sigma_\mu$.

(b) Now define a one-to-one correspondence between real 4-vectors x_μ and Hermitian 2×2 matrices X, where $x_\mu \mapsto X = \sigma_\mu x^\mu$. Find $\det(X)$ in terms of x_μ.

(c) For any $A \in SL(2,\mathbb{C})$, we define a linear transformation $X \mapsto X' = AXA^\dagger = (X')^\dagger$. Show that $x^2 = (x')^2$ and hence this must be a Lorentz transformation, so we can write $(x')^\mu = \Lambda^\mu{}_\nu x^\nu$ where $\Lambda \in SO_+(1,3)$. Thus show $\Lambda^\mu{}_\nu = \mathrm{Tr}(\overline{\sigma}^\mu A \sigma_\nu A^\dagger)/2$.

(d) To find the inverse, show $\sigma_\nu A^\dagger \overline{\sigma}^\nu = 2\mathrm{Tr}(A^\dagger)I \Rightarrow \Lambda^\mu{}_\mu = |\mathrm{Tr}(A)|^2$ and $\sigma_\mu \Lambda^\mu{}_\nu \overline{\sigma}^\nu = 2\mathrm{Tr}(A^\dagger)A$ and hence, for $\mathrm{Tr}(A) = e^{i\alpha}|\mathrm{Tr}(A)|$, $A = e^{i\alpha}\sigma_\mu \Lambda^\mu{}_\nu \overline{\sigma}^\nu / 2\sqrt{\Lambda^\mu{}_\mu}$.

(e) Show that $\det(A) = 1$ determines $e^{i\alpha}$ up to a factor of ± 1. Thus $\pm A \leftrightarrow \Lambda$.

Solution: (a) Throughout, we write $SO_+(1,3)$ for the *proper, orthochronous Lorentz group*, i.e. those Lorentz transformations connected to the identity of the Lorentz group $O(1,3)$ consisting of all matrices Λ such that $\Lambda^T \eta \Lambda = \eta$.

We already showed in Question 7 that any 2×2 complex matrix $A \in \mathrm{Mat}_{2\times 2}(\mathbb{C})$ may be expressed in the form:

$$A = \frac{1}{2}\mathrm{Tr}(A)I + \frac{1}{2}\mathrm{Tr}(\boldsymbol{\sigma}A)\cdot\boldsymbol{\sigma}.$$

Interpreting μ as a Lorentz index, which must be raised and lowered with respect to the Minkowski metric, we see that $\overline{\sigma}^\mu = (I, \boldsymbol{\sigma})$, and hence we can write:

$$A = \frac{1}{2}\mathrm{Tr}(\overline{\sigma}^\mu A)\sigma_\mu,$$

as required.

(b) The next part of the question follows a very similar logic to Question 7. The space of 2×2 Hermitian matrices forms a *real* vector space $\mathcal{H}$, with an orthonormal basis provided by $\{I, \sigma_1, \sigma_2, \sigma_3\}$ (we will use the same notation as Question 7 to highlight the similarities in the proof - but $\mathcal{H}$ denotes the set of Hermitian matrices that could be traceful in this question!). This space is isomorphic to $\mathbb{R}^4$ via the explicit map $\phi : \mathbb{R}^4 \to \mathcal{H}$:

$$\phi(x) = x^\mu \sigma_\mu,$$

as suggested in the question.

In this case, we are encouraged to use the determinant trick from Question 7 to eventually show that the invariant distance is preserved under $SL(2,\mathbb{C})$ conjugation. Here ,the determinant of $\phi(x)$ is given by the invariant distance $x^2 = (x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2$. To see this, recall from Question 7 that for any 2×2 matrix A, we have:

$$\det(A) = \frac{1}{2}\mathrm{Tr}(A)^2 - \frac{1}{2}\mathrm{Tr}(A^2).$$

Hence:

$$\begin{aligned}\det(\phi(x)) &= \frac{1}{2}\mathrm{Tr}(x^0\sigma_0 + x^i\sigma_i)^2 - \frac{1}{2}\mathrm{Tr}(x^\mu\sigma_\mu x^\nu \sigma_\nu)\\ &= 2(x^0)^2 - \left((x^0)^2 + (x^1)^2 + (x^2)^2 + (x^3)^2\right)\\ &= x^2,\end{aligned}$$

using the fact that $\frac{1}{2}\mathrm{Tr}(\sigma_\mu\sigma_\nu) = \delta_{\mu\nu}$ from Question 7 (note that the Kronecker delta does *not* lower indices here, we need the Minkowski metric for that).

(c) Now, as before, fix some $A \in SL(2,\mathbb{C})$. We define $\mathrm{Ad}(A) : \mathcal{H} \to \mathcal{H}$ via:

$$\mathrm{Ad}(A)(x^\mu \sigma_\mu) = A x^\mu \sigma_\mu A^\dagger.$$

This is obviously still a Hermitian matrix, and clearly this map is linear. This map induces a linear map $\phi^{-1} \circ \mathrm{Ad}(A) \circ \phi : \mathbb{R}^4 \to \mathbb{R}^4$ as before. We see that this map preserves the invariant distance:

$$\begin{aligned}
(\phi^{-1} \circ \mathrm{Ad}(A) \circ \phi(x))^2 &= \det(\mathrm{Ad}(A) \circ \phi(x)) && (\text{since } \det(\phi(y)) = y^2 \text{ from above}) \\
&= \det(A\phi(x)A^\dagger) \\
&= \det(\phi(x)) && (A \in SL(2,\mathbb{C})) \\
&= x^2.
\end{aligned}$$

Thus $\phi^{-1} \circ \mathrm{Ad}(A) \circ \phi : \mathbb{R}^4 \to \mathbb{R}^4$ is a linear map which preserves the invariant distance x^2. It follows that it is a *Lorentz transformation*, hence there exists a unique matrix $\Lambda(A) \in O(1,3)$ (with respect to, say, the standard basis) such that:

$$\phi^{-1} \circ \mathrm{Ad}(A) \circ \phi(x) = \Lambda(A)x$$

for all $x \in \mathbb{R}^4$. Applying ϕ to both sides, and recalling the definition of ϕ and $\mathrm{Ad}(A)$, we have:

$$A x^\mu \sigma_\mu A^\dagger = (\Lambda(A)x)^\mu \sigma_\mu = \Lambda(A)^\mu{}_\nu x^\nu \sigma_\mu.$$

for all $x \in \mathbb{R}^4$.

It remains to show that $\Lambda(A) \in SO_+(1,3)$, the part of the Lorentz group connected to the identity (i.e. the *proper, orthochronous* Lorentz group). As usual, we use a continuity argument. Note that the left hand side of the above equation is clearly a continuous function of A. Furthermore, $SL(2,\mathbb{C})$ is a connected group,[5] and since $x^\mu \sigma_\mu = (\Lambda(I)x)^\mu \sigma_\mu$ for all $x \in \mathbb{R}^4$ implies $\Lambda(I) = I$, we deduce that $\Lambda(A) \in SO_+(1,3)$ as required.

In the next part of the question, we derive the explicit form of $\Lambda(A)$ in terms of A. The proof is very similar to Question 7. The equation we derived above holds for all $x \in \mathbb{R}^4$, thus we must be able to rewrite it as:

$$\Lambda(A)^\mu{}_\nu \sigma_\mu = A\sigma_\nu A^\dagger.$$

The left hand side is already written with respect to the orthonormal basis $\{I, \sigma_1, \sigma_2, \sigma_3\}$ of $\mathcal{H}$. Thus we express the right hand side in terms of this orthonormal basis, then compare coefficients. We have:

$$A\sigma_\nu A^\dagger = (I, A\sigma_\nu A^\dagger)I + (\sigma_i, A\sigma_\nu A^\dagger)\sigma_i = (\overline{\sigma}^\mu, A\sigma_\nu A^\dagger)\sigma_\mu = \frac{1}{2}\mathrm{Tr}(\overline{\sigma}^\mu A\sigma_\nu A^\dagger)\sigma_\mu.$$

Thus comparing coefficients, we immediately have:

$$\Lambda(A)^\mu{}_\nu = \frac{1}{2}\mathrm{Tr}(\overline{\sigma}^\mu A\sigma_\nu A^\dagger),$$

as required.

[5] A quick proof is as follows. A special case of the famous *polar decomposition theorem* tells us that any $A \in SL(2,\mathbb{C})$ may be expressed uniquely in the form $A = UP$ for U a unitary matrix and P as positive-definite Hermitian matrix (this is analogous to the modulus-argument form for a complex number). Taking the determinant, we see that $1 = \det(U)\det(P)$. The determinant $\det(U)$ is a pure phase, whilst $\det(P)$ is a positive real number, implying $1 = \det(U)$ and $1 = \det(P)$. It follows that A is the product of an element $U \in SU(2)$ and some positive definite Hermitian matrix with unit determinant. Conversely, if $A = UP$ for $U \in SU(2)$ and P positive definite Hermitian with unit determinant, we have $\det(A) = \det(U)\det(P) = 1$ so $A \in SL(2,\mathbb{C})$.

Any 2×2 Hermitian matrix can be written in the form $x^\mu \sigma_\mu$ for $x \in \mathbb{R}^4$, as we have seen in this question. The condition of having unit determinant corresponds to setting $x^2 = (x^0)^2 - \mathbf{x}^2 = 1$. Being positive definite corresponds to having $\det(x^\mu \sigma_\mu) > 0$ and $\mathrm{Tr}(x^\mu \sigma_\mu) = 2x^0 > 0$, so translates simply to x^0. It follows there is a diffeomorphism between the manifold $SL(2,\mathbb{C})$ and the product manifold $SU(2) \times \{(x^0)^2 - \mathbf{x}^2 = 1, x^0 > 0\}$. But $SU(2)$ is connected, and evidently so is the second manifold (which looks like a hyper-hyperboloid), so $SL(2,\mathbb{C})$ must be connected too (the product of connected manifolds is connected).

(d) In the final two parts of the question, we construct the inverse map which takes us from $\Lambda(A)$ to the coset $\{\pm A\}$ of $SL(2,\mathbb{C})/\mathbb{Z}_2$. Fix $\Lambda \in SO_+(1,3)$. Assuming the map we constructed above is surjective (again, the proof is possible, but long and tedious), there exists some $A(\Lambda) \in SL(2,\mathbb{C})$ such that:

$$Ax^\nu \sigma_\nu A^\dagger = \Lambda^\mu{}_\nu x^\nu \sigma_\mu$$

for all $x \in \mathbb{R}^4$ (where we suppress all Λ dependence). Since this holds for all x, we can rewrite it (as we did above) as:

$$A\sigma_\nu A^\dagger = \Lambda^\mu{}_\nu \sigma_\mu$$

We want to obtain an equation of the form $A = \cdots$, so we need to remove all free indices from the right hand side. Contracting with σ^ν on both sides, we have:

$$A\sigma_\nu A^\dagger \overline{\sigma}^\nu = \sigma_\mu \Lambda^\mu{}_\nu \overline{\sigma}^\nu.$$

We now wish to simplify $\sigma_\nu A^\dagger \overline{\sigma}^\nu$. We can use one of the results of Question 7 to establish this:

$$\begin{aligned}
\sigma_\nu A^\dagger \overline{\sigma}^\nu &= A^\dagger + \sigma_i A^\dagger \sigma^i \\
&= A^\dagger + \sigma_i A^\dagger \sigma_i \\
&= A^\dagger + (2\mathrm{Tr}(A^\dagger)I - A^\dagger) && \text{(from Question 7)} \\
&= 2\mathrm{Tr}(A^\dagger)I.
\end{aligned}$$

It follows that we can write:

$$2\mathrm{Tr}(A^\dagger)A = \sigma_\mu \Lambda^\mu{}_\nu \overline{\sigma}^\nu. \tag{$*$}$$

To finish off, we need to determine the trace $\mathrm{Tr}(A^\dagger)$ in terms of Λ. Taking the trace of the previous equation, we get:

$$2|\mathrm{Tr}(A)|^2 = \Lambda^\mu{}_\nu \mathrm{Tr}(\sigma_\mu \overline{\sigma}^\nu) = 2\Lambda^\mu{}_\mu.$$

It follows that $\mathrm{Tr}(A) = e^{i\alpha}\sqrt{\Lambda^\mu{}_\mu}$ for some phase $e^{i\alpha}$.

Thus we have established the result in the question:

$$A = \frac{e^{i\alpha}\sigma_\mu \Lambda^\mu{}_\nu \overline{\sigma}^\nu}{2\sqrt{\Lambda^\mu{}_\mu}}.$$

(e) To find $e^{i\alpha}$, we use the fact that $\det(A) = 1$. Taking the determinant of ($*$), we have:

$$\det\left(\sigma_\mu \Lambda^\mu{}_\nu \overline{\sigma}^\nu\right) = \det(2\mathrm{Tr}(A^\dagger)A) = 4\mathrm{Tr}(A^\dagger)^2 \det(A) = 4\mathrm{Tr}(A^\dagger)^2.$$

It follows that $\mathrm{Tr}(A^\dagger) = \pm\frac{1}{2}\sqrt{\det\left(\sigma_\mu \Lambda^\mu{}_\nu \overline{\sigma}^\nu\right)}$, and hence $\mathrm{Tr}(A) = \pm\frac{1}{2}\sqrt{\det\left(\sigma_\mu \Lambda^\mu{}_\nu \overline{\sigma}^\nu\right)}^*$, where $\sqrt{\cdot}$ denotes some branch of the square root function (say the principle branch, with values on the cut given by their limit taken from above the cut). Comparing with the previous formula, we see:

$$e^{i\alpha} = \pm\frac{\sqrt{\det\left(\sigma_\mu \Lambda^\mu{}_\nu \overline{\sigma}^\nu\right)}^*}{2\sqrt{\Lambda^\mu{}_\mu}},$$

so indeed the phase is determined up to a sign.

8. For a matrix Lie group G, consider the action of G on itself by conjugation, defined by $g' \mapsto gg'g^{-1}$. Show that the eigenvalues of g' and $gg'g^{-1}$ are the same for all g, so the eigenvalues are invariants of an orbit.

Find the eigenvalues of the $SU(2)$ matrix $\cos(\alpha/2)I - i\sin(\alpha/2)\hat{\boldsymbol{\alpha}} \cdot \boldsymbol{\sigma}$ where $\boldsymbol{\alpha} = \alpha\hat{\boldsymbol{\alpha}}$. Deduce the orbit structure of $SU(2)$ under the action of $SU(2)$ on itself by conjugation.

Solution: To see that g' and $gg'g^{-1}$ have the same eigenvalues, we note that they have the same characteristic equations:

$$\det(gg'g^{-1} - \lambda I) = \det(g(g' - \lambda I)g^{-1}) = \det(g)\det(g' - \lambda I)\det(g)^{-1} = \det(g' - \lambda I),$$

using the standard manipulations $\det(AB) = \det(A)\det(B)$ and $\det(A^{-1}) = \det(A)^{-1}$ of the determinant. Note that this further implies that the eigenvalues have the same algebraic multiplicities.[6]

It is possible to calculate the eigenvalues directly, but a slightly quicker method is to use exponentiation and a bit of quantum mechanics. Recall from Question 7(iii) that:

$$\cos(\alpha/2)I - i\sin(\alpha/2)\hat{\boldsymbol{\alpha}} \cdot \boldsymbol{\sigma} = \exp\left(-\frac{i\alpha}{2}\hat{\boldsymbol{\alpha}} \cdot \boldsymbol{\sigma}\right).$$

Now recall that $\hat{\boldsymbol{\alpha}} \cdot \boldsymbol{\sigma}$ is proportional to the spin operator in the $\hat{\boldsymbol{\alpha}}$ direction for a spin-$1/2$ particle (the proportionality constant is given by $\mathbf{S} = \frac{1}{2}\hbar\boldsymbol{\sigma}$), thus must have eigenvalues ± 1 to coincide with the possible physical values of the spin. It follows that the exponential operator above has eigenvalues:

$$\exp\left(\pm\frac{i\alpha}{2}\right).$$

Using the results from above, we are now asked to derive the orbit structure of $SU(2)$ under conjugation. We begin by noting that given two $SU(2)$ matrices, say $\exp(-i\boldsymbol{\alpha} \cdot \boldsymbol{\sigma}/2)$ and $\exp(-i\boldsymbol{\beta} \cdot \boldsymbol{\sigma}/2)$, their eigenvalues coincide if and only if:

$$\exp\left(\pm\frac{i\alpha}{2}\right) = \exp\left(\pm\frac{i\beta}{2}\right), \qquad \text{or} \qquad \exp\left(\pm\frac{i\alpha}{2}\right) = \exp\left(\mp\frac{i\beta}{2}\right).$$

Without loss of generality, we may take $\alpha, \beta \in [0, 2\pi]$ (over this range, we access all $SU(2)$ matrices). In the former case then, we need $\alpha = \beta + 4n\pi$ for some $n \in \mathbb{Z}$, so $\alpha = \beta$ in this case. In the latter case, we similarly have $\alpha = -\beta$.

It follows from the first result then that if $U \in SU(2)$ is any fixed matrix, we have under conjugation:

$$U\exp\left(-\frac{i\alpha}{2}\hat{\boldsymbol{\alpha}} \cdot \boldsymbol{\sigma}\right)U^{-1} = \exp\left(-\frac{i\alpha}{2}\hat{\boldsymbol{\beta}} \cdot \boldsymbol{\sigma}\right),$$

where $\hat{\boldsymbol{\beta}}$ is some unit vector (any sign in α can be absorbed into $\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}$, so we can assume without loss of generality that $\alpha \in [0, 2\pi]$ too).

To finish our construction of the orbits, we need to say which values of $\hat{\boldsymbol{\beta}}$ are accessible. A straightforward way of accomplishing this is to use the isomorphism $SU(2)/\mathbb{Z}_2 \cong SO(3)$, which we proved in the previous question. The proof of this isomorphism shows us that given any $R \in SO(3)$, there exists some matrix $U \in SU(2)$ such that:

$$U\mathbf{x} \cdot \boldsymbol{\sigma}U^{-1} = (R\mathbf{x}) \cdot \boldsymbol{\sigma}.$$

[6] In fact, it is not too difficult to prove that the eigenvalues of g' and $gg'g^{-1}$ also have the same geometric multiplicities.

In particular, this implies that:

$$\begin{aligned}
U \exp\left(-\frac{i\alpha}{2}\hat{\boldsymbol{\alpha}} \cdot \boldsymbol{\sigma}\right) U^{-1} &= \sum_{r=0}^{\infty} \frac{1}{r!} \left(-\frac{i\alpha}{2}\right)^r U(\hat{\boldsymbol{\alpha}} \cdot \boldsymbol{\sigma})^r U^{-1} \\
&= \sum_{r=0}^{\infty} \frac{1}{r!} \left(-\frac{i\alpha}{2}\right)^r (U(\hat{\boldsymbol{\alpha}} \cdot \boldsymbol{\sigma})U^{-1})^r \\
&= \sum_{r=0}^{\infty} \frac{1}{r!} \left(-\frac{i\alpha}{2}\right)^r (R\hat{\boldsymbol{\alpha}} \cdot \boldsymbol{\sigma})^r \\
&= \exp\left(-\frac{i\alpha}{2}(R\hat{\boldsymbol{\alpha}}) \cdot \boldsymbol{\sigma}\right).
\end{aligned}$$

It follows that any $\hat{\boldsymbol{\beta}}$ is accessible (just choose R such that $\hat{\boldsymbol{\beta}} = R\hat{\boldsymbol{\alpha}}$), and the orbits under conjugation are precisely:

$$\mathcal{O}_\alpha = \left\{\exp\left(-\frac{i\alpha}{2}\hat{\mathbf{n}} \cdot \boldsymbol{\sigma}\right) : \hat{\mathbf{n}} \in S^2\right\},$$

for each $\alpha \in [0, 2\pi]$.

9. (*Optional*) Let V be the fundamental representation of $SO(3)$. Recall that a *rank* r $SO(3)$*-tensor* is an element of the tensor product representation

$$V^{\otimes r} := \underbrace{V \otimes V \otimes ... \otimes V}_{r \text{ times}}.$$

We define $V^{\otimes 0} := \mathbb{C}$ to be the trivial representation of $SO(3)$. If we pick a basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ of V, then there is a natural basis $\{\mathbf{e}_{i_1} \otimes ... \otimes \mathbf{e}_{i_r} : i_1, ..., i_r = 1, 2, 3\}$ for the space $V^{\otimes r}$. In particular, given $T \in V^{\otimes r}$, we may write:

$$T = T_{i_1 i_2 ... i_r} \mathbf{e}_{i_1} \otimes ... \otimes \mathbf{e}_{i_r},$$

where $T_{i_1 i_2 ... i_r}$ are the *components* of the tensor with respect to this basis.

(a) Define a *transposition* $P_{(i,j)} : V^{\otimes r} \to V^{\otimes r}$ (with $1 \leq i < j \leq r$) of the space of rank r $SO(3)$-tensors by:

$$P_{(i,j)}(\mathbf{v}_1 \otimes ... \otimes \mathbf{v}_i \otimes ... \otimes \mathbf{v}_j \otimes ... \otimes \mathbf{v}_r) = \mathbf{v}_1 \otimes ... \otimes \mathbf{v}_j \otimes ... \otimes \mathbf{v}_i \otimes ... \otimes \mathbf{v}_r,$$

and the appropriate extension by linearity. Define a *trace* $T_{(i,j)} : V^{\otimes r} \to V^{\otimes (r-2)}$ (with $1 \leq i < j \leq r$) of the space of rank r $SO(3)$-tensors by:

$$T_{(i,j)}(\mathbf{v}_1 \otimes ... \otimes \mathbf{v}_i \otimes ... \otimes \mathbf{v}_j \otimes ... \otimes \mathbf{v}_r) = (\mathbf{v}_i \cdot \mathbf{v}_j)\, \mathbf{v}_1 \otimes ... \otimes \mathbf{v}_{i-1} \otimes \mathbf{v}_{i+1} \otimes ... \otimes \mathbf{v}_{j-1} \otimes \mathbf{v}_{j+1} \otimes ... \otimes \mathbf{v}_r,$$

and the appropriate extension by linearity. We say that a tensor $T \in V^{\otimes r}$ is *totally symmetric* if $P_{(i,j)}(T) = T$ for all $1 \leq i < j \leq r$, and we say that a tensor $T \in V^{\otimes r}$ is *totally traceless* if $T_{(i,j)}(T) = 0$ for all $1 \leq i < j \leq r$.

Show that a tensor $T \in V^{\otimes r}$ is totally symmetric and totally traceless if and only if its components with respect to some basis satisfy:

$$T_{(i_1 ... i_r)} = T_{i_1 ... i_r}, \qquad T_{kki_3 ... i_r} = 0.$$

(b) Let $W_r \subseteq V^{\otimes r}$ be the subset of totally symmetric, totally traceless tensors in $V^{\otimes r}$. Show that W_r is isomorphic to the $(2r+1)$-dimensional irreducible representation of $SO(3)$.

[Hint: First, show that W_r is an invariant subspace of $V^{\otimes r}$; therefore, it constitutes a valid representation of $SO(3)$. Next, apply the quadratic Casimir of the Lie algebra $\mathfrak{so}(3)$ to W_r and note its value. Finally, check dimensions to conclude.]

(c) Since $SO(3)$ is compact, $V^{\otimes r}$ is completely reducible. Let:

$$V^{\otimes r} = V_1 \oplus V_2 \oplus ... \oplus V_m$$

be a decomposition of $V^{\otimes r}$ into irreducibles (note that the decomposition may not be unique). By part (a), we know that for each $b = 1, ..., m$, there exists some a such that $V_b \cong W_a$. Let $\alpha : W_a \to V_b$ be an isomorphism of these two representations. Show that the components of the image $\alpha(S)$ are given by:

$$\alpha(S)_{j_1 ... j_r} = \alpha_{i_1 ... i_a j_1 ... j_r} S_{i_1 ... i_a},$$

where $\alpha_{i_1 ... i_a j_1 ... j_r}$ are the components of an $SO(3)$ invariant tensor.

(d) Hence, explain why the components T_{ij} of a general rank 2 $SO(3)$-tensor T may be decomposed as:

$$T_{ij} = \delta_{ij} S + \epsilon_{ijk} V_k + B_{ij} \qquad (*)$$

where $\delta_{ij} S$, $\epsilon_{ijk} V_k$, B_{ij} are the components of the projections of T onto irreducible subspaces of $V^{\otimes r}$, and B_{ij} is totally symmetric and totally traceless. By contracting $(*)$ with $SO(3)$ invariants, determine S, V_k and B_{ij} explicitly in terms of T_{ij}.

(e) Perform an analogous decomposition for the components of a rank 3 $SO(3)$-tensor, T_{ijk} (you should note in your construction that the decomposition is not in fact unique).

➾ **Solution:** (a) Consider expanding T with respect to some basis $\{\mathbf{e}_{a_1} \otimes ... \otimes \mathbf{e}_{a_r}\}$ for $V^{\otimes r}$. Acting with the transposition $P_{(i,j)}$, we have:

$$\begin{aligned} P_{(i,j)}(T) &= P_{(i,j)}\left(T_{a_1...a_r}\mathbf{e}_{a_1} \otimes ... \otimes \mathbf{e}_{a_i} \otimes ... \otimes \mathbf{e}_{a_j} \otimes ... \otimes \mathbf{e}_{a_r}\right) \\ &= T_{a_1...a_r}P_{(i,j)}\left(\mathbf{e}_{a_1} \otimes ... \otimes \mathbf{e}_{a_i} \otimes ... \otimes \mathbf{e}_{a_j} \otimes ... \otimes \mathbf{e}_{a_r}\right) \\ &= T_{a_1...a_r}\mathbf{e}_{a_1} \otimes ... \otimes \mathbf{e}_{a_j} \otimes ... \otimes \mathbf{e}_{a_i} \otimes ... \otimes \mathbf{e}_{a_r}, \end{aligned} \tag{$*$}$$

If T is totally symmetric, then this must be equal to T for each pair i, j. In particular, we must have $T_{a_1...a_i...a_j...a_r} = T_{a_1...a_j...a_i...a_r}$ for each pair i, j, and since transpositions generate the symmetric group, it follows that $T_{(a_1...a_r)} = T_{a_1...a_r}$. Conversely, if the components obey $T_{(a_1...a_r)} = T_{a_1...a_r}$, then we can relabel $a_j \leftrightarrow a_i$ in ($*$), and then use symmetry of the components to show $P_{(i,j)}(T) = T$ for any given pair i, j. It follows that T is totally symmetric if and only if its components with respect to some basis obey $T_{(a_1...a_r)} = T_{a_1...a_r}$.

Now consider applying the trace $T_{(i,j)}$ to T in this basis. We have:

$$\begin{aligned} T_{(i,j)}(T) &= T_{(i,j)}\left(T_{a_1...a_r}\mathbf{e}_{a_1} \otimes ... \otimes \mathbf{e}_{a_i} \otimes ... \otimes \mathbf{e}_{a_j} \otimes ... \otimes \mathbf{e}_{a_r}\right) \\ &= T_{a_1...a_r}T_{(i,j)}\left(\mathbf{e}_{a_1} \otimes ... \otimes \mathbf{e}_{a_i} \otimes ... \otimes \mathbf{e}_{a_j} \otimes ... \otimes \mathbf{e}_{a_r}\right) \\ &= T_{a_1...a_r}(\mathbf{e}_{a_i} \cdot \mathbf{e}_{a_j})\mathbf{e}_{a_1} \otimes ... \otimes \mathbf{e}_{a_{i-1}} \otimes \mathbf{e}_{a_{i+1}} \otimes ... \otimes \mathbf{e}_{a_{j-1}} \otimes \mathbf{e}_{a_{j+1}} \otimes ... \otimes \mathbf{e}_{a_r} \\ &= T_{a_1...a_r}\delta_{a_i a_j}\mathbf{e}_{a_1} \otimes ... \otimes \mathbf{e}_{a_{i-1}} \otimes \mathbf{e}_{a_{i+1}} \otimes ... \otimes \mathbf{e}_{a_{j-1}} \otimes \mathbf{e}_{a_{j+1}} \otimes ... \otimes \mathbf{e}_{a_r} \\ &= T_{a_1...a_{i-1}ka_{i+1}...a_{j-1}ka_{j+1}...a_r}\mathbf{e}_{a_1} \otimes ... \otimes \mathbf{e}_{a_{i-1}} \otimes \mathbf{e}_{a_{i+1}} \otimes ... \otimes \mathbf{e}_{a_{j-1}} \otimes \mathbf{e}_{a_{j+1}} \otimes ... \otimes \mathbf{e}_{a_r} \end{aligned} \tag{$\dagger$}$$

If T is totally traceless, then $T_{(1,2)}(T) = T$ immediately implies $T_{kka_3...a_r} = 0$ as required. Conversely, if the components of T with respect to this basis obey *both* $T_{kka_3...a_r} = 0$ *and* $T_{(a_1...a_r)} = T_{a_1...a_r}$, then the second condition implies that the first may be rewritten as $T_{a_1...a_{i-1}ka_{i+1}...a_{j-1}ka_{j+1}...a_r} = 0$ for any pair i, j. It follows that $T_{(i,j)}(T) = 0$ for all pairs i, j.

We conclude that T is totally symmetric and totally traceless if and only if the two given conditions hold for the components of the tensor with respect to some basis.

(b) We follow the advice of the hint. First, notice that W_r is indeed a vector subspace of $V^{\otimes r}$. To show this, note that $0 \in W_r$, where 0 is the zero tensor, since it is obviously totally symmetric and totally traceless; therefore, W_r is non-empty. Next, given any $S^1, S^2 \in W_r$, we can write:

$$S^1 = S^1_{i_1...i_r}\mathbf{e}_{i_1} \otimes ... \otimes \mathbf{e}_{i_r}, \qquad S^2 = S^2_{i_1...i_r}\mathbf{e}_{i_1} \otimes ... \otimes \mathbf{e}_{i_r},$$

where $S^1_{i_1..i_r}, S^2_{i_1...i_r}$ obey the conditions proved in (a). Then, given any scalars λ_1, λ_2, we have:

$$\lambda_1 S^1 + \lambda_2 S^2 = (\lambda_1 S^1_{i_1...i_r} + \lambda_2 S^2_{i_1...i_r})\mathbf{e}_{i_1} \otimes ... \otimes \mathbf{e}_{i_r}.$$

This is totally symmetric and totally traceless, since:

$$(\lambda_1 S^1 + \lambda_2 S^2)_{(i_1...i_r)} = \lambda_1 S^1_{(i_1...i_r)} + \lambda_2 S^2_{(i_1...i_r)} = \lambda_1 S^1_{i_1...i_r} + \lambda_2 S^2_{i_1...i_r} = (\lambda_1 S^1 + \lambda_2 S^2)_{i_1...i_r},$$

and similarly $(\lambda_1 S^1 + \lambda_2 S^2)_{kki_3...i_r} = 0$. Therefore, W_r is indeed a subspace of $V^{\otimes r}$.

To show that W_r is an *invariant* subspace, let $D : SO(3) \to GL(V^{\otimes r})$ be the tensor product representation on $V^{\otimes r}$. Given any $S \in W_r$, we can write:

$$S = S_{i_1...i_r}\mathbf{e}_{i_1} \otimes ... \otimes \mathbf{e}_{i_r},$$

similar to the above, with $S_{i_1...i_r}$ obeying the conditions from (a). Then given any $R \in SO(3)$, we have:

$$\begin{aligned} D(R)S &= S_{i_1...i_r}R\mathbf{e}_{i_1} \otimes ... \otimes R\mathbf{e}_{i_r} \\ &= S_{i_1...i_r}R_{j_1i_1}...R_{j_ri_r}\mathbf{e}_{j_1} \otimes ... \otimes \mathbf{e}_{j_r}. \end{aligned}$$

In particular, we have:

$$(D(R)S)_{j_1...j_r} = R_{j_1i_1}...R_{j_ri_r}S_{i_1...i_r}.$$

We can use this expression for the components of $D(R)S$ to check that $D(R)S$ is totally symmetric and totally traceless. First, we note that:

$$(D(R)S)_{(j_1...j_r)} = \frac{1}{n!}\sum_{\sigma \in S_n} R_{j_{\sigma(1)}i_1}...R_{j_{\sigma(r)}i_r}S_{i_1...i_r}.$$

Now, for each permutation $\sigma \in S_n$, we have:

$$R_{j_{\sigma(1)}i_1}...R_{j_{\sigma(r)}i_r}S_{i_1...i_r} = R_{j_{\sigma(1)}i_1}...R_{j_{\sigma(r)}i_r}S_{i_{\sigma^{-1}(1)}...i_{\sigma^{-1}(r)}},$$

by the total symmetry of the components $S_{i_1...i_r}$. In particular, relabelling $i_1 \to i_{\sigma(1)}$, $i_r \to i_{\sigma(r)}$, we have:

$$R_{j_{\sigma(1)}i_1}...R_{j_{\sigma(r)}i_r}S_{i_{\sigma^{-1}(1)}...i_{\sigma^{-1}(r)}} = R_{j_{\sigma(1)}i_{\sigma(1)}}...R_{j_{\sigma(r)}i_{\sigma(r)}}S_{i_1...i_r} = R_{j_1i_1}...R_{j_ri_r}S_{i_1...i_r}.$$

This applies for each of the terms in the expansion of $(D(R)S)_{(j_1...j_r)}$, and hence we can conclude that:

$$(D(R)S)_{(j_1...j_r)} = R_{j_1i_1}...R_{j_ri_1}S_{i_1...i_r} = (D(R)S)_{j_1...j_r}.$$

Showing the second condition, regarding the trace of the components, is a little easier:

$$\begin{aligned} (D(R)S)_{kkj_3...j_r} &= R_{ki_1}R_{ki_2}R_{j_3i_3}...R_{j_ri_r}S_{i_1...i_r} \\ &= (R^TR)_{i_1i_2}R_{j_3i_3}...R_{j_ri_r}S_{i_1...i_r} \\ &= \delta_{i_1i_2}R_{j_3i_3}...R_{j_ri_r}S_{i_1...i_r} \\ &= R_{j_3i_3}...R_{j_ri_r}S_{kki_3...i_r} \\ &= 0. \end{aligned}$$

It follows that $D(R)S$ is totally symmetric and totally traceless if S is totally symmetric and totally traceless, so indeed W_r is an invariant subspace of $V^{\otimes r}$. In particular, it constitutes a representation of $SO(3)$.

It remains to show that W_r is isomorphic to the $(2r+1)$-dimensional representation of $SO(3)$. Again, we take the advice of the question, and compute the value of the quadratic Casimir of the Lie algebra on W_r. Let $S \in W_r$ be a totally symmetric and totally traceless rank r $SO(3)$-tensor. Expressing S with respect to a basis, we write:

$$S = S_{i_1...i_r}\mathbf{e}_{i_1} \otimes ... \otimes \mathbf{e}_{i_r}$$

as above. Let $d : SO(3) \to \mathfrak{gl}(V^{\otimes r})$ be the Lie algebra representation induced by the Lie group representation $D : SO(3) \to GL(V^{\otimes r})$ above.

Now, recall from lectures that the action of d on S is given by (for each $X \in \mathfrak{so}(3)$):

$$d(X)S = \sum_{l=1}^{r} S_{i_1 \dots i_r} \mathbf{e}_{i_1} \otimes \dots \otimes \mathbf{e}_{i_{l-1}} \otimes X\mathbf{e}_{i_l} \otimes \mathbf{e}_{i_{l+1}} \otimes \dots \otimes \mathbf{e}_{i_r}$$

$$= \sum_{l=1}^{r} X_{\alpha i_l} S_{i_1 \dots i_r} \mathbf{e}_{i_1} \otimes \dots \otimes \mathbf{e}_{i_{l-1}} \otimes \mathbf{e}_{\alpha} \otimes \mathbf{e}_{i_{l+1}} \otimes \dots \otimes \mathbf{e}_{i_r}.$$

Relabelling $\alpha \leftrightarrow i_l$, we have:

$$d(X)S = \sum_{l=1}^{r} X_{i_l \alpha} S_{i_1 \dots i_{l-1} \alpha i_{l+1} \dots i_r} \mathbf{e}_{i_1} \otimes \dots \otimes \mathbf{e}_{i_r},$$

which demonstrates that the components of $d(X)S$ with respect to this basis are given by:

$$(d(X)S)_{i_1 \dots i_r} = \sum_{l=1}^{r} X_{i_l \alpha} S_{i_1 \dots i_{l-1} \alpha i_{l+1} \dots i_r}.$$

We want to find the action of the *quadratic Casimir* $\mathbf{J}^2 := d(J_a)d(J_a)$ on S. In matrix form, the Casimir is represented as $\mathbf{J}^2_{ij} = 2\delta_{ij}$ on the fundamental representation V, i.e. it has eigenvalue $2 = 1(1+1)$ everywhere. On the other hand, the generators J_a are represented by matrices with components $(J_a)_{ij} = -i\epsilon_{aij}$, as specified in Question 5. In particular, we note that:

$$(J_a)_{ij}(J_a)_{kl} = -\epsilon_{aij}\epsilon_{akl} = \delta_{il}\delta_{jk} - \delta_{ik}\delta_{jl}.$$

Now we evaluate the components of $\mathbf{J}^2 S$, which are given with respect to the standard basis above as:

$$(d(J_a)d(J_a)S)_{i_1 \dots i_r} = \sum_{l=1}^{r} (J_a)_{i_l \alpha} (d(J_a)S)_{i_1 \dots i_{l-1} \alpha i_{l+1} \dots i_r}$$

$$= \sum_{l=1}^{r} (J_a)_{i_l \alpha} \left((J_a)_{i_1 \beta} S_{\beta i_2 \dots i_{l-1} \alpha i_{l+1} \dots i_r} + \dots + (J_a)_{\alpha \beta} S_{i_1 \dots i_{l-1} \beta i_{l+1} \dots i_r} + \dots + (J_a)_{i_r \beta} S_{i_1 \dots i_{l-1} \alpha i_{l+1} \dots \beta} \right)$$

$$= \sum_{l=1}^{r} \left((J_a)_{i_l \alpha}(J_a)_{i_1 \beta} S_{\beta i_2 \dots i_{l-1} \alpha i_{l+1} \dots i_r} + \dots + \mathbf{J}^2_{i_l \beta} S_{i_1 \dots i_{l-1} \beta i_{l+1} \dots i_r} + \dots + (J_a)_{i_l \alpha}(J_a)_{i_r \beta} S_{i_1 \dots i_{l-1} \alpha i_{l+1} \dots \beta} \right)$$

$$= \sum_{l=1}^{r} \Big((\delta_{i_l \beta}\delta_{\alpha i_1} - \delta_{i_l i_1}\delta_{\alpha\beta}) S_{\beta i_2 \dots i_{l-1} \alpha i_{l+1} \dots i_r} + \dots + 2\delta_{i_l \beta} S_{i_1 \dots i_{l-1} \beta i_{l+1} \dots i_r} + \dots$$

$$\dots + (\delta_{i_l \beta}\delta_{\alpha i_r} - \delta_{i_l i_r}\delta_{\alpha\beta}) S_{i_1 \dots i_{l-1} \alpha i_{l+1} \dots \beta} \Big)$$

$$= \sum_{l=1}^{r} \left(S_{i_l i_1 \dots i_{l-1} i_1 i_{l+1} \dots i_r} + \dots + 2S_{i_1 \dots i_r} + \dots + S_{i_1 \dots i_{l-1} i_r i_{l+1} \dots i_{r-1} i_l} \right)$$

$$= \sum_{l=1}^{r} (r+1) S_{i_1 \dots i_r}$$

$$= r(r+1) S_{i_1 \dots i_r}$$

Note we used the tracelessness of the components in the fifth line to remove all appearances of $\delta_{\alpha\beta}$, and the symmetry of the components to simplify everything in the sixth line. Hence $\mathbf{J}^2 = r(r+1)\mathrm{id}_{W_r}$ on W_r.

In particular, this implies that W_r is isomorphic to some sum of irreducibles $W_r = \underline{2r+1} \oplus ... \oplus \underline{2r+1}$. To finish, we can check dimensions to show that in fact $W_r \cong \underline{2r+1}$. The space W_r consists of all totally symmetric, totally traceless tensors. Note that the components of a totally symmetric tensor S can be rearranged in the form:

$$S_{i_1...i_r} = S_{\underbrace{1...1}_{a \text{ times}}\underbrace{2...2}_{b \text{ times}}\underbrace{3...3}_{c \text{ times}}},$$

where $a+b+c=r$. In particular, this shows that not all of the components are independent; thinking about the number of possibilities using the 'stars and bars' method, we see that we want to put r indices into 3 boxes (whether they are $1, 2$ or 3). Thus there are a total of:

$$\binom{r+2}{r} = \frac{1}{2}(r+1)(r+2)$$

independent components. It follows that the space of totally symmetric tensors has dimension $\frac{1}{2}(r+1)(r+2)$.

We must now combine this with the fact that we additionally want our tensors to be totally traceless. But contracting a rank r totally symmetric tensor on any two indices, we end up with a rank $r-2$ totally symmetric tensor. Thus the number of degrees of freedom that are lost by demanding total tracelessness corresponds to the dimension of the space of rank $r-2$ totally symmetric tensors, giving:

$$\dim(W_r) = \frac{1}{2}(r+1)(r+2) - \frac{1}{2}(r-1)r = 2r+1,$$

and thus $W_r \cong \underline{2r+1}$ as required.

(c) Recall that an isomorphism of representations $\alpha : W_a \to V_b$ is such that:

$$\alpha(D_a(R)S) = D_b(R)\alpha(S). \qquad (*)$$

for all $S \in W_a$, $R \in SO(3)$, where $D_a : SO(3) \to GL(W_a)$, $D_b : SO(3) \to GL(V_b)$ are the relevant $SO(3)$ representations. Let $\{\mathbf{f}_{i_1} \otimes ... \otimes \mathbf{f}_{i_a}\}$ be a natural basis for $V^{\otimes a}$, so that $\{\mathbf{f}_{(i_1} \otimes ... \otimes \mathbf{f}_{i_a)}\}$ forms a natural basis for W_a. Let $\{\mathbf{e}_{i_1} \otimes ... \otimes \mathbf{e}_{i_r}\}$ be a natural basis for $V^{\otimes r}$, so that any element of V_b can be written with respect to this basis. Let us define the 'components of the map α', namely $\alpha_{i_1...i_a j_1...j_r}$, via the relation:

$$\alpha(\mathbf{f}_{(i_1} \otimes ... \otimes \mathbf{f}_{i_a)}) = \alpha_{i_1...i_a j_1...j_r} \mathbf{e}_{j_1} \otimes ... \otimes \mathbf{e}_{j_r},$$

where without loss of generality we take $i_1 \le i_2 \le ... \le i_a$ to ensure some fixed, defined ordering on the right hand side. We can then define $\alpha_{i_1...i_a j_1...j_r}$ for any ordering of $i_1, ..., i_a$ simply by imposing $\alpha_{(i_1...i_a)j_1...j_r} = \alpha_{i_1...i_a j_1...j_r}$.

Now consider the condition $(*)$. Writing the left hand side in terms of components, we have:

$$\begin{aligned}
\alpha(D_a(R)S) &= \alpha\left(R_{i_1 j_1}...R_{i_a j_a} S_{j_1...j_a} \mathbf{f}_{(i_a} \otimes ... \mathbf{f}_{i_a)}\right) \\
&= R_{i_1 j_1}...R_{i_a j_a} S_{j_1...j_a} \alpha\left(\mathbf{f}_{(i_1} \otimes ... \otimes \mathbf{f}_{i_a)}\right) \\
&= R_{i_1 j_1}...R_{i_a j_a} S_{j_1...j_a} \alpha_{i_1...i_a k_1...k_r} \mathbf{e}_{k_1} \otimes ... \mathbf{e}_{k_r},
\end{aligned}$$

Writing the right hand side in terms of components, we have:

$$\begin{aligned}
D_b(R)\alpha(S) &= D_b(R)\alpha\left(S_{i_1...i_a} \mathbf{f}_{i_1} \otimes ... \mathbf{f}_{i_a}\right) \\
&= S_{i_1...i_a} \alpha_{i_1...i_a j_1...j_r} D_b(R) \mathbf{e}_{j_1} \otimes ... \mathbf{e}_{j_r} \\
&= S_{i_1...i_a} \alpha_{i_1...i_a j_1...j_r} R_{k_1 j_1}...R_{k_r j_r} \mathbf{e}_{k_1} \otimes ... \otimes \mathbf{e}_{k_r}.
\end{aligned}$$

Comparing coefficients, we have (note that we can only project out the symmetric parts when we remove $S_{i_1...i_a}$, but this is fine, since it is taken care of by the total symmetry of $\alpha_{i_1...i_aj_1...j_r}$ on its first a indices):

$$\alpha_{i_1...i_aj_1...j_r}R_{k_1j_1}...R_{k_rj_r} = R_{j_1i_1}...R_{j_ai_a}\alpha_{j_1...j_ak_1...k_r},$$

which implies $R_{l_1i_1}...R_{l_ai_a}R_{k_1j_1}...R_{k_rj_r}\alpha_{i_1...i_aj_1...j_r} = \alpha_{l_1...l_ak_1...k_r}$, so that indeed $\alpha_{i_1...i_aj_1...j_r}$ are the components of an $SO(3)$ invariant tensor. The result follows.

(d) In light of (c), any projection of the components T_{ij} onto the components of an irreducible subspace will take the form $\alpha_{i_1...i_aij}S_{i_1...i_a}$ where $S_{i_1...i_a}$ is totally symmetric and totally traceless. Thus the most general possible decomposition of T_{ij} into irreducible components is:

$$T_{ij} = \alpha_{ij}S + \alpha_{i_1ij}S_{i_1} + \alpha_{i_1i_2ij}S_{i_1i_2} + \cdots$$

where $\alpha_{ij}, \alpha_{i_1ij}, \alpha_{i_1i_2ij}$, etc, are $SO(3)$ invariant tensors. First, observe that this series must terminate at rank 2 totally symmetric and totally traceless tensors; this can be seen by observing that in the tensor product $V \otimes V$, the maximum eigenvalue of $d(J_3)$ is $1+1=2$, so that the largest possible irreducible subspace is the 5-dimensional representation of $SO(3)$ (which has $d(J_3)$ eigenvalues $\{-2,-1,0,1,2\}$. More generally, the decomposition of a rank r $SO(3)$-tensor's components into irreducible components must terminate at rank r totally symmetric and totally traceless tensors.

We can say more. Since $W_r \leq V^{\otimes r}$, we can immediately identify W_r as an invariant subspace of $V^{\otimes r}$. It is even the *unique* copy of $\underline{2r+1}$ inside $V^{\otimes r}$, since there is a unique highest-weight eigenvector in $V^{\otimes r}$, given by $|1\rangle^{\otimes r}$ (where we label the basis elements of V as $\{|-1\rangle, |0\rangle, |1\rangle\}$ in terms of their $d(J_3)$ eigenvalue). Thus, in our rank-2 case, we must be able to write $\alpha_{i_1i_2ij}S_{i_1i_2} = B_{ij}$ where B_{ij} is totally symmetric and totally traceless. Clearly we have $B_{ij} = T_{(ij)} - \frac{1}{3}\delta_{ij}T_{kk}$, the projection of the components of T_{ij} onto the invariant subspace of totally symmetric and totally traceless tensors.

To sort out the other terms, it is necessary to use.a general fact from invariant theory: all invariants of $SO(3)$ can be constructed by taking tensor products of the Kronecker delta and the Levi-Civita symbol. In particular, this implies that $\alpha_{ij} \propto \delta_{ij}$ and $\alpha_{i_1ij} \propto \epsilon_{i_1ij}$.

Putting all this information together, we see that we can write:

$$T_{ij} = \delta_{ij}S + \epsilon_{ijk}V_k + B_{ij},$$

where $B_{ij} = T_{(ij)} - \frac{1}{3}\delta ijT_{kk}$ is totally symmetric and totally traceless, as required.

To obtain S, V_k and B_{ij} explicitly, we can contract with invariant tensors. Contracting with δ_{ij}, we have:

$$\delta_{ij}T_{ij} = 3S \qquad \Rightarrow \qquad S = \frac{1}{3}T_{kk}.$$

Contracting with ϵ_{ijl}, we have:

$$\epsilon_{ijl}T_{ij} = \epsilon_{ijl}\epsilon_{ijk}V_k = 2\delta_{kl}V_k = 2V_l \qquad \Rightarrow \qquad V_l = \frac{1}{2}\epsilon_{ijl}T_{ij}.$$

This completes the decomposition, as required.

(e) Here, the construction is similar, but more involved. By the discussion in part (d), we know that T_{ijk} can be separated into irreducible components via a decomposition of the form:

$$T_{ijk} = \alpha_{ijk}S + \alpha_{i_1ijk}S_{i_1} + \alpha_{i_1i_2ijk}S_{i_1i_2} + \alpha_{i_1i_2i_3ijk}S_{i_1i_2i_3}.$$

Invariant theory dictates that $\alpha_{ijk} \propto \epsilon_{ijk}$. Similarly, we must have:

$$\alpha_{i_1ijk} = \alpha\delta_{i_1i}\delta_{jk} + \beta\delta_{i_1j}\delta_{ik} + \gamma\delta_{i_1k}\delta_{ij},$$

for some α, β, γ and:

$$\alpha_{i_1i_2ijk} = a\delta_{i_1i}\epsilon_{i_2jk} + b\delta_{i_1j}\epsilon_{i_2ik} + c\delta_{i_1k}\epsilon_{i_2ij}$$

for some a, b, c (note we needn't include the term involving $\delta_{i_1i_2}$, or the terms involving i_1, i_2 as indices on the Levi-Civita symbol, because of the contraction with $S_{i_1i_2}$ which is totally symmetric and totally traceless; we can also avoid all terms like $\delta_{i_2i}\epsilon_{i_1jk}$ etc, because we contract with $S_{i_1i_2}$ which is symmetric on i_1, i_2). Thus, we can write T_{ijk} in the form:

$$T_{ijk} = \epsilon_{ijk}S + \delta_{jk}V_i^{(1)} + \delta_{ki}V_j^{(2)} + \delta_{ij}V_k^{(3)} + \epsilon_{jka}B_{ai}^{(1)} + \epsilon_{kia}B_{aj}^{(2)} + \epsilon_{ija}B_{ak}^{(3)} + D_{ijk}.$$

where $B_{ij}^{(1)}, B_{ij}^{(2)}, B_{ij}^{(3)}$ and D_{ijk} are all totally symmetric and totally traceless.

To obtain the irreducibles, we contract with invariant tensors. We have:

- Contracting with ϵ_{ijk}, we have:

$$\epsilon_{ijk}T_{ijk} = 6S \qquad \Rightarrow \qquad S = \frac{1}{6}\epsilon_{ijk}T_{ijk}.$$

- Contracting with $\delta_{jk}, \delta_{ki}, \delta_{ij}$ in turn, we derive the equations:

$$\delta_{jk}T_{ajk} = 3V_a^{(1)} + V_a^{(2)} + V_a^{(3)}$$

$$\delta_{ki}T_{iak} = V_a^{(1)} + 3V_a^{(2)} + V_a^{(3)}$$

$$\delta_{ij}T_{ija} = V_a^{(1)} + V_a^{(2)} + 3V_a^{(3)}.$$

  Solving this system of equations simultaneously, we derive:

$$V_a^{(1)} = \frac{2}{5}\delta_{jk}T_{ajk} - \frac{1}{10}\delta_{ki}T_{iak} - \frac{1}{10}\delta_{ij}T_{ija} = \frac{1}{2}T_{akk} - \frac{3}{5}T_{(akk)},$$

$$V_a^{(2)} = -\frac{1}{10}\delta_{jk}T_{ajk} + \frac{2}{5}\delta_{ki}T_{iak} - \frac{1}{10}\delta_{ij}T_{ija} = \frac{1}{2}T_{kak} - \frac{3}{5}T_{(akk)},$$

$$V_a^{(3)} = -\frac{1}{10}\delta_{jk}T_{ajk} - \frac{1}{10}\delta_{ki}T_{iak} + \frac{2}{5}\delta_{ij}T_{ija} = \frac{1}{2}T_{kka} - \frac{3}{5}T_{(akk)}.$$

- We now contract with ϵ_{jkb}, ϵ_{kib} and ϵ_{ijb} respectively. We find that:

$$\epsilon_{jkb}T_{cjk} - 2\delta_{bc}S - \epsilon_{jcb}V_j^{(2)} - \epsilon_{ckb}V_k^{(3)} = 2B_{bc}^{(1)} - B_{bc}^{(2)} - B_{bc}^{(3)}$$

$$\epsilon_{kib}T_{ick} - 2\delta_{bc}S - \epsilon_{cib}V_i^{(1)} - \epsilon_{kcb}V_k^{(3)} = -B_{bc}^{(1)} + 2B_{bc}^{(2)} - B_{bc}^{(3)}$$

$$\epsilon_{ijb}T_{ijc} - 2\delta_{bc}S - \epsilon_{icb}V_i^{(1)} - \epsilon_{cjb}V_j^{(2)} = -B_{bc}^{(1)} - B_{bc}^{(2)} + 2B_{bc}^{(3)}.$$

To simplify these equations, it is useful to note that the right hand sides are all symmetric on b, c and traceless on b, c. Thus symmetrising the left hand sides on b, c, we end up with the system of equations:

$$\epsilon_{jkb}T_{cjk} + \epsilon_{jkc}T_{bjk} - 4\delta_{bc}S = 4B^{(1)}_{bc} - 2B^{(2)}_{bc} - 2B^{(3)}_{bc}$$

$$\epsilon_{kib}T_{ick} + \epsilon_{kic}T_{ibk} - 4\delta_{bc}S = -2B^{(1)}_{bc} + 4B^{(2)}_{bc} - 2B^{(3)}_{bc}$$

$$\epsilon_{ijb}T_{ijc} + \epsilon_{ijc}T_{ijb} - 4\delta_{bc}S = -2B^{(1)}_{bc} - 2B^{(2)}_{bc} + 4B^{(3)}_{bc}.$$

We note that the equations are not independent. For them to be consistent, we require that the sum of the left hand sides is zero. Relabelling, it is given by:

$$\begin{aligned}
\text{LHS} &= \epsilon_{ijb}T_{cij} + \epsilon_{ijc}T_{bij} + \epsilon_{ijb}T_{jci} + \epsilon_{ijc}T_{jbi} + \epsilon_{ijb}T_{ijc} + \epsilon_{ijc}T_{ijb} - 12\delta_{bc}S \\
&= \epsilon_{ijb}(T_{cij} - T_{icj} + T_{ijc}) + \epsilon_{ijc}(T_{bij} - T_{ibj} + T_{ijb}) - 12\delta_{bc}S \\
&= \frac{1}{2}\epsilon_{ijb}(T_{cij} - T_{cji} - T_{icj} + T_{jci} + T_{ijc} - T_{jic}) + \frac{1}{2}\epsilon_{ijc}(T_{bij} - T_{bji} - T_{ibj} + T_{jbi} + T_{ijb} - T_{jib}) - 12\delta_{bc}S \\
&= 3\epsilon_{ijb}T_{[ijc]} + 3\epsilon_{ijc}T_{[ijb]} - 12\delta_{bc}S.
\end{aligned}$$

But $2\delta_{bc} = \epsilon_{ijb}\epsilon_{ijc}$. Thus we have:

$$3\epsilon_{ijb}T_{[ijc]} + 3\epsilon_{ijc}T_{[ijb]} - 12\delta_{bc}S = 3\epsilon_{ijb}(T_{[ijc]} - \epsilon_{ijc}S) + 3\epsilon_{ijc}(T_{[ijb]} - \epsilon_{ijb}S).$$

Now:

$$\epsilon_{ijc}S = \frac{1}{6}\epsilon_{ijc}\epsilon_{\alpha\beta\gamma}T_{\alpha\beta\gamma} = T_{[ijc]},$$

by thinking about what the last equality means. Thus indeed we have that the equations are consistent. They reduce to a pair:

$$\epsilon_{jkb}T_{cjk} + \epsilon_{jkc}T_{bjk} - 4\delta_{bc}S = 4B^{(1)}_{bc} - 2B^{(2)}_{bc} - 2B^{(3)}_{bc}$$

$$\epsilon_{kib}T_{ick} + \epsilon_{kic}T_{ibk} - 4\delta_{bc}S = -2B^{(1)}_{bc} + 4B^{(2)}_{bc} - 2B^{(3)}_{bc},$$

since the last equation gives us no additional information. We are now free to choose $B^{(3)}_{bc}$; it is an arbitrary symmetric, traceless tensor. We will use it to parametrise $B^{(1)}_{bc}$ and $B^{(2)}_{bc}$. Thus we wish to solve the simultaneous equations:

$$\epsilon_{jkb}T_{cjk} + \epsilon_{jkc}T_{bjk} - 4\delta_{bc}S + 2B^{(3)}_{bc} = 4B^{(1)}_{bc} - 2B^{(2)}_{bc}$$

$$\epsilon_{kib}T_{ick} + \epsilon_{kic}T_{ibk} - 4\delta_{bc}S + 2B^{(3)}_{bc} = -2B^{(1)}_{bc} + 4B^{(2)}_{bc}.$$

The solution is given by:

$$B^{(1)}_{bc} = \frac{1}{3}\left(\epsilon_{jkb}T_{cjk} + \epsilon_{jkc}T_{bjk} - 4\delta_{bc}S + 2B^{(3)}_{bc}\right) + \frac{1}{6}\left(\epsilon_{kib}T_{ick} + \epsilon_{kic}T_{ibk} - 4\delta_{bc}S + 2B^{(3)}_{bc}\right)$$

$$B^{(2)}_{bc} = \frac{1}{6}\left(\epsilon_{jkb}T_{cjk} + \epsilon_{jkc}T_{bjk} - 4\delta_{bc}S + 2B^{(3)}_{bc}\right) + \frac{1}{3}\left(\epsilon_{kib}T_{ick} + \epsilon_{kic}T_{ibk} - 4\delta_{bc}S + 2B^{(3)}_{bc}\right),$$

which simplifies to (via some relabelling):

$$B^{(1)}_{bc} = \frac{1}{3}\left(\epsilon_{ijb}T_{cij} + \epsilon_{ijc}T_{bij} + \epsilon_{ijb}T_{jci} + \epsilon_{ijc}T_{jbi} - 6\delta_{bc}S\right) - \frac{1}{6}\left(\epsilon_{ijb}T_{jci} + \epsilon_{ijc}T_{jbi}\right) + B^{(3)}_{bc}$$

$$B^{(2)}_{bc} = \frac{1}{3}\left(\epsilon_{ijb}T_{cij} + \epsilon_{ijc}T_{bij} + \epsilon_{ijb}T_{jci} + \epsilon_{ijc}T_{jbi} - 6\delta_{bc}S\right) - \frac{1}{6}\left(\epsilon_{ijb}T_{cij} + \epsilon_{ijc}T_{bij}\right) + B^{(3)}_{bc},$$

- Finally, we have $D_{ijk} = T_{(ijk)} - \frac{1}{5}(\delta_{ij}T_{(aak)} + \delta_{ik}T_{(aja)} + \delta_{jk}T_{(iaa)})$, arguing in the same way as in (d).

This completes the decomposition of T_{ijk} into invariant subspaces - phew!

Part III: Symmetries, Fields and Particles
Examples Sheet 3 Solutions

Please send all comments and corrections to mjb318@cam.ac.uk and jmm232@cam.ac.uk.

1. *Schur's Lemma* is the following: 'Let D be an irreducible representation of a group G acting on a complex vector space V. Let A be an operator acting on V which commutes with the action of G, that is, $AD(g) = D(g)A$ for all $g \in G$. Then $A = \lambda I_V$ where λ is a constant and I_V is the identity operator.'

Prove this by showing that any eigenspace of A is an invariant subspace of V, and that there is therefore precisely one eigenspace of A which is the whole of V, and that this gives the desired result.

➾ **Solution:** Since V is a *complex* vector space, A has some eigenvalue λ. Let E_λ be the corresponding eigenspace, which has dimension at least 1. Given $v \in E_\lambda$, we have:

$$AD(g)v = D(g)Av = D(g)\lambda v = \lambda D(g)v,$$

so $D(g)v \in E_\lambda$. Thus E_λ is an invariant subspace of V. But E_λ has dimension at least 1, so $V = E_\lambda$ by irreducibility. □

---

2. The following multiplication rule will be useful in this question (cf. Sheet 1, 4c):

$$(aI + \mathbf{b}\cdot\boldsymbol{\sigma})\left(a'I + \mathbf{b}'\cdot\boldsymbol{\sigma}\right) = (aa' + \mathbf{b}\cdot\mathbf{b}')I + (a\mathbf{b}' + a'\mathbf{b} + i\mathbf{b}\times\mathbf{b}')\cdot\boldsymbol{\sigma}.$$

(a) Show how $B(\theta, \hat{\mathbf{n}}) \in SL(2,\mathbb{C})$, where

$$B(\theta,\hat{\mathbf{n}}) = I\cosh\left(\frac{\theta}{2}\right) + \boldsymbol{\sigma}\cdot\hat{\mathbf{n}}\sinh\left(\frac{\theta}{2}\right), \qquad \hat{\mathbf{n}}^2 = 1,$$

corresponds to a Lorentz boost with velocity $\mathbf{v} = \tanh(\theta)\hat{\mathbf{n}}$.

(b) Show that:

$$\left(I + \frac{1}{2}\boldsymbol{\sigma}\cdot\delta\mathbf{v}\right)B(\theta,\hat{\mathbf{n}}) = B(\theta',\hat{\mathbf{n}}')R,$$

where to first order in $\delta\mathbf{v}$,

$$\theta' = \theta + \delta\mathbf{v}\cdot\hat{\mathbf{n}}, \qquad \hat{\mathbf{n}}' = \hat{\mathbf{n}} + \coth(\theta)\left(\delta\mathbf{v} - \hat{\mathbf{n}}\hat{\mathbf{n}}\cdot\delta\mathbf{v}\right),$$

and R is an infinitesimal rotation given by:

$$R = I + \frac{i}{2}\tanh\left(\frac{\theta}{2}\right)(\delta\mathbf{v}\times\hat{\mathbf{n}})\cdot\boldsymbol{\sigma} = I + \frac{i}{2}\frac{\gamma}{\gamma+1}(\delta\mathbf{v}\times\mathbf{v})\cdot\boldsymbol{\sigma}, \qquad \gamma = \left(1-\mathbf{v}^2\right)^{-\frac{1}{2}}.$$

(c) Show that we must have $\mathbf{v}' = \mathbf{v} + \delta\mathbf{v} - (\mathbf{v}\cdot\delta\mathbf{v})\mathbf{v}$. [NB $(\mathbf{a}\cdot\boldsymbol{\sigma})(\mathbf{b}\cdot\boldsymbol{\sigma}) = (\mathbf{a}\cdot\mathbf{b})I + i\boldsymbol{\sigma}\cdot(\mathbf{a}\times\mathbf{b})$.]

(d) By considering boosts by velocities $\mathbf{v}$, $\mathbf{w}$ followed by boosts by $-\mathbf{w}$, $-\mathbf{v}$, find a physical interpretation of this question.

➾ **Solution:** (a) We consider the action of $B(\theta,\hat{\mathbf{n}})$ on $x^\mu\sigma_\mu$, where $\sigma_\mu = (I, \boldsymbol{\sigma})$, defined by:

$$x^\mu\sigma_\mu \mapsto B(\theta,\hat{\mathbf{n}})x^\mu\sigma_\mu B(\theta,\hat{\mathbf{n}})^\dagger.$$

It is sufficient to consider the action on each of $I, \boldsymbol{\sigma}$ separately in turn. Let's also introduce the notation $c = \cosh(\theta/2)$, $s = \sinh(\theta/2)$ to save on writing. Then for I, we have:

$$
\begin{aligned}
B(\theta, \hat{\mathbf{n}}) I B(\theta, \hat{\mathbf{n}})^\dagger &= (cI + s\boldsymbol{\sigma}\cdot\hat{\mathbf{n}})\,(cI + s\boldsymbol{\sigma}\cdot\hat{\mathbf{n}}) \\
&= \left(c^2 + s^2\hat{\mathbf{n}}^2\right) I + 2cs(\boldsymbol{\sigma}\cdot\hat{\mathbf{n}}) \\
&= I\cosh(\theta) + \boldsymbol{\sigma}\cdot\hat{\mathbf{n}}\sinh(\theta),
\end{aligned}
$$

where in the last line we used $\cosh(\theta) = c^2 + s^2 = \cosh^2(\theta/2) + \sinh^2(\theta/2), \sinh(\theta) = 2cs = 2\cosh(\theta/2)\sinh(\theta/2)$.

Now consider the transformation of $\mathbf{a}\cdot\boldsymbol{\sigma}$ under $B(\theta, \hat{\mathbf{n}})$:

$$
\begin{aligned}
B(\theta, \hat{\mathbf{n}})\mathbf{a}\cdot\boldsymbol{\sigma} B(\theta, \hat{\mathbf{n}})^\dagger &= (cI + s\boldsymbol{\sigma}\cdot\hat{\mathbf{n}})\,\mathbf{a}\cdot\boldsymbol{\sigma}\,(cI + s\boldsymbol{\sigma}\cdot\hat{\mathbf{n}}) \\
&= (s\mathbf{a}\cdot\hat{\mathbf{n}}I + (c\mathbf{a} + is\hat{\mathbf{n}}\times\mathbf{a})\cdot\boldsymbol{\sigma})\,(cI + s\boldsymbol{\sigma}\cdot\hat{\mathbf{n}}) \\
&= 2\mathbf{a}\cdot\hat{\mathbf{n}}csI + \left(s^2(\mathbf{a}\cdot\hat{\mathbf{n}})\hat{\mathbf{n}} + c^2\mathbf{a} + ics\hat{\mathbf{n}}\times\mathbf{a} + ics\mathbf{a}\times\hat{\mathbf{n}} - s^2(\hat{\mathbf{n}}\times\mathbf{a})\times\hat{\mathbf{n}}\right)\cdot\boldsymbol{\sigma} \\
&= \mathbf{a}\cdot\hat{\mathbf{n}}\sinh(\theta)I + \left(s^2(\mathbf{a}\cdot\hat{\mathbf{n}})\hat{\mathbf{n}} + c^2\mathbf{a} + s^2(\hat{\mathbf{n}}\cdot\mathbf{a})\hat{\mathbf{n}} - s^2\mathbf{a}\right)\cdot\boldsymbol{\sigma} \\
&= \mathbf{a}\cdot\hat{\mathbf{n}}\sinh(\theta)I + (\mathbf{a} + (\cosh(\theta) - 1)(\mathbf{a}\cdot\hat{\mathbf{n}})\hat{\mathbf{n}})\cdot\boldsymbol{\sigma}
\end{aligned}
$$

where in the second-from-last line we used *Lagrange's formula* for the vector triple product: $\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) = (\mathbf{a}\cdot\mathbf{c})\mathbf{b} - (\mathbf{a}\cdot\mathbf{b})\mathbf{c}$. Putting everything together, we see that $B(\theta, \hat{\mathbf{n}})x^\mu\sigma_\mu B(\theta, \hat{\mathbf{n}})^\dagger = x'^\mu\sigma_\mu$, with x' given by:

$$
x'^\mu = \begin{pmatrix} \cosh(\theta) & \hat{\mathbf{n}}^T\sinh(\theta) \\ \hat{\mathbf{n}}\sinh(\theta) & I_3 + (\cosh(\theta) - 1)\hat{\mathbf{n}}\hat{\mathbf{n}}^T \end{pmatrix} x^\mu,
$$

which indeed corresponds to a Lorentz transformation with rapidity θ and direction $\hat{\mathbf{n}}$, i.e. velocity $\mathbf{v} = \tanh(\theta)\hat{\mathbf{n}}$.

(b) In this part of the question, we are asked to show:

$$
\left(I + \frac{1}{2}\boldsymbol{\sigma}\cdot\delta\mathbf{v}\right) B(\theta, \hat{\mathbf{n}}) = B(\theta', \hat{\mathbf{n}}')R,
$$

where θ', $\hat{\mathbf{n}}'$ and R are all appropriately defined. We simply check that both sides are equal to one another. On the left hand side, we have:

$$
\begin{aligned}
\left(I + \frac{1}{2}\boldsymbol{\sigma}\cdot\delta\mathbf{v}\right) B(\theta, \hat{\mathbf{n}}) &= \left(I + \frac{1}{2}\boldsymbol{\sigma}\cdot\delta\mathbf{v}\right)(cI + s\boldsymbol{\sigma}\cdot\hat{\mathbf{n}}) \\
&= \left(c + \frac{1}{2}s\delta\mathbf{v}\cdot\hat{\mathbf{n}}\right) I + \left(s\hat{\mathbf{n}} + \frac{1}{2}c\delta\mathbf{v} + \frac{1}{2}is\delta\mathbf{v}\times\hat{\mathbf{n}}\right)\cdot\boldsymbol{\sigma}
\end{aligned}
$$

Let's compare this with the expansion of the other side, to first order in $\delta\mathbf{v}$. First we must work out $B(\theta', \hat{\mathbf{n}}')$ to first order in $\delta\mathbf{v}$. We have:

$$
\begin{aligned}
B(\theta', \hat{\mathbf{n}}') &= I\cosh\left(\frac{\theta'}{2}\right) + \boldsymbol{\sigma}\cdot\hat{\mathbf{n}}'\sinh\left(\frac{\theta'}{2}\right) \\
&= I\cosh\left(\frac{\theta + \delta\mathbf{v}\cdot\hat{\mathbf{n}}}{2}\right) + \boldsymbol{\sigma}\cdot(\hat{\mathbf{n}} + \coth(\theta)(\delta\mathbf{v} - \hat{\mathbf{n}}(\hat{\mathbf{n}}\cdot\delta\mathbf{v})))\sinh\left(\frac{\theta + \delta\mathbf{v}\cdot\hat{\mathbf{n}}}{2}\right)
\end{aligned}
$$

$$
\begin{aligned}
&= I\left(\cosh\left(\frac{\theta}{2}\right) + \frac{1}{2}\delta\mathbf{v}\cdot\hat{\mathbf{n}}\sinh\left(\frac{\theta}{2}\right)\right) + \boldsymbol{\sigma}\cdot\hat{\mathbf{n}}\sinh\left(\frac{\theta}{2}\right) + \frac{1}{2}(\boldsymbol{\sigma}\cdot\hat{\mathbf{n}})(\delta\mathbf{v}\cdot\hat{\mathbf{n}})\cosh\left(\frac{\theta}{2}\right)\\
&\quad + (\boldsymbol{\sigma}\cdot\delta\mathbf{v} - (\boldsymbol{\sigma}\cdot\hat{\mathbf{n}})(\delta\mathbf{v}\cdot\hat{\mathbf{n}}))\coth(\theta)\sinh\left(\frac{\theta}{2}\right) + O(\delta\mathbf{v}^2).
\end{aligned}
$$

We can simplify this slightly by using the identity:

$$
\coth(\theta)\sinh\left(\frac{\theta}{2}\right) = \frac{\cosh(\theta)\sin(\theta/2)}{\sinh(\theta)} = \frac{\cosh^2(\theta/2)+\sinh^2(\theta/2)}{2\cosh(\theta/2)} = \frac{1}{2}\cosh\left(\frac{\theta}{2}\right) + \frac{1}{2}\tanh\left(\frac{\theta}{2}\right)\sinh\left(\frac{\theta}{2}\right).
$$

Hence, writing $c = \cosh(\theta/2)$, $s = \sinh(\theta/2)$ and $s/c = \tanh(\theta/2)$, we have:

$$
\begin{aligned}
B(\theta',\hat{\mathbf{n}}') &= \left(c + \frac{1}{2}s\delta\mathbf{v}\cdot\hat{\mathbf{n}}\right)I + s\boldsymbol{\sigma}\cdot\hat{\mathbf{n}} + \frac{1}{2}c(\boldsymbol{\sigma}\cdot\hat{\mathbf{n}})(\delta\mathbf{v}\cdot\hat{\mathbf{n}}) + \frac{1}{2}\left(\boldsymbol{\sigma}\cdot\delta\mathbf{v} - (\boldsymbol{\sigma}\cdot\hat{\mathbf{n}})(\delta\mathbf{v}\cdot\hat{\mathbf{n}})\right)\left(c + \frac{s^2}{c}\right) + O(\delta\mathbf{v}^2)\\
&= \left(c + \frac{1}{2}s\delta\mathbf{v}\cdot\hat{\mathbf{n}}\right)I + \left(s\hat{\mathbf{n}} + \frac{1}{2}c\delta\mathbf{v} + \frac{s^2}{2c}(\delta\mathbf{v} - \hat{\mathbf{n}}(\delta\mathbf{v}\cdot\hat{\mathbf{n}}))\right)\cdot\boldsymbol{\sigma} + O(\delta\mathbf{v}^2)
\end{aligned}
$$

We must multiply this expression by the rotation:

$$
R = I + \frac{i}{2}\tanh\left(\frac{\theta}{2}\right)(\delta\mathbf{v}\times\hat{\mathbf{n}})\cdot\boldsymbol{\sigma} = I + \frac{is}{2c}(\delta\mathbf{v}\times\hat{\mathbf{n}})\cdot\boldsymbol{\sigma}.
$$

We find:

$$
\begin{aligned}
B(\theta',\hat{\mathbf{n}}')R &= \left[\left(c + \frac{1}{2}s\delta\mathbf{v}\cdot\hat{\mathbf{n}}\right)I + \left(s\hat{\mathbf{n}} + \frac{1}{2}c\delta\mathbf{v} + \frac{s^2}{2c}(\delta\mathbf{v} - \hat{\mathbf{n}}(\delta\mathbf{v}\cdot\hat{\mathbf{n}}))\right)\cdot\boldsymbol{\sigma}\right]\left[I + \frac{is}{2c}(\delta\mathbf{v}\times\hat{\mathbf{n}})\cdot\boldsymbol{\sigma}\right]\\
&= \left(c + \frac{1}{2}s\delta\mathbf{v}\cdot\hat{\mathbf{n}}\right)I + \left(\frac{1}{2}is(\delta\mathbf{v}\times\hat{\mathbf{n}}) + s\hat{\mathbf{n}} + \frac{1}{2}c\delta\mathbf{v} + \frac{s^2}{2c}(\delta\mathbf{v} - \hat{\mathbf{n}}(\delta\mathbf{v}\cdot\hat{\mathbf{n}})) - \frac{s^2}{2c}\hat{\mathbf{n}}\times(\delta\mathbf{v}\times\hat{\mathbf{n}})\right)\cdot\boldsymbol{\sigma}\\
&= \left(c + \frac{1}{2}s\delta\mathbf{v}\cdot\hat{\mathbf{n}}\right)I + \left(\frac{1}{2}is(\delta\mathbf{v}\times\hat{\mathbf{n}}) + s\hat{\mathbf{n}} + \frac{1}{2}c\delta\mathbf{v}\right)\cdot\boldsymbol{\sigma},
\end{aligned}
$$

where in the last line, we noted that $\hat{\mathbf{n}}\times(\delta\mathbf{v}\times\hat{\mathbf{n}}) = \delta\mathbf{v} - \hat{\mathbf{n}}(\delta\mathbf{v}\cdot\hat{\mathbf{n}})$ by Lagrange's formula for the vector triple product. Hence we have indeed verified the identity:

$$
\left(I + \frac{1}{2}\boldsymbol{\sigma}\cdot\delta\mathbf{v}\right)B(\theta,\hat{\mathbf{n}}) = B(\theta',\hat{\mathbf{n}}')R.
$$

We are also given an alternative form of the rotation; we will briefly check that this is equivalent. Starting from the right hand side, we note that $\gamma = (1-\mathbf{v}^2)^{-1/2} = (1-\tanh^2(\theta))^{-1/2} = \cosh(\theta)$, and hence:

$$
R = I + \frac{i}{2}\frac{\gamma}{\gamma+1}(\delta\mathbf{v}\times\mathbf{v})\cdot\boldsymbol{\sigma} = I + \frac{i}{2}\frac{\cosh(\theta)\tanh(\theta)}{\cosh(\theta)+1}(\delta\mathbf{v}\times\hat{\mathbf{n}})\cdot\boldsymbol{\sigma} = I + \frac{i}{2}\frac{\sinh(\theta)}{\cosh(\theta)+1}(\delta\mathbf{v}\times\hat{\mathbf{n}})\cdot\boldsymbol{\sigma}.
$$

Noting the identity:

$$
\frac{\sinh(\theta)}{\cosh(\theta)+1} = \frac{2\sinh(\theta/2)\cosh(\theta/2)}{\cosh^2(\theta/2)+\sinh^2(\theta/2)+1} = \frac{2\sinh(\theta/2)\cosh(\theta/2)}{2\cosh^2(\theta/2)} = \tanh\left(\frac{\theta}{2}\right),
$$

we see that the two given forms of the rotation R are equivalent.

(c) The velocity of the primed boost is $\mathbf{v}' = \tanh(\theta')\hat{\mathbf{n}}'$. Hence inserting the formula for θ' and $\hat{\mathbf{n}}'$, we have:

$$\begin{aligned}
\mathbf{v}' &= \tanh\left(\theta + \delta\mathbf{v}\cdot\hat{\mathbf{n}}\right)\left(\hat{\mathbf{n}} + \coth(\theta)(\delta\mathbf{v} - (\delta\mathbf{v}\cdot\hat{\mathbf{n}})\hat{\mathbf{n}})\right) \\
&= \tanh(\theta)\hat{\mathbf{n}} + (\delta\mathbf{v}\cdot\hat{\mathbf{n}})\mathrm{sech}^2(\theta)\hat{\mathbf{n}} + \delta\mathbf{v} - (\delta\mathbf{v}\cdot\hat{\mathbf{n}})\hat{\mathbf{n}} \\
&= \mathbf{v} + \delta\mathbf{v} + \left(\mathrm{sech}^2(\theta) - 1\right)(\delta\mathbf{v}\cdot\hat{\mathbf{n}})\hat{\mathbf{n}} && (\text{since } \mathbf{v} = \tanh(\theta)\hat{\mathbf{n}}) \\
&= \mathbf{v} + \delta\mathbf{v} - \tanh^2(\theta)(\delta\mathbf{v}\cdot\hat{\mathbf{n}})\hat{\mathbf{n}} && (\text{using } \mathrm{sech}^2(\theta) - 1 = -\tanh^2(\theta)) \\
&= \mathbf{v} + \delta\mathbf{v} - (\delta\mathbf{v}\cdot\mathbf{v})\mathbf{v} && (\text{since } \mathbf{v} = \tanh(\theta)\hat{\mathbf{n}})
\end{aligned}$$

as required.

(d) This question involved a lot of algebra, but actually tells us something interesting about the geometry of Lorentz boosts and rotations. Let's forget what we did in the question, and try to compute the result of:

$$B(\mathbf{w})B(\mathbf{v}),$$

that is, a boost by velocity $\mathbf{v}$ followed by a boost by velocity $\mathbf{w}$. If we assume we start from the rest frame S, then we know from basic special relativity that the resulting doubly-boosted frame S' appears to move with a new velocity $\mathbf{v}\oplus\mathbf{w}$ relative to the rest frame S, with $\mathbf{v}\oplus\mathbf{w}$ given by:

$$\mathbf{v}\oplus\mathbf{w} = \frac{1}{1+\mathbf{v}\cdot\mathbf{w}}\left(\left(1 + \frac{\gamma_{\mathbf{v}}}{1+\gamma_{\mathbf{v}}}\mathbf{v}\cdot\mathbf{w}\right)\mathbf{v} + \frac{1}{\gamma_{\mathbf{v}}}\mathbf{w}\right),$$

where $\gamma_{\mathbf{v}} = (1-\mathbf{v}^2)^{-1/2}$ is the standard gamma-factor. This is the classic formula for the *addition of relativistic velocities*.

However, perhaps surprisingly, the resulting frame S' is *not* the boost of the rest frame S by $\mathbf{v}\oplus\mathbf{w}$, i.e.

$$B(\mathbf{w})B(\mathbf{v}) \neq B(\mathbf{v}\oplus\mathbf{w}).$$

Indeed, this *cannot* be the case for the following reason. Suppose that we did a boost by $\mathbf{v}$, followed by a boost by $\mathbf{w}$. We suppose, aiming for a contradiction, that the result is a boost of velocity $\mathbf{v}\oplus\mathbf{w}$. Now let's think about what happens when we reverse the boosts. We could undo them one at a time, boosting by $-\mathbf{w}$, then by $-\mathbf{v}$. By assumption, this is equivalent to the boost by $(-\mathbf{w})\oplus(-\mathbf{v})$. But this also must be the same as undoing everything altogether, i.e. boosting by $-\mathbf{v}\oplus\mathbf{w}$. This suggests the identity:

$$-\mathbf{v}\oplus\mathbf{w} = (-\mathbf{w})\oplus(-\mathbf{v}).$$

But this identity is not true (just substitute into the formula above to check)! This is a famous 'paradox', which seemingly shows the violation of the basic idea in special relativity that a boost by $\mathbf{v}$ is inverse to a boost by $-\mathbf{v}$ (which is actually a fundamental assumption in deriving the Lorentz transformation laws directly from the invariance of the speed of light in all inertial frames).

Of course, this is not actually a paradox. As already stated, the key assumption that 'two boosts give another boost' is flawed. So what saves us? We know that the frame S' must appear to be travelling with velocity $\mathbf{v}\oplus\mathbf{w}$ from the rest frame S, by the formula for relativistic velocity addition. The only other thing permitted to happen to the frame S' is some change in its *orientation* relative to the rest frame S. That is, there must exist an orthogonal matrix R such that:

$$B(\mathbf{w})B(\mathbf{v}) = B(\mathbf{v}\oplus\mathbf{w})R,$$

or an orthogonal matrix R' such that $B(\mathbf{w})B(\mathbf{v}) = R'B(\mathbf{v}\oplus\mathbf{w})$. It can be shown that the matrices R, R' are in fact *rotations*, called *Thomas-Wigner rotations*.[1] In this question, we used infinitesimal methods to verify a given form for a Thomas-Wigner rotation.

[1] More information is available on the Wikipedia page here: https://en.wikipedia.org/wiki/Wigner_rotation.

3. A field $\phi(x)$ transforms under the action of a Poincaré transformation (Λ, a) such that

$$U[\Lambda, a]\phi(x)U[\Lambda, a]^{-1} = \phi(\Lambda x + a).$$

For an infinitesimal transformation $\Lambda^\mu{}_\nu = \delta^\mu{}_\nu + \omega^\mu{}_\nu$, and correspondingly $U[\Lambda, a] = 1 - \frac{i}{2}\omega^{\mu\nu}M_{\mu\nu} - ia^\mu P_\mu$.

(a) Show that

$$[M_{\mu\nu}, \phi(x)] = -i(x_\mu \partial_\nu - x_\nu \partial_\mu)\phi(x), \qquad [P_\mu, \phi(x)] = i\partial_\mu \phi(x).$$

(b) Verify that $M_{\mu\nu} \mapsto i(x_\mu\partial_\nu - x_\nu\partial_\mu)$ and $P_\mu \mapsto -i\partial_\mu$ satisfy the algebra for $[M_{\mu\nu}, M_{\sigma\rho}]$ and $[M_{\mu\nu}, P_\sigma]$ expected for the Poincaré group.

➾ **Solution:** (a) Let us expand the transformation rule, supposing that $\Lambda = 1 + \omega$ with ω infinitesimal, and that a is infinitesimal. On the left hand side, we have:

$$U[\Lambda, a]\phi(x)U[\Lambda, a]^{-1} = \left(1 - \frac{i}{2}\omega^{\mu\nu}M_{\mu\nu} - ia^\mu P_\mu\right)\phi(x)\left(1 + \frac{i}{2}\omega^{\mu\nu}M_{\mu\nu} + ia^\mu P_\mu\right) \quad \text{(expansion of } U[\Lambda, a], U[\Lambda, a]^{-1}\text{)}$$

$$= \phi(x) - \frac{i}{2}\omega^{\mu\nu}[M_{\mu\nu}, \phi(x)] - ia^\mu[P_\mu, \phi(x)] + O(a^2, \omega^2, a\omega).$$

On the right hand side, we have:

$$\begin{aligned}
\phi(\Lambda x + a) &= \phi(x + \omega x + a) && \text{(assuming } \Lambda = 1 + \omega \text{ infinitesimal)}\\
&= \phi(x) + (\omega x + a)^\mu \partial_\mu \phi(x) + \cdots && \text{(Taylor expanding)}\\
&= \phi(x) + \omega^{\mu\nu}x_\nu \partial_\mu \phi(x) + a^\mu \partial_\mu \phi(x) + \cdots && \text{(inserting indices)}
\end{aligned}$$

Since these expressions must agree for all a, we have one of the required commutators (comparing the terms of order a):

$$-ia^\mu[P^\mu, \phi(x)] = a^\mu \partial_\mu \phi(x) \qquad \Rightarrow \qquad [P^\mu, \phi(x)] = i\partial_\mu \phi(x),$$

Comparing the terms of order ω, we have:

$$-\frac{i}{2}\omega^{\mu\nu}[M_{\mu\nu}, \phi(x)] = \omega^{\mu\nu}x_\nu\partial_\mu\phi(x) \qquad \Rightarrow \qquad -\frac{i}{2}[M_{\mu\nu} - M_{\nu\mu}, \phi(x)] = (x_\nu\partial_\mu - x_\mu\partial_\nu)\,\phi(x).$$

Note we can only deduce that the antisymmetric parts of both sides agree, since $\omega^{\mu\nu}$ is a generic *antisymmetric* tensor, so projects out the antisymmetric part of both sides. We can simplify the left hand side by recalling that the generators of the Lorentz algebra are antisymmetric, i.e. $M_{\mu\nu} = -M_{\nu\mu}$. It follows that we have the remaining required commutator:

$$[M_{\mu\nu}, \phi(x)] = -i\left(x_\mu\partial_\nu - x_\nu\partial_\mu\right)\phi(x).$$

(b) Recall that the Poincaré algebra is given by:

$$[M_{\mu\nu}, M_{\sigma\rho}] = i\left(\eta_{\nu\sigma}M_{\mu\rho} - \eta_{\mu\sigma}M_{\nu\rho} - \eta_{\nu\rho}M_{\mu\sigma} + \eta_{\mu\rho}M_{\nu\sigma}\right), \qquad [M_{\mu\nu}, P_\sigma] = i\left(\eta_{\nu\sigma}P_\mu - \eta_{\mu\sigma}P_\nu\right),$$

together with the identity $[P_\mu, P_\nu] = 0$, which holds trivially for this representation since partial derivatives commute. We just need to check that the first two commutation relations of the algebra are satisfied by the choices:

$$M_{\mu\nu} = i\left(x_\mu\partial_\nu - x_\nu\partial_\mu\right), \qquad P_\mu = -i\partial_\mu.$$

Before diving in, let's establish some preliminary results:

Theorem: Let $f(x)$ be an arbitrary function of spacetime. Then we have:

$$[x_\mu \partial_\nu, \partial_\sigma] f(x) = -\eta_{\mu\sigma}\partial_\nu f(x), \qquad \text{and} \qquad [x_\mu\partial_\nu, x_\sigma\partial_\rho] f(x) = \eta_{\nu\sigma} x_\mu \partial_\rho f(x) - \eta_{\rho\mu} x_\sigma \partial_\nu f(x).$$

Proof: For the first relation, we have:

$$[x_\mu \partial_\nu, \partial_\sigma] f(x) = x_\mu \partial_\nu \partial_\sigma f(x) - \partial_\sigma (x_\mu \partial_\nu f(x)) = -\eta_{\mu\sigma}\partial_\nu f(x),$$

and for the second relation, we have:

$$[x_\mu\partial_\nu, x_\sigma\partial_\rho] f(x) = x_\mu \partial_\nu \left(x_\sigma \partial_\rho f(x)\right) - x_\sigma \partial_\rho \left(x_\mu \partial_\nu f(x)\right) = \eta_{\nu\sigma} x_\mu \partial_\rho f(x) - \eta_{\rho\mu} x_\sigma \partial_\nu f(x). \quad \square$$

We now start the question proper; we begin by checking the second commutation relation. Acting on some arbitrary function of spacetime $f(x)$, we have:

$$\begin{aligned} [M_{\mu\nu}, P_\sigma] f(x) &= [x_\mu \partial_\nu - x_\nu \partial_\mu, \partial_\sigma] f(x) \\ &= -\eta_{\mu\sigma} \partial_\nu f(x) + \eta_{\nu\sigma} \partial_\mu f(x) \\ &= i \left(\eta_{\nu\sigma} P_\mu - \eta_{\mu\sigma} P_\nu\right) f(x), \end{aligned}$$

hence the given operators are indeed a representation of this commutation relation.

It remains to check the commutation relation for the Lorentz algebra. Again, acting on some arbitrary function of spacetime $f(x)$, we have:

$$\begin{aligned} [M_{\mu\nu}, M_{\sigma\rho}] f(x) &= -[x_\mu\partial_\nu - x_\nu \partial_\mu, x_\sigma \partial_\rho - x_\rho \partial_\sigma] f(x) \\ &= -(\eta_{\nu\sigma} x_\mu \partial_\rho - \eta_{\rho\mu} x_\sigma \partial_\nu) f(x) + (\eta_{\nu\rho} x_\mu \partial_\sigma - \eta_{\sigma\mu} x_\rho \partial_\nu) f(x) \\ &\quad + (\eta_{\mu\sigma} x_\nu \partial_\rho - \eta_{\rho\nu} x_\sigma \partial_\mu) f(x) - (\eta_{\mu\rho} x_\nu \partial_\sigma - \eta_{\sigma\nu} x_\rho \partial_\mu) f(x) \\ &= i \left(i\eta_{\nu\sigma} (x_\mu \partial_\rho - x_\rho \partial_\mu) - i\eta_{\mu\sigma}(x_\nu \partial_\rho - x_\rho \partial_\nu) - i\eta_{\nu\rho}(x_\mu \partial_\sigma - x_\sigma \partial_\rho) + i\eta_{\mu\rho}(x_\nu\partial_\sigma - x_\sigma \partial_\rho)\right) f(x) \\ &= i \left(\eta_{\nu\sigma} M_{\mu\rho} - \eta_{\mu\sigma} M_{\nu\rho} - \eta_{\nu\rho} M_{\mu\sigma} + \eta_{\mu\rho} M_{\nu\sigma}\right) f(x), \end{aligned}$$

and hence the commutation relations are again satisfied by this representation of the operators.

❈ **Comments:** Let's derive the transformation law for a scalar field used in this question directly from the theory of Poincaré representations which we studied in lectures. We work with the mass $m^2 > 0$, spin $s = 0$ irreducible representation of the Poincaré group, which takes the form:

$$\mathcal{H} = \{|p\rangle : p \in \mathbb{R}^4, p^2 = m^2\}.$$

The interpretation of this space is that it is the Hilbert space for a free, spin-0, massive particle. We recall from lectures that the elements of the unitary representation U of the Poincaré group act on this space as:

$$U[\Lambda, a] \, |p\rangle = e^{ia\cdot \Lambda p} \, |\Lambda p\rangle \,,$$

where we perform the Lorentz transformation Λ first, then the four-translation a.

In this question, we wish to construct the associated *quantum field* that gives rise to such particles, and then establish its transformation properties. By the basic axioms of quantum mechanics, we know that generic multi-particle states inhabit the associated *Fock space*:

$$\mathcal{F} = \mathbb{C} \oplus \mathcal{H} \oplus \mathcal{H}^2 \oplus \mathcal{H}^3 \oplus \cdots ,$$

where $\mathcal{H}^n$ is the n-fold tensor product.[2] We can define creation and annihilation operators on Fock space $\mathcal{F}$ via:

$$a_{\tilde{p}}^\dagger : \mathcal{H}^n \to \mathcal{H}^{n+1}, \qquad a_{\tilde{p}}^\dagger \left|p_1, ..., p_n\right\rangle = \left|p_1, ..., p_n, \tilde{p}\right\rangle ,$$

with the appropriate extension by linearity, and

$$a_{\tilde{p}} : \mathcal{H}^{n+1} \to \mathcal{H}^n, \qquad a_{\tilde{p}} \left|p_1, ..., p_n\right\rangle = \begin{cases} 0 & \text{if } p_i \neq \tilde{p} \text{ for any } i; \\ \left|p_1, ..., p_{i-1}, p_{i+1} ... p_n\right\rangle & \text{if } p_i = \tilde{p}, \end{cases}$$

again with the appropriate extension by linearity. These creation and annihilation operators satisfy nice transformation properties:

Proposition: The creation and annihilation operators transform under Poincaré transformations as:

$$U[\Lambda, a] a_p^\dagger U[\Lambda, a]^{-1} = e^{ia\cdot\Lambda p} a_{\Lambda p}^\dagger, \qquad U[\Lambda, a] a_p U[\Lambda, a]^{-1} = e^{-ia\cdot\Lambda p} a_{\Lambda p}.$$

Proof: Consider the action of $U[\Lambda, a] a_p^\dagger U[\Lambda, a]^{-1}$ on the general (rotated) basis element $U[\Lambda, a]\left|p_1 ... p_n\right\rangle$. We have:

$$\begin{aligned} U[\Lambda, a] a_p^\dagger U[\Lambda, a]^{-1} U[\Lambda, a] \left|p_1 ... p_n\right\rangle &= U[\Lambda, a] a_p^\dagger \left|p_1 ... p_n\right\rangle = U[\Lambda, a] \left|p_1 ... p_n, p\right\rangle \\ &= e^{ia\cdot\Lambda p_1} ... e^{ia\cdot\Lambda p_n} e^{ia\cdot\Lambda p} \left|\Lambda p_1 ... \Lambda p_n, \Lambda p\right\rangle = e^{ia\cdot\Lambda p} a_{\Lambda p}^\dagger U[\Lambda, a] \left|p_1 ... p_n\right\rangle , \end{aligned}$$

hence the equality follows. Similarly for the annihilation operator. □

We can now construct a *quantum field operator*[3] on Fock space as some linear combination of the creation and annihilation operators:

$$\phi(x) = \int \frac{d^3\mathbf{p}}{(2\pi)^3 2E_\mathbf{p}} \left(e^{-ip\cdot x} a_p + e^{ip\cdot x} a_p^\dagger\right) .$$

Under a Poincaré transformation $U[\Lambda, a]$, the quantum field transforms as:

$$\begin{aligned} \phi(x) \mapsto U[\Lambda, a]\phi(x) U[\Lambda, a]^{-1} &= \int \frac{d^3\mathbf{p}}{(2\pi)^3 2E_\mathbf{p}} \left(e^{-i(p\cdot x - a\cdot\Lambda p)} a_{\Lambda p} + e^{i(p\cdot x + \Lambda p\cdot a)} a_{\Lambda p}^\dagger\right) \\ &= \int \frac{d^3\tilde{\mathbf{p}}}{(2\pi)^3 2E_{\tilde{\mathbf{p}}}} \left(e^{-i\tilde{p}\cdot(\Lambda x + a)} a_{\tilde{p}} + e^{i\tilde{p}\cdot(\Lambda x + a)} a_{\tilde{p}}^\dagger\right) \qquad (\text{substituting } \tilde{p} = \Lambda p) \\ &= \phi(\Lambda x + a). \end{aligned}$$

Note also we used the invariance of the inner product on Minkowski spacetime, e.g. $\Lambda^{-1}p \cdot x = p \cdot \Lambda x$. Thus, we have established the scalar field transformation law directly from the results in lectures on transformations of states in the $m^2 > 0, s = 0$ irreducible representation of the Poincaré group.

[2] Technically, we require the symmetrised tensor product, to account for the fact that we have multiple identical bosonic particles. For a full discussion, including a proof of the spin-statistics theorem in this formalism, see Weinberg's *Quantum Theory of Fields Volume I*.

[3] Of course, you might reasonably ask *why* we are defining a quantum field operator here. The reason is due to *locality* - one can show that any operator on Fock space can be written in terms of creation and annihilation operators, but *local* ones must be written in terms of specific linear combinations of the creation and annihilation operators, namely *quantum fields*. Hence for our theory to be defined in terms of a local Hamiltonian operating on Fock space we are *forced* to work with quantum fields.

4. Show that there is a choice of basis for $\mathfrak{so}(4)$ consisting of 4×4 antisymmetric matrices that contain precisely two non-zero entries: 1 and -1. Evaluate the commutation relations of these generators. By choosing a new basis consisting of sums and differences of pairs of the old $\mathfrak{so}(3)$ generators, show that $\mathfrak{so}(4) \cong \mathfrak{so}(3) \oplus \mathfrak{so}(3)$.

➼ **Solution:** Recall from lectures that $\mathfrak{so}(4)$ is the set of 4×4 real, antisymmetric matrices. This Lie algebra is six-dimensional, since there are six independent off-diagonal entries which we can pick.

Let's try to pick a clever basis to minimise our work in the next part of the question. One obvious thing we can do is embed the Lie algebra $\mathfrak{so}(3)$ into the Lie algebra $\mathfrak{so}(4)$ by choosing the first three elements of our basis to be the standard basis of $\mathfrak{so}(3)$ (given on Sheet 2, Question 5 as $(t_i)_{jk} = -i(T_i)_{jk} = -\epsilon_{ijk}$ - note the need to convert to an anti-Hermitian matrix in this question, since we are working with real Lie algebras) embedded into the top left 3×3 submatrices of our 4×4 matrices:

$$t_1 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad t_2 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad t_3 = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

By construction, these matrices obey the standard commutation relations $[t_i, t_j] = \epsilon_{ijk} t_k$ (which one can check via an explicit calculation like the one in Sheet 2, Question 5).

We still need three more matrices to complete the basis. We choose:

$$s_1 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}, \qquad s_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}, \qquad s_3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}.$$

In order to evaluate the commutation relations between the s_i and the t_i, it is helpful to use the result of Question 7, that the matrices T_{ij} whose only non-zero entries are ones in the (i, j)th entry obey the commutation relations:

$$[T_{ij}, T_{kl}] = \delta_{kj} T_{il} - \delta_{il} T_{kj}.$$

Identifying $t_1 = T_{32} - T_{23}$, $t_2 = T_{13} - T_{31}$, $t_3 = T_{21} - T_{12}$, $s_1 = T_{14} - T_{41}$, $s_2 = T_{24} - T_{42}$ and $s_3 = T_{34} - T_{43}$, we see that in general we must compute the commutation relations of $e_{ij} = T_{ij} - T_{ji}$. We have:

$$\begin{aligned}
[e_{ij}, e_{kl}] &= [T_{ij}, T_{kl}] - [T_{ij}, T_{lk}] - [T_{ji}, T_{kl}] + [T_{ji}, T_{lk}] \\
&= \delta_{kj} T_{il} - \delta_{il} T_{kj} - \delta_{lj} T_{ik} + \delta_{ik} T_{lj} - \delta_{ki} T_{jl} + \delta_{jl} T_{ki} + \delta_{li} T_{jk} - \delta_{jk} T_{li} \\
&= \delta_{jk} e_{il} - \delta_{il} e_{kj} - \delta_{jl} e_{ik} + \delta_{ik} e_{lj}.
\end{aligned}$$

This allows us to read off the commutators. Since $s_i = e_{i4}$ and $t_i = -\frac{1}{2}\epsilon_{ijk} e_{jk}$ (for $i, j, k = 1, 2, 3$), we have:

$$\begin{aligned}
[s_i, s_j] &= [e_{i4}, e_{j4}] \\
&= \delta_{4j} e_{i4} - \delta_{i4} e_{j4} - e_{ij} + \delta_{ij} e_{44} \\
&= -e_{ij} && (e_{44} = 0, i, j \in \{1, 2, 3\}) \\
&= -\frac{1}{2}(e_{ij} - e_{ji}) && (e_{ij} = -e_{ji}) \\
&= -\frac{1}{2}\epsilon_{ijk}\epsilon_{klm} e_{lm} \\
&= \epsilon_{ijk} t_k.
\end{aligned}$$

We also have:

$$\begin{aligned}
[s_i, t_j] &= -\frac{1}{2}\epsilon_{jkl}[e_{i4}, e_{kl}] \\
&= -\frac{1}{2}\epsilon_{jkl}\left(\delta_{4k}e_{il} - \delta_{il}e_{k4} - \delta_{4l}e_{ik} + \delta_{ik}e_{l4}\right) \\
&= -\frac{1}{2}\epsilon_{j4l}e_{il} + \frac{1}{2}\epsilon_{jki}e_{k4} + \frac{1}{2}\epsilon_{jk4}e_{ik} - \frac{1}{2}\epsilon_{jil}e_{l4} \\
&= \frac{1}{2}\epsilon_{jki}e_{k4} - \frac{1}{2}\epsilon_{jik}e_{k4} \qquad\qquad (i,j,k \in \{1,2,3\}) \\
&= \epsilon_{ijk}e_{k4} \\
&= \epsilon_{ijk}s_k.
\end{aligned}$$

Hence, we have evaluated all commutators in this basis as required (without multiplying a single matrix!). We have:

$$[t_i, t_j] = \epsilon_{ijk}t_k, \qquad [s_i, s_j] = \epsilon_{ijk}t_k, \qquad [s_i, t_j] = \epsilon_{ijk}s_k.$$

Now define $u_i^{\pm} = \frac{1}{2}(t_i \pm s_i)$. With this definition, we have:

$$[u_i^{\pm}, u_j^{\pm}] = \frac{1}{4}\left([t_i, t_j] \pm [t_i, s_j] \pm [s_i, t_j] + [s_i, s_j]\right) = \frac{1}{4}\left(\epsilon_{ijk}t_k \pm \epsilon_{ijk}s_k \pm \epsilon_{ijk}s_k + \epsilon_{ijk}t_k\right) = \epsilon_{ijk}\cdot\frac{1}{2}\left(t_k \pm s_k\right) = \epsilon_{ijk}u_k^{\pm}.$$

Thus, u_i^+ and u_i^- separately obey the $\mathfrak{so}(3)$ commutation relations. Furthermore, we have:

$$[u_i^{\pm}, u_j^{\mp}] = \frac{1}{4}\left([t_i, t_j] \mp [t_i, s_j] \pm [s_i, t_j] - [s_i, s_j]\right) = \frac{1}{4}\left(\epsilon_{ijk}t_k \mp \epsilon_{ijk}s_k \pm \epsilon_{ijk}s_k - \epsilon_{ijk}t_k\right) = 0,$$

so these $\mathfrak{so}(3)$ subalgebras are completely commuting. We deduce $\mathfrak{so}(4) \cong \mathfrak{so}(3) \oplus \mathfrak{so}(3)$ as required. In particular, this shows that $\mathfrak{so}(4)$ is *semisimple* since it is the direct sum of simple Lie algebras.

5. Let $\{T^i{}_j\}$ be $n \times n$ matrices such that $T^i{}_j$ has 1 in the i'th row and j'th column and is zero otherwise.

(a) Show that they satisfy the Lie algebra $[T^i{}_j, T^k{}_l] = \delta^k{}_j T^i{}_l - \delta^i{}_l T^k{}_j$.

(b) Define $X = T^i{}_j X^j{}_i$ with arbitrary components $X^j{}_i$. Determine the adjoint matrix $(X^{\text{ad}})^n{}_{m,}{}^k{}_l$ by

$$[X, T^k{}_l] = T^m{}_n (X^{\text{ad}})^n{}_{m,}{}^k{}_l,$$

and show that

$$\kappa(X,Y) = \mathrm{tr}(X^{\text{ad}}Y^{\text{ad}}) = 2\left(n\sum_{i,j} X^j{}_i Y^i{}_j - \sum_i X^i{}_i \sum_j Y^j{}_j\right).$$

(c) Show that $1 + \epsilon X \in U(n)$ for infinitesimal ϵ if $(X^j{}_i)^* = -X^i{}_j$.

(d) Hence show that in this case

$$\kappa(X,X) = -2n\sum_{i,j} |\hat{X}^j{}_i|^2, \quad \hat{X}^j{}_i = X^j{}_i - \frac{1}{n}\delta^j{}_i \sum_k X^k{}_k,$$

and therefore $\kappa(X,X) \Leftrightarrow X^{\text{ad}} = 0$.

(e) What restrictions must be made for $SU(n)$? Verify that in this case the generators satisfy $\kappa(X,X) < 0$, and deduce the group is semi-simple.

Solution: (a) We have that $(T^i{}_j)_a{}^b = \delta^i{}_a \delta^b{}_j$. Thus

$$\left(T^i{}_j T^k{}_l\right)_a{}^b = \left(T^i{}_j\right)_a{}^c \left(T^k{}_l\right)_c{}^b = \delta^i{}_a \delta^c{}_j \delta^k{}_c \delta^b{}_l = \delta^k{}_j \delta^i{}_a \delta^b{}_l = \delta^k{}_j \left(T^i{}_l\right)_a{}^b.$$

Thus $T^i{}_j T^k{}_l = \delta^k_j T^i{}_l$. This means that we have

$$[T^i{}_j, T^k{}_l] = T^i{}_j T^k{}_l - T^k{}_l T^i{}_j = \delta^k{}_j T^i{}_l - \delta^i{}_l T^k{}_j.$$

(b) We now have that $X = T^i{}_j X^j{}_i$. This means

$$T^m{}_n (X^{\text{ad}})^n{}_{m,}{}^k{}_l = [X, T^k{}_l] = X^j{}_i [T^i{}_j, T^k{}_l] = X^j{}_i\left(\delta^k{}_j T^i{}_l - \delta^i{}_l T^k{}_j\right) = X^k{}_i T^i{}_l - X^j{}_l T^k{}_j.$$

Equating coefficients on both sides gives us

$$(X^{\text{ad}})^n{}_{m,}{}^k{}_l = \delta^n{}_l X^k{}_m - \delta^k{}_m X^n{}_l.$$

We now need to calculate the killing form on the generators X and Y

$$\begin{aligned}
\kappa(X,Y) &= \mathrm{Tr}(X^{\text{ad}}Y^{\text{ad}}) \\
&= (X^{\text{ad}})^n{}_{m,}{}^k{}_l (Y^{\text{ad}})^l{}_{k,}{}^m{}_n \\
&= \left(\delta^n{}_l X^k{}_m - \delta^k{}_m X^n{}_l\right)\left(\delta^l{}_n Y^m{}_k - \delta^m{}_k Y^l{}_n\right) \\
&= (\delta^n{}_l \delta^l{}_n X^k{}_m Y^m{}_k + \delta^k{}_m \delta^k{}_m X^n{}_l Y^l{}_n - \delta^k{}_m \delta^l{}_n X^n{}_l Y^m{}_k - \delta^n{}_l \delta^m{}_k X^k{}_m Y^l{}_n).
\end{aligned}$$

After contracting the indices (using that $\delta^n_l \delta^l_n = n$) and relabelling the dummy indices we get:

$$\kappa(X,Y) = 2(nX^j{}_i Y^i{}_j - X^i{}_i Y^j{}_j).$$

This is the desired result.

(c) For $1+\epsilon X \in U(N)$, we need $(1+\epsilon X)^\dagger(1+\epsilon X) = 1$. Expanding this to first order gives $1+\epsilon(X+X^\dagger) = I$ and thus $X^\dagger = -X$. Since $X = T^i{}_j X^j{}_i$, we have that $X^\dagger = T^j{}_i (X^j{}_i)^* = T^i{}_j (X^i{}_j)^*$. So in $X^\dagger = -X$ equating coefficients of $T^i{}_j$ on both sides gives $(X^i{}_j)^* = -X^j{}_i$, which after a relabeling of indices is the desired result.

(d) We now want to find the form $\kappa(X,X)$, in the case where $X^\dagger = -X$, that is find it in the case of $U(n)$. It is easiest to do this in the reverse way to what should normally be done. That is, to start with the solution they gave us, and show that this is equal to $\kappa(X,X)$. Then argue that the steps can be followed in reverse. Thus, we have that

$$\begin{aligned} RHS &= -2n\sum_{i,j} |\hat{X}^j{}_i|^2 \\ &= -2n\sum_{i,j}\left(X^j{}_i - \frac{1}{n}\delta^j{}_i X^k{}_k\right)\left((X^j{}_i)^* - \frac{1}{n}\delta^j{}_i (X^r{}_r)^*\right). \end{aligned}$$

Using the expression we found in part (c) we get

$$RHS = -2n\sum_{i,j}\left(X^j{}_i - \frac{1}{n}\delta^j{}_i X^k{}_k\right)\left(-X^i{}_j + \frac{1}{n}\delta^j{}_i X^r{}_r\right).$$

Expanding this we get

$$RHS = -2n\sum_{i,j}\left(-X^j{}_i X^i{}_j + \frac{1}{n}\delta^j{}_i X^k{}_k X^i{}_j + \frac{1}{n}X^j{}_i\delta^j{}_i X^r{}_r - \frac{1}{n^2}\delta^j{}_i\delta^j{}_i X^k{}_k X^r{}_r\right)$$

Contracting indices and using that $\sum_{i,j}\delta^j{}_i\delta^j{}_i = n$, we get

$$\begin{aligned} RHS &= -2n\sum_{i,j}\left(-X^j{}_i X^i{}_j + \frac{1}{n}X^k{}_k X^j{}_j + \frac{1}{n}X^j{}_j X^r{}_r - \frac{1}{n}X^k{}_k X^r{}_r\right) \\ &= 2(nX^j{}_i X^i{}_j - X^k{}_k X^r{}_r) \\ &= \kappa(X,X). \end{aligned}$$

This is as we were required to show.

If $\kappa(X,X) = 0$, then $\hat{X}^j{}_i = 0$ and $X^j{}_i = \frac{1}{n}\delta^j{}_i X^r{}_r$. We thus have that

$$\begin{aligned} (X^{\text{ad}})^n{}_{m,}{}^k{}_l &= \delta^n{}_l X^k{}_m - \delta^k{}_m X^n{}_l \\ &= \delta^n{}_l\delta^k{}_m \frac{1}{n}X^r{}_r - \delta^k{}_m\delta^n{}_l\frac{1}{n}X^r{}_r \\ &= 0. \end{aligned}$$

Note that this does not imply that $U(n)$ is semi-simple because $X^{\text{ad}} = 0$ for $X = I$ as well as $X = 0$.

Conversely, $X^{\text{ad}} = 0$ gives $\kappa(X,X) = \mathrm{Tr}(X^{\text{ad}}X^{\text{ad}}) = 0$ immediately.

(e) For $SU(N)$ we require that $\det(1+\epsilon X) = 1$. Taking the Taylor expansion of this we get $\det(1+\epsilon X) \approx 1+\epsilon\mathrm{Tr}(X) = 1$, and therefore that $\mathrm{Tr}(X) = 0$, *i.e.* $X^r{}_r = 0$. Thus in this case $\hat{X}^j{}_i = X^j{}_i$, and we have that

$$\kappa(X,X) = -2n\sum_{i,j}|X^j{}_i|^2.$$

Now we get, for $X \neq 0$, that $\kappa(X,X) < 0$. This proves that the group is semi-simple.

6. For a simple Lie algebra $\mathfrak{g}$, with generators X_a such that $[X_a, X_b] = f_{abc}X_c$, where f_{abc} is totally antisymmetric, let $\tilde{T}_a$ be matrices forming a basis for a representation R of $\mathfrak{g}$, and assume $\tilde{T}_a\tilde{T}_a = C_R I$. Define:

$$\langle X_a, X_b\rangle = \mathrm{Tr}(\tilde{T}_a\tilde{T}_b)\frac{\dim(\mathfrak{g})}{C_R\dim(R)}.$$

(a) Let $\mathfrak{g} = \mathfrak{su}(2)$. Evaluate $\langle J_3, J_3\rangle$ in the jth irreducible representation of $SU(2)$ and show that the result is independent of j.

(b) For $\mathfrak{su}(3)$ show that the Gell-Mann representation given by $\tilde{T}_a = \frac{i}{2}\lambda_a$ gives the same value for $\langle X_a, X_b\rangle$ as the adjoint representation $(T_a^{\mathrm{ad}})_{bc} = f_{abc}$. [The Gell-Mann matrices are:

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$

$$\lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \quad \lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad \lambda_8 = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}.]$$

(c) It can be shown that the Killing form on a simple Lie algebra is the unique ad-invariant symmetric bilinear form on a simple Lie algebra, up to an overall scalar multiple. How do you interpret your calculations above in relation to this fact?

Solution: (a) From lectures, we know that the spin-j matrix representation of J_3 is simply the $(2j+1)\times(2j+1)$ diagonal matrix:

$$J_3 = \mathrm{diag}\{-j, -j+1, ..., j-1, j\},$$

and the quadratic Casimir is $\mathbf{J}^2 = J_1^2 + J_2^2 + J_3^2 = j(j+1)I_{2j+1}$.

The dimension of the Lie algebra is $\dim(\mathcal{L}) = 2^2 - 1 = 3$, and the dimension of the representation is $\dim(R) = 2j+1$. The expression for the quadratic Casimir immediately implies $C_R = j(j+1)$. Hence we only need to evaluate the trace $\mathrm{Tr}(J_3^2)$, which is given by:

$$\mathrm{Tr}(J_3^2) = \mathrm{Tr}(\mathrm{diag}\{(-j)^2, (-j+1)^2, ..., (j-1)^2, j^2\}) = \sum_{i=0}^{2j}(i-j)^2 = \sum_{i=0}^{2j} i^2 - 2j\sum_{i=0}^{2j} i + 2j^3$$
$$= \frac{j(2j+1)(4j+1)}{3} - 2j^2(2j+1) + j^2(2j+1) = \frac{j(2j^2+3j+1)}{3} = \frac{j(j+1)(2j+1)}{3}.$$

using a standard formula for the sum of squares. Putting everything together, we have:

$$\langle J_3, J_3\rangle = \frac{3j(j+1)(2j+1)}{3j(j+1)(2j+1)} = 1,$$

which is indeed independent of j.

(b) For this question, all we need know about the Gell-Mann matrices is that they are constructed such that they make the Killing form 'orthonormal' in the following sense:

$$\mathrm{Tr}(\lambda_a \lambda_b) = 2\delta_{ab},$$

which you can verify explicitly using the forms of the Gell-Mann matrices given above (but please don't do this unless you are particularly masochistic). We can now evaluate $\langle X_a, X_b \rangle$. Since the matrices are 3×3, the dimension of the representation is $\dim(R) = 3$. The dimension of the Lie algebra is $\dim(\mathcal{L}) = \dim(SU(3)) = 3^2 - 1 = 8$. The trace $\mathrm{Tr}(\tilde{T}_a \tilde{T}_b)$ is given by:

$$\mathrm{Tr}(\tilde{T}_a \tilde{T}_b) = -\frac{1}{4}\mathrm{Tr}(\lambda_a \lambda_b) = -\frac{1}{2}\delta_{ab},$$

and the quadratic Casimir can be determined from the trace using:

$$3C_R = \mathrm{Tr}(C_R I) = \mathrm{Tr}(\tilde{T}_a \tilde{T}_a) = -\frac{1}{2}\delta_{aa} = -\frac{8}{2} = -4,$$

since there are 8 indices a on the Gell-Mann matrices λ_a, and I is the 3×3 identity matrix. It follows that $C_R = -4/3$. Hence we have $\langle X_a, X_b \rangle = \delta_{ab}$.

Now consider the adjoint representation T_a^{ad} instead, given by $(T_a^{\mathrm{ad}})_{bc} = f_{abc}$, where f_{abc} are the totally-antisymmetric structure constants of $\mathfrak{su}(3)$. The matrices T_a^{ad} are now 8×8 matrices instead, since they act on the Lie algebra itself, hence $\dim(R) = 8$. The Lie algebra is still of dimension $\dim(\mathcal{L}) = \dim(SU(3)) = 3^2 - 1 = 8$. The required trace is given by:

$$\mathrm{Tr}(T_a^{\mathrm{ad}} T_b^{\mathrm{ad}}) = f_{aij} f_{bji}.$$

However, this can be rewritten in the form as the quadratic Casimir using total antisymmetry:

$$f_{aij} f_{bji} = f_{iaj} f_{ijb} = (T_i^{\mathrm{ad}})_{aj} (T_i^{\mathrm{ad}})_{jb} = C_R \delta_{ab},$$

Hence we deduce $\mathrm{Tr}(T_a^{\mathrm{ad}} T_b^{\mathrm{ad}}) = C_R \delta_{ab}$. In particular, we don't even need to evaluate C_R, since it will cancel out in the final answer; so we simply have $\langle X_a, X_b \rangle = \delta_{ab}$, which is the same as the result we found for the fundamental representation.

(c) To explain these results, we note that $\langle X_a, X_b \rangle$ is certainly symmetric, and bilinear by definition since it is defined on generators and extended by linearity. It is also ad-invariant, since for any third generator X_c, we have (using the fact that representations of Lie algebras respect the Lie bracket):

$$\langle [X_c, X_a], X_b \rangle = \mathrm{Tr}([\tilde{T}_c, \tilde{T}_a]\tilde{T}_b)\frac{\dim(\mathfrak{g})}{C_R \dim(R)} = -\mathrm{Tr}(\tilde{T}_a[\tilde{T}_c, \tilde{T}_b])\frac{\dim(\mathfrak{g})}{C_R \dim(R)} = -\langle X_a, [X_c, X_b] \rangle.$$

By the result given in the question then, we have that:

$$\langle X_a, X_b \rangle = \lambda_R \kappa(X_a, X_b),$$

where κ is the Killing form and λ_R is a constant of proportionality, dependent on the representation. To determine the constant of proportionality, we sum over a, b; on the left hand side we have:

$$\langle X_a, X_a \rangle = \mathrm{Tr}(\tilde{T}_a \tilde{T}_a)\frac{\dim(\mathfrak{g})}{C_R \dim(R)} = \mathrm{Tr}(C_R I)\frac{\dim(\mathfrak{g})}{C_R \dim(R)} = \dim(\mathfrak{g}).$$

On the right hand side we have:

$$\lambda_R \kappa(X_a, X_a) = \lambda_R \mathrm{Tr}(T_a^{\mathrm{ad}} T_a^{\mathrm{ad}}) = \lambda_R \mathrm{Tr}(C_{\mathrm{ad}} I) = \lambda_R \dim(\mathfrak{g}) C_{\mathrm{ad}}.$$

Hence we see that $\lambda_R = 1/C_{\mathrm{ad}}$, and we have:

$$\langle X_a, X_b \rangle = \frac{1}{C_{\mathrm{ad}}}\kappa(X_a, X_b).$$

Thus all the results in the different representations in this question *had* to be the same - we were just evaluating the Killing form in lots of different ways!

7. The Lie algebra of $U(n)$ may be represented by a basis consisting of the n^2-n off diagonal matrices $(E_{ij})_{kl}=\delta_{ik}\delta_{jl}$ for $i\neq j$ and also the n diagonal matrices $(h_i)_{kl}=\delta_{ik}\delta_{kl}$, no sum on k, where $i,j,k,l=1,\cdots,n$. For $SU(n)$ it is necessary to restrict to traceless matrices given by h_i-h_j for some i,j. The $n-1$ independent h_i-h_j correspond to the Cartan subalgebra.

(a) Show that
$$[h_i,E_{jk}]=(\delta_{ij}-\delta_{ik})E_{jk},\quad [E_{ij},E_{ji}]=h_i-h_j\quad \text{(no summation convention)}.$$

(b) Let $\mathbf{e}_i$ be orthogonal n-dimensional unit vectors, $(\mathbf{e}_i)_j=\delta_{ij}$. Show that E_{ij} is associated with the root vector $\mathbf{e}_i-\mathbf{e}_j$ while E_{ji} corresponds to the root vector $\mathbf{e}_j-\mathbf{e}_i$.

(c) Hence show that there are $n(n-1)$ root vectors belonging to the $n-1$ dimensional hyperplane orthogonal to $\sum_i \mathbf{e}_i$.

(d) Verify that we may take as simple roots
$$\alpha_1=\mathbf{e}_1-\mathbf{e}_2,\quad \alpha_2=\mathbf{e}_2-\mathbf{e}_3,\quad \ldots,\quad \alpha_i=\mathbf{e}_i-\mathbf{e}_{i+1},\quad \ldots,\quad \alpha_{n-1}=\mathbf{e}_{n-1}-\mathbf{e}_n,$$
by showing that all roots may be expressed in terms of α_i with either positive or negative integer coefficients.

(e) Determine the Cartan matrix and write down the corresponding Dynkin diagram.

Solution: (a) Without summing over indices, we have
$$\begin{aligned}([h_i,E_{jk}])_{mn}&=\sum_p((h_i)_{mp}(E_{jk})_{pn}-(E_{jk})_{mp}(h_i)_{pn})\\&=\sum_p(\delta_{im}\delta_{mp}\delta_{jp}\delta kn-\delta_{jm}\delta_{kp}\delta_{ip}\delta_{in})\\&=(\delta_{im}\delta_{mj}\delta_{kn}-\delta_{jm}\delta_{ik}\delta_{in}).\end{aligned}$$
We now use that $\delta_{im}\delta_{mj}=\delta_{ij}\delta_{jm}$ and $\delta_{ik}\delta_{in}=\delta_{ik}\delta_{kn}$ to get
$$\begin{aligned}([h_i,E_{jk}])_{mn}&=(\delta_{ij}-\delta_{ik})\delta_{jm}\delta kn\\&=(\delta_{ij}-\delta_{ik})(E_{jk})_{mn}.\end{aligned}$$

This then proves our first formula.

For the second,
$$\begin{aligned}([E_{ij},E_{ji}])_{mn}&=\sum_p((E_{ij})_{mp}(E_{ji})_{pn}-(E_{ji})_{mp}(E_{ij})_{pn})\\&=\sum_p(\delta_{im}\delta_{jp}\delta_{jp}\delta_{in}-\delta_{jm}\delta_{ip}\delta_{ip}\delta_{jn})\\&=\delta_{im}\delta_{jj}\delta_{in}-\delta_{jm}\delta_{ii}\delta_{jn}\\&=\delta_{im}\delta_{in}-\delta_{jm}\delta_{jn}\\&=\delta_{im}\delta_{mn}-\delta_{jm}\delta_{mn}\\&=(h_i-h_j)_{mn},\end{aligned}$$

which proves the second expression.

(b) The root associated with E_{ij} is the map $\alpha_{ij} \in H^*$ (the dual of the Cartan subalgebra), such that $[h, E_{ij}] = \alpha_{ij}(h)E_{ij}$, for all $h \in H$. The basis $\mathbf{e}_i$ is the dual basis to h_i, meaning that $\mathbf{e}_i(h_k) = \delta_{ik}$. Since $[h_k, E_{ij}] = (\delta_{ik} - \delta_{jk})E_{ij} = (\mathbf{e}_i(h_k) - \mathbf{e}_j(h_k))E_{ij}$ we have that $\alpha_{ij} = \mathbf{e}_i - \mathbf{e}_j$; likewise $\alpha_{ji} = \mathbf{e}_j - \mathbf{e}_i$. Whilst we technically used elements of the Cartan subalgebra of $\mathfrak{u}(n)$ rather than $\mathfrak{su}(n)$, we note that for the basis:

$$h_k - \frac{1}{n}I,$$

for the Cartan subalgebra of $\mathfrak{su}(n)$, where I is the $n \times n$ identity matrix, we have $[h_k - \frac{1}{n}I, E_{ij}] = [h_k, E_{ij}]$. Thus α_{ij} are the roots for $\mathfrak{su}(n)$, as required.

(c) For each E_{ij} we have a root vector, there are $n(n-1)$ of these, so we have $n(n-1)$ root vectors. Note that $(\mathbf{e}_i - \mathbf{e}_j) \cdot (\sum_i \mathbf{e}_i) = 1 - 1 = 0$, so they belong to the $n-1$ dimensional hyperplane orthogonal to $(\sum_i \mathbf{e}_i)$.

(d) Consider α_{ij} for $i < j$ we have that $\alpha_{ij} = \alpha_i + \alpha_{i+1} + \cdots + \alpha_{j-1}$ (where here $\alpha_i = \alpha_{i,i+1}$), these are the 'positive roots'. Since $\alpha_{ji} = -\alpha_{ij} = -(\alpha_i + \alpha_{i+1} + \cdots + \alpha_{j-1})$, these are the negative roots.

(e) The Cartan subalgebra is given by

$$A_{ij} = \frac{2(\alpha_i, \alpha_j)}{(\alpha_j, \alpha_j)}.$$

For us $(\alpha_i, \alpha_j) \propto \alpha_i \cdot \alpha_j$, where $\cdot$ is the dot product with respect to the basis $\mathbf{e}_i$. We have that

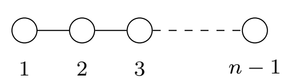
$$(\alpha_i, \alpha_j) = \begin{cases} -1, & \text{if } j = i-1 \\ 2, & \text{if } i = j, \\ -1 & \text{if } j = i+1 \\ 0, & \text{otherwise} \end{cases}.$$

Thus the Cartan matrix is given by

$$A = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 & 0 \\ 0 & -1 & 2 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 2 & -1 \\ 0 & 0 & 0 & \cdots & 0 & -1 & 2 \end{pmatrix}.$$

This is an $(n-1) \times (n-1)$ matrix.

A Dynkin diagram in this case has $n-1$ nodes (equal to the rank of the algebra), with the ith and the jth node linked by $-A_{ij}$ lines for $i < j$. In some cases, one needs to take account of the fact that some simple roots may be of different lengths, but this is not the case here. Our Dynkin diagram is then given by



8. A Lie algebra has simple roots $\boldsymbol{\alpha}_1, ..., \boldsymbol{\alpha}_r$. The fundamental weights $\{\boldsymbol{\mu}_i\}$ satisfy:

$$\frac{2\boldsymbol{\mu}_i \cdot \boldsymbol{\alpha}_j}{\boldsymbol{\alpha}_j \cdot \boldsymbol{\alpha}_j} = \delta_{ij}.$$

(a) Show that $\boldsymbol{\alpha}_i = \sum_j K_{ij}\boldsymbol{\mu}_j$ where K_{ij} is the Cartan matrix.

(b) A rank two Lie algebra has simple roots $\boldsymbol{\alpha}_1 = (1,0)$ and $\boldsymbol{\alpha}_2 = (-1,1)$. What is the Cartan matrix?

(c) Assuming any other positive roots are equal in length to either one of the simple roots, show that $\boldsymbol{\alpha}_3 = \boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2$ and $\boldsymbol{\alpha}_4 = 2\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2$ are the other positive roots.

(d) Draw the root diagram, and show that the dimension of the Lie algebra is ten.

(e) Construct the fundamental weights $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$.

(f) How is the highest weight of the representation whose weights coincide with the roots of the Lie algebra related to the fundamental weights?

Solution: (a) Since the simple roots of the Lie algebra span, we can write:

$$\boldsymbol{\mu}_i = \sum_j B_{ij}\boldsymbol{\alpha}_j$$

for any fundamental weight $\boldsymbol{\mu}_i$. Inserting this into the 'quantisation condition' given in the question, we have:

$$\delta_{ij} = \sum_k B_{ik}\frac{2\boldsymbol{\alpha}_k \cdot \boldsymbol{\alpha}_j}{\boldsymbol{\alpha}_j \cdot \boldsymbol{\alpha}_j} = \sum_k B_{ik}K_{kj},$$

and hence we see that B is the inverse of the Cartan matrix K. Thus inverting the equation expressing $\boldsymbol{\mu}_i$ in terms of $\boldsymbol{\alpha}_j$, we have:

$$\boldsymbol{\alpha}_i = \sum_j K_{ij}\boldsymbol{\mu}_j,$$

as required.

(b) Recall that the Cartan matrix elements are defined by:

$$K_{ij} = \frac{2\boldsymbol{\alpha}_i \cdot \boldsymbol{\alpha}_j}{\boldsymbol{\alpha}_j \cdot \boldsymbol{\alpha}_j}.$$

Hence inserting the given simple roots into this expression, we see that the Cartan matrix is:

$$K = \begin{pmatrix} 2 & -1 \\ -2 & 2 \end{pmatrix}.$$

(c) Let Δ be the set of simple roots. All roots of a Lie algebra can be generated from the simple roots Δ using the following algorithm:

Generating all roots from simple roots:

(1) Let $S = \Delta$.

(2) For each pair of *distinct* roots $\boldsymbol{\alpha} \in \Delta, \boldsymbol{\beta} \in S$, determine the maximum and minimum integers n_+, n_- such that $\boldsymbol{\beta} + n_-\boldsymbol{\alpha}$ and $\boldsymbol{\beta} + n_+\boldsymbol{\alpha}$ are roots. For all integers n such that $n_- \leq n \leq n_+$, we have that $\boldsymbol{\beta} + n\boldsymbol{\alpha}$ is a root; add all of these roots to the set S. A useful result that we can apply here is that the integers n_-, n_+ obey the quantisation condition:

$$n_- + n_+ = -\frac{2\boldsymbol{\alpha}\cdot\boldsymbol{\beta}}{\boldsymbol{\beta}^2}.$$

It is also useful to remember that: (i) if $\boldsymbol{\gamma}$ is a root, then a constant multiple $\lambda\boldsymbol{\gamma}$ is a root if and only if $\lambda = \pm 1$; (ii) if $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are simple roots, then $\boldsymbol{\beta} - \boldsymbol{\alpha}$ is not a root.

(3) Repeat (2) until we generate no new elements. At this point S contains all positive roots; the full root set is given by $S \cup (-S)$.

We apply the algorithm starting with $\Delta = \{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2\}$; before starting it's useful to note that $\boldsymbol{\alpha}_1^2 = 1, \boldsymbol{\alpha}_2^2 = 2$ and $\boldsymbol{\alpha}_1 \cdot \boldsymbol{\alpha}_2 = -1$. Setting $S = \{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2\}$, we have:

- Examine roots of the form $\boldsymbol{\alpha}_2 + n\boldsymbol{\alpha}_1$. We have that $\boldsymbol{\alpha}_2$ is a root, and $\boldsymbol{\alpha}_2 - \boldsymbol{\alpha}_1$ is not a root, since $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2$ are simple roots. Hence $n_- = 0$. In particular, it follows that:

$$n_+ = -\frac{2\boldsymbol{\alpha}_1\cdot\boldsymbol{\alpha}_2}{\boldsymbol{\alpha}_1^2} = 2.$$

  Hence we have that $\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2, 2\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2$ are also roots which must be added to the set S.

- The set S is now given by $S = \{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2, 2\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2\}$. Considering roots of the form $s + n\boldsymbol{\alpha}_1$ for $s \in S$, i.e. roots of the form $\boldsymbol{\alpha}_2 + n\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2 + n\boldsymbol{\alpha}_1$ and $2\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2 + n\boldsymbol{\alpha}_1$, will not give us anything new, so we should instead start considering roots of $s + n\boldsymbol{\alpha}_2$.

  First, consider roots of type $\boldsymbol{\alpha}_1 + n\boldsymbol{\alpha}_2$. Here, $\boldsymbol{\alpha}_1$ is a root, but $\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2$ is not a root, hence $n_- = 0$. It follows that:

$$n_+ = -\frac{2\boldsymbol{\alpha}_1\cdot\boldsymbol{\alpha}_2}{\boldsymbol{\alpha}_2^2} = 1.$$

  Hence we generate the roots $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2$ which are already in S.

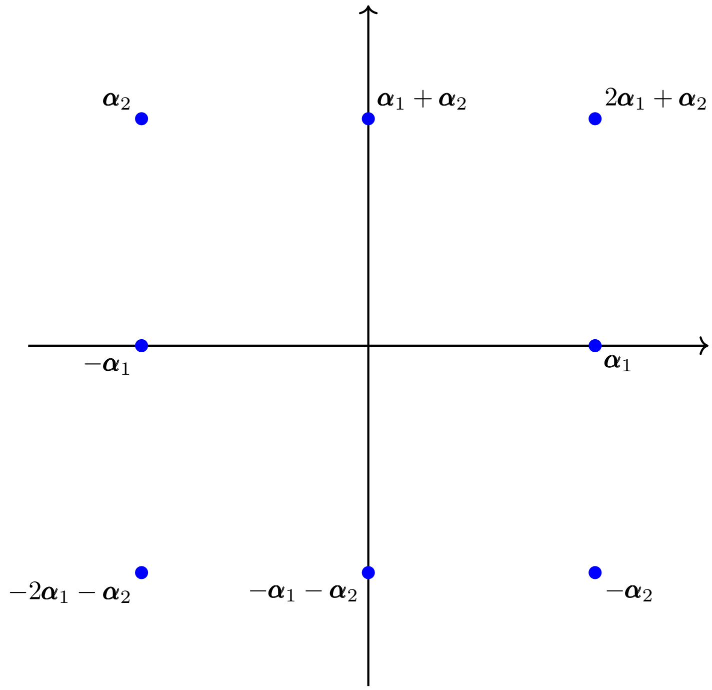
  Considering roots of type $\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2 + n\boldsymbol{\alpha}_2 = \boldsymbol{\alpha}_1 + (n+1)\boldsymbol{\alpha}_2$ will not give us anything new, and hence we are left considering roots of type $2\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2 + n\boldsymbol{\alpha}_2 = 2\boldsymbol{\alpha}_1 + (n+1)\boldsymbol{\alpha}_2$. When $n = 0$ we have a root, but when $n = -1$ we have that $2\boldsymbol{\alpha}_1$ is not a root because it is twice a root of a Lie algebra. Hence we have $n_- = 0$. Furthermore, when $n = 1$ we have that $2\boldsymbol{\alpha}_1 + 2\boldsymbol{\alpha}_2 = 2(\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2)$ is not a root because it is twice a root of a Lie algebra. Hence we have $n_+ = 0$, and it follows that the only root generated this way is $2\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2$.

At this point, we have generated no new roots, and so the algorithm terminates. It follows that the complete root set is:

$$S = \{\pm\boldsymbol{\alpha}_1, \pm\boldsymbol{\alpha}_2, \pm(\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2), \pm(2\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2)\} = \{\pm(1,0), \pm(-1,1), \pm(0,1), \pm(1,1)\},$$

as required.

(d) The root diagram is a square:



The dimension of the Lie algebra is given by the size of the root system, 8, plus the dimension of the Cartan subalgebra, 2 (which is equal to the number of simple roots, which form a basis for the dual of the Cartan subalgebra), hence is 10.

(e) The fundamental weights can be computed by inverting the Cartan matrix. We simply solve the linear system of equations:

$$\begin{aligned}\boldsymbol{\alpha}_1 &= 2\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 \\ \boldsymbol{\alpha}_2 &= -2\boldsymbol{\mu}_1 + 2\boldsymbol{\mu}_2.\end{aligned}$$

Adding the equations, we have

$$\boldsymbol{\mu}_2 = \boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2 = (0, 1).$$

Substituting this expression into the first equation, we have:

$$\boldsymbol{\mu}_1 = \frac{\boldsymbol{\alpha}_1 + \boldsymbol{\mu}_2}{2} = \boldsymbol{\alpha}_1 + \frac{1}{2}\boldsymbol{\alpha}_2 = (1/2, 1/2).$$

(f) The roots of the Lie algebra are the weights of the adjoint representation. Hence the highest weight vector of the adjoint representation is the same as the highest root. Recall, a highest root is a positive root whose associated step operator is annihilated by the adjoint action of the step operators of all other positive roots. Since:

$$[E_{\boldsymbol{\alpha}}, E_{\boldsymbol{\beta}}] = 0$$

if $\boldsymbol{\alpha} + \boldsymbol{\beta}$ is not a root and $\boldsymbol{\alpha} + \boldsymbol{\beta} \neq 0$, it is easy to see from the above diagram of the root system that $(1, 1) = 2\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2$ is the highest root. Hence the highest weight of the adjoint representation is $(1, 1) = 2\boldsymbol{\mu}_1$.

9. Consider the Lie algebra with exactly 2 simple roots $\boldsymbol{\alpha}_1 = (1, 0)$ and $\boldsymbol{\alpha}_2 = \frac{1}{2}(-3, \sqrt{3})$.

(a) Determine the fundamental weights $\mathbf{w}_1$ and $\mathbf{w}_2$.

Let $|q_1, q_2\rangle$ be a state corresponding to the weight $q_1\mathbf{w}_1 + q_2\mathbf{w}_2$.

(b) Assuming $E_{i\pm}$, H_i are the $SU(2)$ generators associated with the roots $\boldsymbol{\alpha}_i$, construct a basis for the representation space starting from a highest weight vector (i) $|1, 0\rangle$ and (ii) $|0, 1\rangle$, by the successive action of E_{1-} and E_{2-} on the highest weight state.

(c) Show that the dimensions of the spaces are respectively 7 and 14 (in the second case there are two independent states with $q_1 = q_2 = 0$).

(d) Construct the weight diagram and in the 14 dimensional case show that it coincides with the root diagram.

➾ **Solution:** (a) We begin by constructing the Cartan matrix. Using the formula:

$$K_{ij} = \frac{2\boldsymbol{\alpha}_i \cdot \boldsymbol{\alpha}_j}{\boldsymbol{\alpha}_j \cdot \boldsymbol{\alpha}_j},$$

we have that the Cartan matrix is given by:

$$K = \begin{pmatrix} 2 & -1 \\ -3 & 2 \end{pmatrix}.$$

Using Question 5(a), we see that the fundamental weights $\mathbf{w}_1, \mathbf{w}_2$ are given by solving the equations:

$$\boldsymbol{\alpha}_1 = 2\mathbf{w}_1 - \mathbf{w}_2,$$

$$\boldsymbol{\alpha}_2 = -3\mathbf{w}_1 + 2\mathbf{w}_2.$$

The solution is given by:

$$\mathbf{w}_1 = 2\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2 = \tfrac{1}{2}(1, \sqrt{3}), \qquad \mathbf{w}_2 = 3\boldsymbol{\alpha}_1 + 2\boldsymbol{\alpha}_2 = (0, \sqrt{3}).$$

(b) We use the following important results from lectures (which we do not prove):

The theorem of the highest weight: Let $\mathbf{w}_1, ..., \mathbf{w}_r$ be the fundamental weights of a simple complex Lie algebra $\mathfrak{g}$. Then:

(i) The irreducible representations of the Lie algebra are in bijection with the space $\mathbb{N}^r$ (here, $\mathbb{N}$ is the set $\{0, 1, 2, ...\}$). The labels $\boldsymbol{\Lambda} \in \mathbb{N}^r$ of the irreducible representation $d_{\boldsymbol{\Lambda}}$ are called the *Dynkin indices* of the representation.

(ii) The representation space of the irreducible representation $d_{\boldsymbol{\Lambda}}$ contains a vector $|\boldsymbol{\Lambda}\rangle$ (the *highest weight state*), unique up to multiplication by a complex scalar, such that:

$$d(E_{\boldsymbol{\alpha}})\,|\boldsymbol{\Lambda}\rangle = \mathbf{0} \text{ for all } \boldsymbol{\alpha} \in \Phi^+, \qquad \text{and} \qquad d(H_i)\,|\boldsymbol{\Lambda}\rangle = \Lambda_i\,|\boldsymbol{\Lambda}\rangle\,,$$

where Φ^+ is the set of positive roots of $\mathfrak{g}$, and H_i are the generators of the Cartan subalgebra of $\mathfrak{g}$.

(iii) The representation space of the irreducible representation $d_{\mathbf{\Lambda}}$ is spanned by vectors of the form:

$$d(E_{-\boldsymbol{\alpha}_{i_1}})...d(E_{-\boldsymbol{\alpha}_{i_n}})\,|\mathbf{\Lambda}\rangle\,,$$

where $\boldsymbol{\alpha}_{i_1},...,\boldsymbol{\alpha}_{i_n} \in \Phi_s$ are some simple roots $\mathfrak{g}$. In particular, every weight of the representation can be written in the form:

$$\boldsymbol{\mu} = \mathbf{\Lambda} - \sum_{i=1}^{r} \mu_i \boldsymbol{\alpha}_i,$$

where $\boldsymbol{\alpha}_1,...,\boldsymbol{\alpha}_r$ are the simple roots of $\mathfrak{g}$, and $\mu_i \in \mathbb{N}$.

Lemma: Suppose that $\mathfrak{g}$ is a simple complex Lie algebra with simple roots $\boldsymbol{\alpha}_1,...,\boldsymbol{\alpha}_r$ and fundamental weights $\mathbf{w}_1,...,\mathbf{w}_r$. Let d be a representation of $\mathfrak{g}$, and let:

$$\boldsymbol{\lambda} = \sum_{i=1}^{r} \lambda_i \mathbf{w}_i$$

be a weight of the representation, with $\lambda_i \in \mathbb{N}$. Then:

$$\boldsymbol{\lambda} - \sum_{i=1}^{r} m_i \boldsymbol{\alpha}_i$$

is a weight of the representation if $0 \leq m_i \leq \lambda_i$ and $m_i \in \mathbb{N}$.

Together, they imply the following algorithm for generating all the weights of an irreducible representation of a simple complex Lie algebra $\mathfrak{g}$ from the highest weight $\mathbf{\Lambda}$:

Generating all weights of an irreducible representation: Let $d_{\mathbf{\Lambda}}$ be the irreducible representation of the simple complex Lie algebra $\mathfrak{g}$ with Dynkin indices $\mathbf{\Lambda}$. Let $|\mathbf{\Lambda}\rangle$ be a highest weight state of the representation, with weight $\mathbf{\Lambda}$. To generate all weights, we use the following algorithm:

(1) Set $S = \{\mathbf{\Lambda}\}$.

(2) For each element $\boldsymbol{\mu} \in S$, write:

$$\boldsymbol{\mu} = \sum_{i=1}^{r} \mu_i \mathbf{w}_i,$$

where $\mathbf{w}_1,...,\mathbf{w}_r$ are the fundamental weights of the Lie algebra $\mathfrak{g}$. Add:

$$\boldsymbol{\mu} - \sum_{i=1}^{r} m_i \boldsymbol{\alpha}_i$$

to S for all $m_i \in \mathbb{N}$ such that $0 \leq m_i \leq \lambda_i$, where $\boldsymbol{\alpha}_1,...,\boldsymbol{\alpha}_r$ are the simple roots of the Lie algebra.

This procedure corresponds to applying the lowering operators $E_{-\boldsymbol{\alpha}_1}$, ..., $E_{-\boldsymbol{\alpha}_r}$ in specific sequences (we conflate $d(E_{-\boldsymbol{\alpha}})$ with $E_{-\boldsymbol{\alpha}}$ for notational convenience), so we can produce associated states:

$$|\Lambda - \boldsymbol{\alpha}_{i_1} - ... - \boldsymbol{\alpha}_{i_n}; m\rangle := E_{-\boldsymbol{\alpha}_{i_1}}...E_{-\boldsymbol{\alpha}_{i_n}}\,|\mathbf{\Lambda}\rangle$$

where m is some additional label accounting for the fact that different sequences of lowering operator applications can result in different states with the same weights (i.e. there can be degeneracy). These states span the representation space.

Checking whether two sequences of lowering operator applications actually result in equivalent states is rather difficult, and should be considered beyond the scope of the course. This is important in part (b)(ii) and part (c).

(i) Applying the algorithm to the highest weight state $|1,0\rangle$, which has eigenvalue $\mathbf{w}_1$ of $\mathbf{H} = (H_1, H_2)$, we have the following:

- We start from the state $|1,0\rangle$. We can apply E_{1-} precisely once and we cannot apply E_{2-}. We generate:

$$E_{1-}|1,0\rangle = |1-2,0+1\rangle = |-1,1\rangle .$$

- We now start from the state $|-1,1\rangle$. We cannot apply E_{1-} and we can apply E_{2-} precisely once. We generate:

$$E_{2-}|-1,1\rangle = |-1+3,1-2\rangle = |2,-1\rangle .$$

- We now start from the state $|2,-1\rangle$. We can apply E_{1-} precisely two times and we cannot apply E_{2-}. We generate:

$$E_{1-}|2,-1\rangle = |2-2,-1+1\rangle = |0,0\rangle$$
$$E_{1-}|0,0\rangle = |0-2,0+1\rangle = |-2,1\rangle .$$

- We now start from the states $|0,0\rangle$ and $|-2,1\rangle$. We cannot apply the lowering operators to the first state. For the second state, we cannot apply E_{1-} and we can apply E_{2-} precisely once. We generate:

$$E_{2-}|-2,1\rangle = |-2+3,1-2\rangle = |1,-1\rangle .$$

- Finally, we start from the state $|1,-1\rangle$. We can apply E_{1-} precisely once and we cannot apply E_{2-}. We generate:

$$E_{1-}|1,-1\rangle = |1-2,-1+1\rangle = |-1,0\rangle .$$

In all, we see that we have generated the states:

$$\begin{aligned}
|1,0\rangle & \\
|-1,1\rangle &= E_{1-}|1,0\rangle \\
|2,-1\rangle &= E_{2-}E_{1-}|1,0\rangle \\
|0,0\rangle &= E_{1-}E_{2-}E_{1-}|1,0\rangle \\
|-2,1\rangle &= (E_{1-})^2 E_{2-}E_{1-}|1,0\rangle \\
|1,-1\rangle &= E_{2-}(E_{1-})^2 E_{2-}E_{1-}|1,0\rangle \\
|-1,0\rangle &= E_{1-}E_{2-}(E_{1-})^2 E_{2-}E_{1-}|1,0\rangle .
\end{aligned}$$

None of the states we have obtained have the same $\mathbf{H} = (H_1, H_2)$ eigenvalues, so there is no degeneracy. Thus a basis for the space is:

$$\{|1,0\rangle, |-1,1\rangle, |2,-1\rangle, |0,0\rangle, |-2,1\rangle, |1,-1\rangle, |-1,0\rangle\}.$$

(ii) Applying the algorithm to the highest weight state $|0,1\rangle$, which has eigenvalue $\mathbf{w}_2$ of $\mathbf{H} = (H_1, H_2)$, we have the following:

- We start from the state $|0,1\rangle$. We can apply E_{2-} precisely once and we cannot apply E_{1-}. We generate:

$$E_{2-}\,|0,1\rangle = |0+3,1-2\rangle = |3,-1\rangle\,.$$

- We now start from the state $|3,-1\rangle$. We can apply E_{1-} three times and we cannot apply E_{2-}. We generate:

$$E_{1-}\,|3,-1\rangle = |3-2,-1+1\rangle = |1,0\rangle\,,$$

$$E_{1-}\,|1,0\rangle = |1-2,0+1\rangle = |-1,1\rangle\,,$$

$$E_{1-}\,|-1,1\rangle = |-1-2,1+1\rangle = |-3,2\rangle\,.$$

- We already know what happens when we start from $|1,0\rangle$ and $|-1,1\rangle$, because we applied the algorithm to these states in part (i); we generate $\{|1,0\rangle\,, |-1,1\rangle\,, |2,-1\rangle\,, |0,0\rangle\,, |-2,1\rangle\,, |1,-1\rangle\,, |-1,0\rangle\}$. Thus we only need to start from $|-3,2\rangle$. We can apply E_{2-} precisely twice, but we cannot apply E_{1-}. We generate:

$$E_{2-}\,|-3,2\rangle = |-3+3,2-2\rangle = |0,0\rangle$$

$$E_{2-}\,|0,0\rangle = |0+3,0-2\rangle = |3,-2\rangle\,.$$

- We cannot start from $|0,0\rangle$, because we cannot apply E_{1-} or E_{2-}. On the other hand, we can start from $|3,-2\rangle$ and apply E_{1-} precisely three times (but we cannot apply E_{2-}). We generate:

$$E_{1-}\,|3,-2\rangle = |3-2,-2+1\rangle = |1,-1\rangle\,,$$

$$E_{1-}\,|1,-1\rangle = |1-2,-1+1\rangle = |-1,0\rangle\,,$$

$$E_{1-}\,|-1,0\rangle = |-1-2,0+1\rangle = |-3,1\rangle\,.$$

- Again, we have already started from $|1,-1\rangle$ and $|-1,0\rangle$ in previous parts, and we have already seen the states we obtain. We now start from $|-3,1\rangle$ instead. We can apply E_{2-} precisely once, but we cannot apply E_{1-}. We generate:

$$E_{2-}\,|-3,1\rangle = |-3+3,1-2\rangle = |0,-1\rangle\,.$$

This time, some states have been generated twice, so there is a possibility of degeneracy; this arises because some sequences of lowering could produce a different state to other sequences of lowering, e.g. $|0,0;1\rangle = (E_{2-})^2(E_{1-})^3E_{2-}\,|0,1\rangle$ versus $|0,0;2\rangle = E_{1-}E_{2-}(E_{1-})^2E_{2-}\,|0,1\rangle$, adding an additional label to represent the fact that these two states were obtained by different sequences of the lowering operators. As stated above, in general, it is rather difficult to decide when two sequences of lowering operator applications are in fact equivalent; this should be considered beyond the scope of the course.[4]

Instead, we shall use the hint from part (c); we are told that $|0,0\rangle$ is degenerate in this representation, so the above two sequences of lowering operator applications are inequivalent. Assuming there is no further degeneracy, we have constructed a basis:

$$\{|0,1\rangle\,, |3,-1\rangle\,, |1,0\rangle\,, |-1,1\rangle\,, |-3,2\rangle\,, |2,-1\rangle\,, |0,0;1\rangle\,, |0,0;2\rangle\,, |-2,1\rangle\,, |1,-1\rangle\,, |-1,0\rangle\,, |3,-2\rangle\,, |-3,1\rangle\,, |0,-1\rangle\}$$

for the representation space.

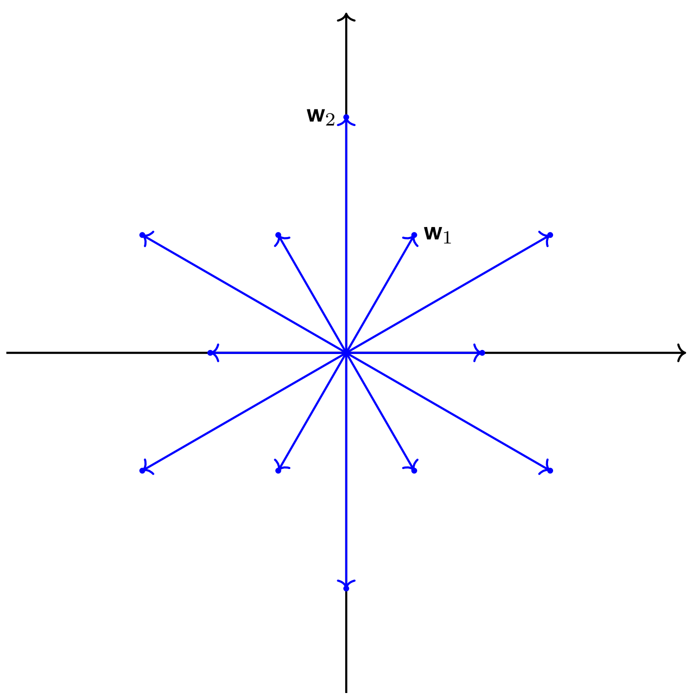
(c) Counting the sizes of the bases in parts (b)(i) and (b)(ii), we get the dimensions 7 and 14 respectively, as required.

[4] There are some general facts that can help us, though. For example, it is a general fact that the weights at the edges of the weight diagram are never degenerate.

(d) The weights in the 14 dimensional case take the explicit forms:

$$\begin{aligned}
\mathbf{w}_2 &= (0, \sqrt{3}) \\
3\mathbf{w}_1 - \mathbf{w}_2 &= \tfrac{1}{2}(3, \sqrt{3}) \\
\mathbf{w}_1 &= \tfrac{1}{2}(1, \sqrt{3}) \\
-\mathbf{w}_1 + \mathbf{w}_2 &= \tfrac{1}{2}(-1, \sqrt{3}) \\
-3\mathbf{w}_1 + 2\mathbf{w}_2 &= \tfrac{1}{2}(-3, \sqrt{3}) \\
2\mathbf{w}_1 - \mathbf{w}_2 &= (1, 0) \\
&\mathbf{0} \\
-2\mathbf{w}_1 + \mathbf{w}_2 &= \\
\mathbf{w}_1 - \mathbf{w}_2 &= \tfrac{1}{2}(1, -\sqrt{3}) \\
-\mathbf{w}_1 &= \tfrac{1}{2}(-1, -\sqrt{3}) \\
3\mathbf{w}_1 - 2\mathbf{w}_2 &= \tfrac{1}{2}(3, -\sqrt{3}) \\
-3\mathbf{w}_1 + \mathbf{w}_2 &= \tfrac{1}{2}(-3, -\sqrt{3}) \\
-\mathbf{w}_2 &= (0, 0\sqrt{3})
\end{aligned}$$

Thus the weight diagram is a six-point star:



We now compare to the root diagram. To construct all the roots, we use the algorithm presented in Question 5(c).

We start with $\Delta = \{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2\}$; before starting, it's useful to note that $\boldsymbol{\alpha}_1^2 = 1$, $\boldsymbol{\alpha}_2^2 = 3$ and $\boldsymbol{\alpha}_1 \cdot \boldsymbol{\alpha}_2 = -3/2$. Setting $S = \{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2\}$, we have:

- Examine roots of the form $\boldsymbol{\alpha}_2 + n\boldsymbol{\alpha}_1$. We have that $\boldsymbol{\alpha}_2$ is a root, and $\boldsymbol{\alpha}_2 - \boldsymbol{\alpha}_1$ is not a root, since $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2$ are simple roots. Hence $n_- = 0$. In particular, it follows that:

  $$n_+ = -\frac{2\boldsymbol{\alpha}_1 \cdot \boldsymbol{\alpha}_2}{\boldsymbol{\alpha}_1^2} = 3.$$

  Hence we have that $\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2, 2\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2$ and $3\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2$ are also roots which must be added to the set S.

- The set S is now given by $S = \{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2, 2\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2, 3\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2\}$. Considering roots of the form $s + n\boldsymbol{\alpha}_1$ for $s \in S$, i.e. roots of the form $\boldsymbol{\alpha}_2 + n\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2 + n\boldsymbol{\alpha}_1, 2\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2 + n\boldsymbol{\alpha}_1$ and $3\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2 + n\boldsymbol{\alpha}_1$, will not give us anything new, so we should instead start considering roots of $s + n\boldsymbol{\alpha}_2$.

  First, consider roots of type $\boldsymbol{\alpha}_1 + n\boldsymbol{\alpha}_2$. Here, $\boldsymbol{\alpha}_1$ is a root, but $\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2$ is not a root, hence $n_- = 0$. It follows that:

  $$n_+ = -\frac{2\boldsymbol{\alpha}_1 \cdot \boldsymbol{\alpha}_2}{\boldsymbol{\alpha}_2^2} = 1.$$

  Hence we generate the roots $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2$ which are already in S.

  Considering roots of type $\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2 + n\boldsymbol{\alpha}_2 = \boldsymbol{\alpha}_1 + (n+1)\boldsymbol{\alpha}_2$ will not give us anything new, and hence we are left considering roots of the type $2\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2 + n\boldsymbol{\alpha}_2 = 2\boldsymbol{\alpha}_1 + (n+1)\boldsymbol{\alpha}_2$ and the type $3\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2 + n\boldsymbol{\alpha}_2 = 3\boldsymbol{\alpha}_1 + (n+1)\boldsymbol{\alpha}_2$.

  In the first case, $2\boldsymbol{\alpha}_1 + (n+1)\boldsymbol{\alpha}_2$, when $n = 0$ we have a root, but when $n = -1$ we have that $2\boldsymbol{\alpha}_1$ is not a root because it is twice a root of a Lie algebra. Hence we have $n_- = 0$. Furthermore, when $n = 1$ we have that $2\boldsymbol{\alpha}_1 + 2\boldsymbol{\alpha}_2 = 2(\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2)$ is not a root because it is twice a root of a Lie algebra. Hence we have $n_+ = 0$, and it follows that the only root generated this way is $2\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2$.

  In the second case, $3\boldsymbol{\alpha}_1 + (n+1)\boldsymbol{\alpha}_2$, when $n = 0$ we have a root, but when $n = -1$ we have that $3\boldsymbol{\alpha}_1$ is not a root because it is three times a root of a Lie algebra. Hence we have $n_- = 0$. To obtain n_+, we evaluate:

  $$n_+ = -\frac{(3\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2) \cdot \boldsymbol{\alpha}_2}{\boldsymbol{\alpha}_2^2} = -\frac{3\boldsymbol{\alpha}_1 \cdot \boldsymbol{\alpha}_2}{\boldsymbol{\alpha}_2^2} - 1 = 2 - 1 = 1.$$

  Thus we have that $3\boldsymbol{\alpha}_1 + 2\boldsymbol{\alpha}_2$ is the only new root generated which needs to be added to S. Furthermore, root strings passing through this root won't generate anything new, so the algorithm terminates here.

It follows that the complete root set is:

$$\begin{aligned} S &= \{\pm\boldsymbol{\alpha}_1, \pm\boldsymbol{\alpha}_2, \pm(\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2), \pm(2\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2), \pm(3\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2), \pm(3\boldsymbol{\alpha}_1 + 2\boldsymbol{\alpha}_2)\} \\ &= \{\pm(2\mathbf{w}_1 - \mathbf{w}_2), \pm(-3\mathbf{w}_1 + 2\mathbf{w}_2), \pm(-\mathbf{w}_1 + \mathbf{w}_2), \pm\mathbf{w}_1, \pm(3\mathbf{w}_1 - \mathbf{w}_2), \pm\mathbf{w}_2\}. \end{aligned}$$

It follows that the roots exactly coincide with the (non-zero) weights we obtained earlier. Thus the 14-dimensional representation above is in fact the *adjoint* representation of the relevant Lie algebra.[5]

[5] Which in this case is G_2, by examining the Cartan classification.

Part III: Symmetries, Fields and Particles
Examples Sheet 4 Solutions

Please send all comments and corrections to mjb318@cam.ac.uk and jmm232@cam.ac.uk.

1. A Lie algebra has Cartan subalgebra $\underline{H} = (H_1, ..., H_r)$ and the remaining generators are $E_{\underline{\alpha}}$, corresponding to roots $\underline{\alpha}$, where $[\underline{H}, E_{\underline{\alpha}}] = \underline{\alpha} E_{\underline{\alpha}}$. Assume $[E_{\underline{\alpha}}, E_{-\underline{\alpha}}] = H_{\underline{\alpha}} = 2\underline{\alpha} \cdot \underline{H}/\underline{\alpha}^2$. For a root $\underline{\beta}$, $E_{\underline{\beta}}$ satisfies:

$$[E_{\underline{\alpha}}, E_{\underline{\beta}}] = 0, \qquad [H_{\underline{\alpha}}, E_{\underline{\beta}}] = nE_{\underline{\beta}}, \qquad \underbrace{[E_{-\underline{\alpha}}, [\dots, [E_{-\underline{\alpha}}}_{r \text{ times}}, E_{\underline{\beta}}], \dots] = E_{\underline{\beta} - r\underline{\alpha}}.$$

(a) Show that

$$[E_{\underline{\alpha}}, E_{\underline{\beta} - r\underline{\alpha}}] = r(n - r + 1) E_{\underline{\beta} - (r-1)\underline{\alpha}}.$$

(b) For n an integer, show that we may assume $E_{\underline{\beta} - (n+1)\underline{\alpha}} = 0$.

Solution: (a) Recall that the general formula for the commutator of a coroot $H_{\underline{\alpha}}$ with a step operator $E_{\underline{\beta}}$ is given by:

$$[H_{\underline{\alpha}}, E_{\underline{\beta}}] = \frac{2\underline{\alpha} \cdot \underline{\beta}}{\underline{\alpha}^2} E_{\underline{\beta}}$$

where $\underline{\alpha} \cdot \underline{\beta}$ is the inner product on the roots induced by the Killing form. Hence for the roots $\underline{\alpha}$, $\underline{\beta}$ in this question, we can immediately deduce that:

$$\frac{2\underline{\alpha} \cdot \underline{\beta}}{\underline{\alpha}^2} = n.$$

We can now prove the main result using induction. The base case $r = 0$ is trivial, since $[E_{\underline{\alpha}}, E_{\underline{\beta}}] = 0$. Now assume the result holds for $r = k$, and consider the case $r = k + 1$. We have:

$$\begin{aligned}
[E_{\underline{\alpha}}, E_{\underline{\beta}-(k+1)\underline{\alpha}}] &= [E_{\underline{\alpha}}, [E_{-\underline{\alpha}}, E_{\underline{\beta}-k\underline{\alpha}}]] && \text{(by definition)} \\
&= -[E_{-\underline{\alpha}}, [E_{\underline{\beta}-k\underline{\alpha}}, E_{\underline{\alpha}}]] - [E_{\underline{\beta}-k\underline{\alpha}}, [E_{\underline{\alpha}}, E_{-\underline{\alpha}}]] && \text{(by the Jacobi identity)} \\
&= k(n-k+1)[E_{-\underline{\alpha}}, E_{\underline{\beta}-(k-1)\underline{\alpha}}] - \frac{2\underline{\alpha}}{\underline{\alpha}^2} \cdot [E_{\underline{\beta}-k\underline{\alpha}}, \underline{H}] && \text{(induction hypothesis, definition of } H_{\underline{\alpha}}) \\
&= k(n-k+1)E_{\underline{\beta}-k\underline{\alpha}} + \frac{2\underline{\alpha}}{\underline{\alpha}^2} \cdot (\underline{\beta} - k\underline{\alpha}) E_{\underline{\beta}-k\underline{\alpha}} \\
&= k(n-k+1)E_{\underline{\beta}-k\underline{\alpha}} + (n-2k)E_{\underline{\beta}-k\underline{\alpha}} && \text{(using above result)} \\
&= (k+1)(n-k)E_{\underline{\beta}-k\underline{\alpha}}.
\end{aligned}$$

Hence the result follows by induction.

(b) Using part (a), we note that the subspace $W = \bigoplus_{r \in \mathbb{Z}} \operatorname{span}_{\mathbb{C}}\{E_{\underline{\beta}-r\underline{\alpha}}\}$ is an $\mathfrak{su}_{\mathbb{C}}(2)$ representation, since:

$$\operatorname{ad}(H_{\underline{\alpha}})E_{\underline{\beta}-r\underline{\alpha}} \propto E_{\underline{\beta}-r\underline{\alpha}}, \qquad \operatorname{ad}(E_{\pm\underline{\alpha}})E_{\underline{\beta}-r\underline{\alpha}} \propto E_{\underline{\beta}-(r\pm 1)\underline{\alpha}},$$

where $[H_{\underline{\alpha}}, E_{\pm\underline{\alpha}}] = \pm 2E_{\pm\underline{\alpha}}$ and $[E_{\underline{\alpha}}, E_{-\underline{\alpha}}] = H_{\underline{\alpha}}$. Assuming that the Lie algebra is finite-dimensional, the subspace W is a finite-dimensional irreducible representation of $\mathfrak{su}_{\mathbb{C}}(2)$, and hence there exists integral n such that $E_{\underline{\beta}-(n+1)\underline{\alpha}} = 0$.

2. Decompose the following tensor product representations of $A_2 = \mathfrak{su}(3)$ into irreducible components: (a) $\mathbf{3} \otimes \overline{\mathbf{3}}$ and (b) $\mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3}$. Discuss the connections between these irreducible representations and the quark model of light mesons and baryons [c.f. Wikipedia]. You might also think about the spin-$\frac{3}{2}$ baryons Δ^{++} and Ω^-, whose quark content is, respectively, uuu and sss, and what Pauli's exclusion principle implies about the quantum numbers of those quarks.

➼ **Solution:** Before beginning this question, recall from the Cartan classification that the Cartan matrix for A_2 is given by:

$$\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix},$$

which tells us that the fundamental weights are related to the roots of A_2 via:

$$\alpha_1 = 2\omega_1 - \omega_2,$$

$$\alpha_2 = -\omega_1 + 2\omega_2.$$

It is also useful to recall from lectures that the weight set of the fundamental representation of $A_2 = \mathfrak{su}(3)$ is:

$$S_{(1,0)} = \{\omega_1, \omega_2 - \omega_1, -\omega_2\},$$

and the weight set of the anti-fundamental representation of $A_2 = \mathfrak{su}(3)$ is:

$$S_{(0,1)} = \{\omega_2, \omega_1 - \omega_2, -\omega_1\}.$$

Let's now consider $\mathbf{3} \otimes \bar{\mathbf{3}}$. The weight set of this tensor product, allowing for repeated weights, is given by adding all weights of $\mathbf{3}$ to all weights of $\bar{\mathbf{3}}$ in all possible ways:

$$S_{\mathbf{3}\otimes\bar{\mathbf{3}}} = \{\omega_1 + \omega_2, 2\omega_1 - \omega_2, 0, 2\omega_2 - \omega_1, 0, \omega_2 - 2\omega_1, 0, \omega_1 - 2\omega_2, -\omega_1 - \omega_2\}.$$

To decompose this into irreducibles, we find dominant weights amongst this set and construct their corresponding weight sets. Observe that $\omega_1 + \omega_2$ is a dominant weight in this set. Using the algorithm presented in Examples Sheet 1, Question 8, we can construct the corresponding weight set as follows:

- Start with $\omega_1 + \omega_2$. From here, we may subtract α_1 once and α_2 once to give two more weights. The results are $-\omega_1 + 2\omega_2$ and $2\omega_1 - \omega_2$.
- Now beginning with $-\omega_1 + 2\omega_2$, we may subtract α_2 exactly twice to give two more weights. The results are 0 and $\omega_1 - 2\omega_2$.
- Instead beginning with $2\omega_1 - \omega_2$, we may subtract α_1 exactly twice to give two more weights. The results are 0 and $-2\omega_1 + \omega_2$.
- We are now left with $\omega_1 - 2\omega_2$ and $-2\omega_1 + \omega_2$, from which we may subtract α_1 and α_2 respectively to obtain two more weights. The result is $-\omega_1 - \omega_2$ in both cases.

Overall then, the weight set is:

$$S_{(1,1)} = \{\omega_1 + \omega_2, -\omega_1 + 2\omega_2, 2\omega_1 - \omega_2, 0, \omega_1 - 2\omega_2, -2\omega_1 + \omega_2, -\omega_1 - \omega_2\}.$$

We have to take care here because there may be repeated weights. It was shown in lectures that in fact the representation with Dynkin labels $(1,1)$ corresponds to the *adjoint* representation of the Lie algebra in this case. In particular, this implies that 0 is a repeated weight because the rank of A_2 is two, i.e. the Cartan subalgebra is two-dimensional. Thus including repeated weights, the weight set is in fact:

$$S_{(1,1)} = \{\omega_1 + \omega_2, -\omega_1 + 2\omega_2, 2\omega_1 - \omega_2, 0, 0, \omega_1 - 2\omega_2, -2\omega_1 + \omega_2, -\omega_1 - \omega_2\}.$$

With this calculation performed, we can conclude that:

$$S_{\mathbf{3}\otimes\bar{\mathbf{3}}} = S_{(1,1)} \cup S_{(0,0)},$$

which immediately yields $\mathbf{3} \otimes \bar{\mathbf{3}} = \mathbf{8} \otimes \mathbf{1}$.

To obtain $\mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3}$, we again repeatedly add the weights of $\mathbf{3}$ to themselves. In particular, for $\mathbf{3} \otimes \mathbf{3}$ we have the weights (allowing for repeats):

$$S_{\mathbf{3}\otimes\mathbf{3}} = \{2\omega_1, \omega_2, \omega_1 - \omega_2, \omega_2, 2\omega_2 - 2\omega_1, -\omega_1, \omega_1 - \omega_2, -\omega_1, -2\omega_2\}$$

and thus:

$$\begin{aligned} S_{\mathbf{3}\otimes\mathbf{3}\otimes\mathbf{3}} = \{&3\omega_1, \omega_1 + \omega_2, 2\omega_1 - \omega_2, \omega_1 + \omega_2, 2\omega_2 - \omega_1, 0, 2\omega_1 - \omega_2, 0, \omega_1 - 2\omega_2, \\ &\omega_2 + \omega_1, 2\omega_2 - \omega_1, 0, 2\omega_2 - \omega_1, 3\omega_2 - 3\omega_1, \omega_2 - 2\omega_1, 0, \omega_2 - 2\omega_1, -\omega_1 - \omega_2 \\ &2\omega_1 - \omega_2, 0, \omega_1 - 2\omega_2, 0, \omega_2 - 2\omega_1, -\omega_1 - \omega_2, \omega_1 - 2\omega_2, -\omega_1 - \omega_2, -3\omega_2\} \end{aligned}$$

In this case, the dominant weights are $3\omega_1$, $\omega_1 + \omega_2$ and 0. In particular, we should compute the weight set corresponding to $3\omega_1$ next, to identify which weights we can remove. We follow the standard algorithm:

- Given $3\omega_1$, we may subtract α_1 precisely three times to obtain three more weights. We obtain $\omega_1 + \omega_2$, $-\omega_1 + 2\omega_2$ and $-3\omega_1 + 3\omega_2$.
- Given $\omega_1 + \omega_2$, we may subtract α_2 precisely once. We obtain $2\omega_1 - \omega_2$. We may then subtract α_1 precisely twice; we obtain 0 and $-2\omega_1 + \omega_2$. We may finish this chain by subtracting α_2 precisely once, giving $-\omega_1 - \omega_2$.
- Next, given $-\omega_1 + 2\omega_2$, we may subtract α_2 precisely twice. We obtain 0 and $\omega_1 - 2\omega_2$. From here, we can subtract α_1 exactly once. We obtain $-\omega_1 - \omega_2$, where we terminate.
- Finally, given $-3\omega_1 + 3\omega_2$, we may subtract α_2 precisely three times. We obtain $-2\omega_1 + \omega_2$, $-\omega_1 - \omega_2$ and $-3\omega_2$, where everything terminates.

Thus the complete weight set, *not* accounting for possibly degenerate roots, is given by:

$$S_{(3,0)} = \{3\omega_1, \omega_1 + \omega_2, 2\omega_2 - \omega_1, 3\omega_2 - 3\omega_1, 2\omega_1 - \omega_2, 0, \omega_2 - 2\omega_1, \omega_1 - 2\omega_2, -\omega_2 - \omega_1, -3\omega_2\}$$

We shall assume that there are no degenerate weights in this case.[1] Overall then, we see that:

$$S_{\mathbf{3}\otimes\mathbf{3}\otimes\mathbf{3}} = S_{(3,0)} \cup S_{(1,1)} \cup S_{(1,1)} \cup S_{(0,0)},$$

so that we can immediately conclude:

$$\mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3} = \mathbf{10} \oplus \mathbf{8} \oplus \mathbf{8} \oplus \mathbf{1}.$$

In baryonic physics, we assume an $SU(3)$ flavour symmetry between the three lightest flavours of quarks, u, d and s. The flavour state of a quark is then described as a vector in the representation space $\mathbf{3}$; the first component refers to the u contribution, the second to the d contribution, and the third to the s contribution. Similarly, the representation space $\bar{\mathbf{3}}$ describes the flavour state of an *anti*-quark. Overall then, $\mathbf{3} \otimes \bar{\mathbf{3}}$ describes the flavour state of a composite system of a quark and anti-quark, called a *meson*.

The decomposition $\mathbf{3} \otimes \bar{\mathbf{3}} = \mathbf{8} \oplus \mathbf{1}$ shows us how physical mesons can transform under flavour symmetry. The eight weights of $\mathbf{8}$ correspond to the conserved quantum numbers of the meson *octet*, comprising kaons, pions and the η meson. The weight of $\mathbf{1}$ corresponds to the conserved quantum numbers of the η' meson.

[1] In fact, this can be proved using the *Weyl dimension formula* in this case, but this is beyond the scope of the course. For $A_2 = \mathfrak{su}(3)$, this states that the dimension of the irreducible with Dynkin indices (n, m) is given by $\frac{1}{2}(n+1)(m+1)(n+m+1)$.

On the other hand, $\mathbf{3}\otimes\mathbf{3}\otimes\mathbf{3}$ describes the flavour state of a composite system of three quarks, called a *baryon*. The decomposition $\mathbf{3}\otimes\mathbf{3}\otimes\mathbf{3}=\mathbf{10}\oplus\mathbf{8}\oplus\mathbf{8}\oplus\mathbf{1}$ shows us that physical baryons either lie in the baryon *decuplet*, the baryon *octet* (the two copies of the representation $\mathbf{8}$ do not give rise to separate baryons), and the baryon *singlet* (called the Λ).

Finally, we are asked about the quantum numbers of quarks in the two baryons Δ^{++} and Ω^-. These are completely flavour symmetric since their quark contents are uuu and sss respectively. Let us further assume that their wavefunctions reside in spaces of the form:

$$(\text{spatial})\otimes(\text{flavour})\otimes(\text{spin}),$$

In a ground state, these baryons will have spherically symmetric spatial wavefunctions; in particular, the spatial wavefunction will be totally symmetric on exchange of the constituent quarks. The spin space is:

$$\underline{\tfrac{1}{2}}\otimes\underline{\tfrac{1}{2}}\otimes\underline{\tfrac{1}{2}},$$

since quarks have spin-$1/2$. But since each factor of this tensor product is two-dimensional, totally antisymmetric elements of this space do not exist (and given that the Δ^{++} and Ω^- are spin-$3/2$ particles, we must lie in the top spin states anyway, which are in fact totally *symmetric*). Therefore, assuming only spatial, flavour and spin degrees of freedom for the baryon, it is impossible for the overall wavefunction of the three quarks comprising the baryons to be *antisymmetric*, which is required by the spin-statistics theorem. This implies the existence of an additional quantum number, namely *colour*.

3. Find the smallest-dimension, non-trivial, irreducible representation of B_2. Decompose the tensor product of two copies of this representation into irreps of B_2, giving the dimension of each component.

➾ **Solution:** The Lie algebra $B_2 = \mathfrak{so}(5)$ has Cartan matrix:

$$\begin{pmatrix} 2 & -1 \\ -2 & 2 \end{pmatrix},$$

by considering the Cartan classification. In particular, this is the same Lie algebras the one considered in Examples Sheet 3, Question 8(b). We saw in this question that the roots of the Lie algebra can be expressed in terms of the fundamental weights as:

$$\boldsymbol{\alpha}_1 = 2\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2,$$

$$\boldsymbol{\alpha}_2 = -2\boldsymbol{\mu}_1 + 2\boldsymbol{\mu}_2.$$

The theorem of the highest weight tells us that the irreducible representations of B_2 are in bijection with $\mathbb{N}^2$. To find the smallest irreducible representation, we consider the representations with the following Dynkin indices:

- $(1,0)$. We can obtain all the weights of the representation using the algorithm presented in Examples Sheet 3, Question 9(b). Start with the set $S_{(1,0)} = \{\boldsymbol{\mu}_1\}$. We can lower once with $\boldsymbol{\alpha}_1$, adding the weight $\boldsymbol{\mu}_1 - \boldsymbol{\alpha}_1 = -\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2$. From this weight, we can lower once with $\boldsymbol{\alpha}_2$, adding the weight $-\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2 - \boldsymbol{\alpha}_2 = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$. We can now lower again once with $\boldsymbol{\alpha}_1$, adding the weight $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 - \boldsymbol{\alpha}_1 = -\boldsymbol{\mu}_1$; the algorithm terminates here. Overall we have the weight set:
$$S_{(1,0)} = \{\boldsymbol{\mu}_1, -\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2, \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2, -\boldsymbol{\mu}_1\}.$$
Therefore, this representation is four-dimensional.

- $(0,1)$. Similarly, we start with the weight set $S_{(0,1)} = \{\boldsymbol{\mu}_2\}$. We can lower once with $\boldsymbol{\alpha}_2$, adding the weight $\boldsymbol{\mu}_2 - \boldsymbol{\alpha}_2 = 2\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$. We can now lower *twice* with $\boldsymbol{\alpha}_1$, adding the weights $2\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 - \boldsymbol{\alpha}_1 = 0$ and $-\boldsymbol{\alpha}_1 = -2\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2$. Next, we can lower once with $\boldsymbol{\alpha}_2$, adding the weight $-2\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2 - \boldsymbol{\alpha}_2 = -\boldsymbol{\mu}_2$; the algorithm terminates here. Overall we have the weight set:
$$S_{(0,1)} = \{\boldsymbol{\mu}_2, 2\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2, 0, -2\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2, -\boldsymbol{\mu}_2\}.$$
Therefore, this representation is five-dimensional.

- More generally, suppose that we start with Dynkin indices (n,m). The highest weight of the corresponding irreducible representation is $n\boldsymbol{\mu}_1 + m\boldsymbol{\mu}_2$. We already know from Examples Sheet 3, Question 8(e), that if we start with highest weight $2\boldsymbol{\mu}_1$, we generate the ten-dimensional adjoint representation. If we take $n > 2$, we can lower three times with $\boldsymbol{\alpha}_1$ immediately, generating at least a four-dimensional representation, excluding the possibility of anything smaller (indeed a quick check shows that we actually generate something larger).

  On the other hand, if we start with $2\boldsymbol{\mu}_2$ as our highest weight, we can lower twice with $\boldsymbol{\alpha}_2$ to get $2\boldsymbol{\mu}_1$ amongst our weights. This means we will at least generate the ten-dimensional adjoint representation weights again if we start with Dynkin indices $(0,2)$. Choosing $m > 2$, we can lower three times with $\boldsymbol{\alpha}_2$ immediately, generating at least a four-dimensional representation (and indeed we do get something larger on a quick check).

  The only remaining possibility is starting with the highest weight $\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2$. Lowering once with $\boldsymbol{\alpha}_1$, we get the weight $-\boldsymbol{\mu}_1 + 2\boldsymbol{\mu}_2$; on the other hand, lowering once with $\boldsymbol{\alpha}_2$ we get the weight $3\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$. Continuing, we generate more than four distinct weights, and hence get a larger dimensional representation than four.

We conclude that the smallest non-trivial irreducible representation of B_2 is four-dimensional. You might naturally ask if there is an intuitive reason for the smallest non-trivial irreducible representation of $B_2 = \mathfrak{so}(5)$ being four-dimensional and *not* five-dimensional. This is a consequence of the accidental isomorphism $B_2 \cong C_2 = \mathfrak{sp}(4,\mathbb{C})$; this shows that representations of the set of 5×5 matrices $\mathfrak{so}(5)$ can be naturally related to representations of the set of 4×4 matrices $\mathfrak{sp}(4,\mathbb{C})$.

Now let's consider taking the tensor product $\mathbf{4} \otimes \mathbf{4}$. The weight set of this tensor product, allowing for repeated weights, is given by adding all weights of $\mathbf{4}$ to all weights of $\mathbf{4}$ in all possible ways:

$$S = \{2\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, 2\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2, 0, \boldsymbol{\mu}_2, -2\boldsymbol{\mu}_1 + 2\boldsymbol{\mu}_2, 0, -2\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2, 2\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2, 0, 2\boldsymbol{\mu}_1 - 2\boldsymbol{\mu}_2, -\boldsymbol{\mu}_2, 0, -2\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2, -\boldsymbol{\mu}_2, -2\boldsymbol{\mu}_1\}.$$

We saw on Examples Sheet 3, Question 8(e), that the highest weight of the adjoint representation is $2\boldsymbol{\mu}_1$ (i.e. the adjoint representation has Dynkin indices $(0,2)$). The remaining weights of the representation are constructed using the standard algorithm. We lower twice with $\boldsymbol{\alpha}_1$ to add the weights $\boldsymbol{\mu}_2$ and $-2\boldsymbol{\mu}_1 + 2\boldsymbol{\mu}_2$. We can lower the first of these once with $\boldsymbol{\alpha}_2$ to obtain $2\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$, and we can lower the second of these twice with $\boldsymbol{\alpha}_2$ to obtain 0 and $2\boldsymbol{\mu}_1 - 2\boldsymbol{\mu}_2$. Both these weights can be lowered twice with $\boldsymbol{\alpha}_1$ to obtain $0, -2\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2$ and $-\boldsymbol{\mu}_2, -2\boldsymbol{\mu}_1$ respectively. Finally, $-2\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2$ can be lowered once with $\boldsymbol{\alpha}_2$ to obtain $-\boldsymbol{\mu}_2$. Overall, we find that:

$$S_{(0,2)} = \{2\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, -2\boldsymbol{\mu}_1 + 2\boldsymbol{\mu}_2, 2\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2, 0, 0, 2\boldsymbol{\mu}_1 - 2\boldsymbol{\mu}_2, -2\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2, -\boldsymbol{\mu}_2, -2\boldsymbol{\mu}_1\},$$

with zero the only weight that is repeated (see Examples Sheet 2, Question 8(d), for the reasoning). In particular, we see that the weight set of $\mathbf{4} \otimes \mathbf{4}$ can be decomposed as:

$$S = S_{(0,2)} \cup S_{(0,1)} \cup S_{(0,0)}.$$

It follows that we have $\mathbf{4} \otimes \mathbf{4} = \mathbf{10} \oplus \mathbf{5} \oplus \mathbf{1}$.

4. Consider a gauge theory whose gauge group G is a matrix Lie group. The corresponding gauge field,

$$A_\mu : \mathbb{R}^{1,3} \to \mathfrak{g}$$

transforms as:

$$A_\mu \mapsto A'_\mu = gA_\mu g^{-1} - (\partial_\mu g)g^{-1},$$

under a gauge transformation $g : \mathbb{R}^{1,3} \to G$. For the case $G = SU(N)$, check that $A'_\mu(x)$ takes values in the Lie algebra $\mathfrak{g}$. Explain why this is true for any matrix Lie group G. Writing $g = \exp(\epsilon X)$, with $\epsilon \ll 1$, show that the corresponding infinitesimal gauge transformation coincides with the one defined in the lectures.

Solution: We begin by showing that $A'_\mu \in \mathfrak{su}(N)$ for $A_\mu \in \mathfrak{su}(N)$ and $g \in SU(N)$ (recall that $\mathfrak{su}(N)$ is the set of $N \times N$ traceless anti-Hermitian matrices). Before beginning, it is useful to note that:

$$0 = \partial_\mu(I) = \partial_\mu(gg^{-1}) = (\partial_\mu g)g^{-1} + g\partial_\mu(g^{-1}) \qquad \Rightarrow \qquad (\partial_\mu g)g^{-1} = -g\partial_\mu(g^{-1}).$$

We now check that A'_μ is anti-Hermitian:

$$\begin{aligned}(A'_\mu)^\dagger &= g^{-\dagger}A_\mu^\dagger g^\dagger - g^{-\dagger}\partial_\mu g^\dagger \\ &= -gA_\mu g^{-1} - g\partial_\mu(g^{-1}) \\ &= -gA_\mu g^{-1} + (\partial_\mu g)g^{-1} \\ &= -A'_\mu,\end{aligned}$$

using the fact that $g^{-1} = g^\dagger$ for $g \in SU(N)$, and $A_\mu^\dagger = -A_\mu$ for $A_\mu \in \mathfrak{su}(N)$. To finish, we must check that A'_μ is traceless:

$$\begin{aligned}\mathrm{Tr}(A'_\mu) &= \mathrm{Tr}\left(gA_\mu g^{-1} - (\partial_\mu g)g^{-1}\right) \\ &= \mathrm{Tr}\left(gA_\mu g^{-1}\right) - \mathrm{Tr}\left((\partial_\mu g)g^{-1}\right) \\ &= \mathrm{Tr}\left(A_\mu\right) - \mathrm{Tr}\left((\partial_\mu g)g^{-1}\right) && \text{(cyclicity of trace)} \\ &= -\mathrm{Tr}\left((\partial_\mu g)g^{-1}\right) && \text{(tracelessness of } A_\mu)\end{aligned}$$

To finish, write $g(x) = \exp(X(x))$ for some $X(x) \in \mathfrak{su}(N)$. Then:

$$(\partial_\mu g)g^{-1} = (\partial_\mu X(x))\exp(X(x))\exp(-X(x)) = \partial_\mu X(x).$$

But $X(x) \in \mathfrak{su}(N)$ is traceless, so $\mathrm{Tr}\left((\partial_\mu g)g^{-1}\right) = \mathrm{Tr}(\partial_\mu X(x)) = \partial_\mu \mathrm{Tr}(X(x)) = 0$, and we're done.

Next, we show that for G a general matrix Lie group, we have that $A'_\mu \in \mathfrak{g}$. We shall show that $gA_\mu g^{-1}$ and $(\partial_\mu g)g^{-1}$ are members of $\mathfrak{g}$ separately.

First, fix some $g_0 \in G$ and some $X \in \mathfrak{g}$. Consider any curve $g(t) \in G$ satisfying $g(0) = g_0$, and construct a new curve $g(t)e^{tX}g(t)^{-1}$; this is certainly in G since $X \in \mathfrak{g}$ and the exponential map has codomain G. The tangent to this curve at $t = 0$ (corresponding to $g(0)g(0)^{-1} = I$, the identity) is given by:

$$\frac{d}{dt}\left(g(t)e^{tX}g(t)^{-1}\right)\bigg|_{t=0} = \frac{d}{dt}(g(t))\bigg|_{t=0} g(0)^{-1} + g(0)\frac{d}{dt}(g(t)^{-1})\bigg|_{t=0} + g(0)Xg(0)^{-1}.$$

But note that:

$$0 = \frac{d}{dt}(I)\bigg|_{t=0} = \frac{d}{dt}(g(t)g(t)^{-1})\bigg|_{t=0} = \frac{d}{dt}(g(t))\bigg|_{t=0} g(0)^{-1} + g(0)\frac{d}{dt}(g(t)^{-1})\bigg|_{t=0}.$$

Hence, everything simplifies to:

$$\frac{d}{dt}\left(g(t)e^{tX}g(t)^{-1}\right)\bigg|_{t=0} = g(0)Xg(0)^{-1} = g_0 X g_0^{-1}.$$

Thus, by definition, $g_0 X g_0^{-1} \in \mathfrak{g}$. Since g_0, X were arbitrary, we have $gA_\mu g^{-1} \in \mathfrak{g}$ generally for any $g(x) \in G$ and $A_\mu(x) \in \mathfrak{g}$, for each $\mu = 0, 1, 2, 3$.

Now let x_0 be an arbitrary point in spacetime, and let $x(t)$ be a curve in spacetime satisfying $x(0) = x_0$. Consider the curve in G given by $g(x(t))g(x_0)^{-1}$. Then at $t = 0$, i.e. at the identity $I \in G$, the tangent to the curve is given by:

$$\frac{d}{dt}\left(g(x(t))g(x_0)^{-1}\right)\bigg|_{t=0} = \dot{x}^\mu(0)(\partial_\mu g)(x_0)g(x_0)^{-1}.$$

Since $\dot{x}^\mu(0)$ is arbitrary (by choosing different curves $x(t)$), we have that $(\partial_\mu g)(x_0)g(x_0)^{-1} \in \mathfrak{g}$ for each $\mu = 0, 1, 2, 3$. But since x_0 was arbitrary, it follows that $(\partial_\mu g(x))g(x)^{-1} \in \mathfrak{g}$ for all spacetime points x.

In the final part of this question, we re-derive the infinitesimal results presented in lectures. We write $g = \exp(\epsilon X)$ and work to linear order throughout, giving:

$$\begin{aligned} A'_\mu &= gA_\mu g^{-1} - (\partial_\mu g)g^{-1} \\ &= (1 + \epsilon X)A_\mu(1 - \epsilon X) - \epsilon(\partial_\mu X)(1 - \epsilon X) + O(\epsilon^2) \\ &= A_\mu + \epsilon[X, A_\mu] - \epsilon\partial_\mu X + O(\epsilon^2), \end{aligned}$$

implying that $\delta A_\mu = -\epsilon\partial_\mu X + \epsilon[X, A_\mu]$, as required.

5. For a group with a Lie algebra with a basis $\{T_a\}$ such that $[T_a, T_b] = f^c{}_{ab} T_c$, let $\kappa_{ab} = (T_a, T_b)$ where $(\cdot, \cdot)$ is an invariant symmetric bilinear form, so that $([X, Y], Z) = (-Y, [X, Z])$.

(i) If D_μ is an appropriate covariant derivative involving a gauge field A^a_μ, verify:

$$\partial_\mu (X(x), Y(x)) = (D_\mu X(x), Y(x)) + (X(x), D_\mu Y(x)).$$

(ii) Let $T^\mu{}_\nu = (F^{\mu\sigma}, F_{\nu\sigma}) - \frac{1}{4}\delta^\mu{}_\nu (F^{\sigma\rho}, F_{\sigma\rho})$. Show, using the Bianchi identity, $\partial_\mu T^\mu{}_\nu = (D_\mu F^{\mu\sigma}, F_{\nu\sigma})$.

(iii) For a variation δA^a_μ, obtain also:

$$\delta \left(\frac{1}{4} \epsilon^{\mu\nu\sigma\rho} (F_{\mu\nu}, F_{\sigma\rho}) \right) = \partial_\mu \epsilon^{\mu\nu\sigma\rho} (\delta A_\nu, F_{\sigma\rho}).$$

(iv) By letting $A_\mu \mapsto t A_\mu$ and differentiating with respect to t, and then integrating, show that:

$$\frac{1}{4} \epsilon^{\mu\nu\sigma\rho} (F_{\mu\nu}, F_{\sigma\rho}) = \partial_\mu \left[\epsilon^{\mu\nu\sigma\rho} \left(A_\nu, \partial_\sigma A_\rho + \frac{1}{3}[A_\sigma, A_\rho] \right) \right].$$

➻ **Solution:** (a) Since $X(x), Y(x)$ are being inserted into the slots of the given symmetric bilinear form, which takes as its arguments elements of the Lie algebra, we have that the covariant derivative D_μ is acting in the *adjoint* representation:

$$D_\mu X(x) := \partial_\mu X(x) + \lambda [A_\mu(x), X(x)],$$

where λ is some gauge coupling and $A_\mu(x)$ is a gauge field.

Now working from the right hand side of the given identity to the left hand side, we have:

$$\begin{aligned}
(D_\mu X(x), Y(x)) + (X(x), D_\mu Y(x)) &= (\partial_\mu X(x) + \lambda [A_\mu(x), X(x)], Y(x)) + (X(x), \partial_\mu Y(x) + \lambda [A_\mu(x), Y(x)]) \\
&= (\partial_\mu X(x), Y(x)) + (X(x), \partial_\mu Y(x)) \\
&\qquad + \lambda \left(([A_\mu(x), X(x)], Y(x)) + (X(x), [A_\mu(x), Y(x)]) \right) \\
&= (\partial_\mu X(x), Y(x)) + (X(x), \partial_\mu Y(x)),
\end{aligned}$$

where the last line follows by invariant of the symmetric bilinear form. To finish, expand $X(x) = X^a(x) T_a$, $Y(x) = Y^a(x) T_a$:

$$\begin{aligned}
(\partial_\mu X(x), Y(x)) + (X(x), \partial_\mu Y(x)) &= \left(\partial_\mu X^a(x) Y^b(x) + X^a(x) \partial_\mu Y^b(x) \right) (T_a, T_b) \\
&= \partial_\mu (X^a(x) Y^b(x)) (T_a, T_b) \\
&= \partial_\mu (X(x), Y(x)),
\end{aligned}$$

as required. $\square$

(b) We begin by recalling the *Bianchi identity*:

$$D_\mu F_{\nu\rho} + D_\nu F_{\rho\mu} + D_\rho F_{\mu\nu} = 0. \tag{$*$}$$

To prove this identity, we start from the *Jacobi identity* for the covariant derivative:

$$[D_\mu, [D_\nu, D_\rho]] + [D_\nu, [D_\rho, D_\mu]] + [D_\rho, [D_\mu, D_\nu]] = 0;$$

this can be proved simply by writing out all the terms on the left hand side (you don't need to expand the covariant derivatives though!). Indeed, the Jacobi identity holds whenever the commutator of two objects, $[X, Y] = XY - YX$ involves an *associative* product on the right hand side (which is indeed the case here, since $D_\mu(D_\nu D_\rho) = (D_\mu D_\nu)D_\rho$.

For an arbitrary element of the adjoint representation X, we can now note:

$$\begin{aligned}
[D_\mu, [D_\nu, D_\rho]]X &= \lambda[D_\mu, F_{\nu\rho}]X && \text{(definition of } F_{\mu\nu}\text{)} \\
&= \lambda D_\mu(F_{\nu\rho}X) - \lambda F_{\nu\rho}D_\mu X \\
&= \lambda D_\mu F_{\nu\rho} X.
\end{aligned}$$

Thus $[D_\mu, [D_\nu, D_\rho]] = \lambda D_\mu F_{\nu\rho}$, and the Bianchi identity ($*$) follows.

We now examine $\partial_\mu T^\mu{}_\nu$. Computing this quantity directly, we have:

$$\begin{aligned}
\partial_\mu T^\mu{}_\nu &= \partial_\mu(F^{\mu\sigma}, F_{\nu\sigma}) - \frac{1}{4}\delta^\mu{}_\nu \partial_\mu(F^{\sigma\rho}, F_{\sigma\rho}) \\
&= (D_\mu F^{\mu\sigma}, F_{\nu\sigma}) + (F^{\mu\sigma}, D_\mu F_{\nu\sigma}) - \frac{1}{4}(D_\nu F^{\sigma\rho}, F_{\sigma\rho}) - \frac{1}{4}(F^{\sigma\rho}, D_\nu F_{\sigma\rho}) && \text{(part (i))} \\
&= (D_\mu F^{\mu\sigma}, F_{\nu\sigma}) + (F^{\rho\sigma}, D_\rho F_{\nu\sigma}) - \frac{1}{2}(D_\nu F_{\sigma\rho}, F^{\sigma\rho}) && \text{(symmetry)} \\
&= (D_\mu F^{\mu\sigma}, F_{\nu\sigma}) + (F^{\rho\sigma}, D_\rho F_{\nu\sigma}) + \frac{1}{2}(D_\sigma F_{\rho\nu}, F^{\sigma\rho}) + \frac{1}{2}(D_\rho F_{\nu\sigma}, F^{\sigma\rho}) && \text{(Bianchi identity)} \\
&= (D_\mu F^{\mu\sigma}, F_{\nu\sigma}) + (F^{\rho\sigma}, D_\rho F_{\nu\sigma}) - \frac{1}{2}(F^{\sigma\rho}, D_\sigma F_{\nu\rho}) - \frac{1}{2}(F^{\rho\sigma}, D_\rho F_{\nu\sigma}) \\
&= (D_\mu F^{\mu\sigma}, F_{\nu\sigma}),
\end{aligned}$$

as required.

This part of the question is interesting, because the equations of motion for free Yang-Mills theory are $D_\mu F^{\mu\sigma} = 0$. In particular, we can conclude that $\partial_\mu T^\mu{}_\nu = 0$, so that $T^\mu{}_\nu$ is a conserved current for each ν.

(c) We use the fact that the infinitesimal variation of $F_{\mu\nu}$ under a gauge transformation is:

$$\delta F_{\mu\nu} = D_\mu \delta A_\nu - D_\nu \delta A_\mu.$$

To prove this, we consider an infinitesimal variation of the field strength tensor:

$$\begin{aligned}
\delta F_{\mu\nu} &= \delta\left(\partial_\mu A_\nu - \partial_\nu A_\mu + \lambda[A_\mu, A_\nu]\right) \\
&= \delta\left(\partial_\mu A_\nu - \partial_\nu A_\mu + \lambda(A_\mu A_\nu - A_\nu A_\mu)\right) \\
&= \partial_\mu \delta A_\nu - \partial_\nu \delta A_\mu + \lambda \delta A_\mu A_\nu + \lambda A_\mu \delta A_\nu - \lambda \delta A_\nu A_\mu - \lambda A_\nu \delta A_\mu \\
&= (\partial_\mu \delta A_\nu + \lambda[A_\mu, \delta A_\nu]) - (\partial_\nu \delta A_\mu + \lambda[A_\nu, \delta A_\mu]) \\
&= D_\mu \delta A_\nu - D_\nu \delta A_\mu.
\end{aligned}$$

Using the infinitesimal variation in the field strength tensor, and the Bianchi identity, we can establish the required result:

$$\begin{aligned}
\delta\left(\frac{1}{4}\epsilon^{\mu\nu\sigma\rho}(F_{\mu\nu}, F_{\sigma\rho})\right) &= \frac{1}{4}\kappa_{ab}\epsilon^{\mu\nu\sigma\rho}\delta\left(F^a_{\mu\nu}F^b_{\sigma\rho}\right) \\
&= \frac{1}{4}\kappa_{ab}\epsilon^{\mu\nu\sigma\rho}\left((D_\mu \delta A^a_\nu - D_\nu \delta A^a_\mu)F^b_{\sigma\rho} + F^a_{\mu\nu}(D_\sigma \delta A^b_\rho - D_\rho \delta A^b_\sigma)\right) \\
&= \kappa_{ab}\epsilon^{\mu\nu\sigma\rho}(D_\mu \delta A^a_\nu)F^b_{\sigma\rho} && \text{(relabelling indices)} \\
&= \epsilon^{\mu\nu\sigma\rho}(D_\mu \delta A_\nu, F_{\sigma\rho}) \\
&= \epsilon^{\mu\nu\sigma\rho}\left(\partial_\mu(\delta A_\nu, F_{\sigma\rho}) - (\delta A_\nu, D_\mu F_{\sigma\rho})\right) && \text{(part (i))} \\
&= \partial_\mu \epsilon^{\mu\nu\sigma\rho}(\delta A_\nu, F_{\sigma\rho}) - (\delta A_\nu, \epsilon^{\mu\nu\sigma\rho} D_\mu F_{\sigma\rho}).
\end{aligned}$$

The first term is the one we want. The remaining term is zero, by the Bianchi identity (for each fixed ν, we are antisymmetrising $D_\mu F_{\sigma\rho}$ over μ, σ, ρ, and since $F_{\sigma\rho}$ is already antisymmetric, we get precisely the relevant Bianchi quantity in each case). Thus we have established:

$$\delta\left(\frac{1}{4}\epsilon^{\mu\nu\sigma\rho}(F_{\mu\nu}, F_{\sigma\rho})\right) = \partial_\mu \epsilon^{\mu\nu\sigma\rho}(\delta A_\nu, F_{\sigma\rho}),$$

as required.

(d) For the final part, we follow the instructions of the question and begin by sending $A_\mu \mapsto tA_\mu$, then differentiating with respect to t. First, it will be useful to define:

$$F^t_{\mu\nu} = t\partial_\mu A_\nu - t\partial_\nu A_\mu + \lambda t^2[A_\mu, A_\nu].$$

By part (iii), we have:

$$\delta\left(\frac{1}{4}\epsilon^{\mu\nu\sigma\rho}(F^t_{\mu\nu}, F^t_{\sigma\rho})\right) = \partial_\mu \epsilon^{\mu\nu\sigma\rho}(\delta(tA_\nu), F^t_{\sigma\rho}),$$

Dividing through by δt, a small change in t, we have:

$$\frac{\delta}{\delta t}\left(\frac{1}{4}\epsilon^{\mu\nu\sigma\rho}(F^t_{\mu\nu}, F^t_{\sigma\rho})\right) = \partial_\mu \epsilon^{\mu\nu\sigma\rho}\left(\frac{\delta}{\delta t}(tA_\nu), F^t_{\sigma\rho}\right).$$

Taking the limit as $\delta t \to 0$, we get t-derivatives, yielding:

$$\frac{d}{dt}\left(\frac{1}{4}\epsilon^{\mu\nu\sigma\rho}(F^t_{\mu\nu}, F^t_{\sigma\rho})\right) = \partial_\mu \epsilon^{\mu\nu\sigma\rho}\left(A_\nu, F^t_{\sigma\rho}\right).$$

We now continue with the advice of the question, and integrate t from 0 to 1. On the left hand side, we have:

$$\int_0^1 dt\, \frac{d}{dt}\left(\frac{1}{4}\epsilon^{\mu\nu\sigma\rho}(F^t_{\mu\nu}, F^t_{\sigma\rho})\right) = \frac{1}{4}\epsilon^{\mu\nu\sigma\rho}(F_{\mu\nu}, F_{\sigma\rho}).$$

On the right hand side, we have:

$$\begin{aligned}\int_0^1 dt\, \partial_\mu \epsilon^{\mu\nu\sigma\rho}\left(A_\nu, F^t_{\sigma\rho}\right) &= \int_0^1 dt\, \partial_\mu \epsilon^{\mu\nu\sigma\rho}\left(A_\nu, t\partial_\sigma A_\rho - t\partial_\rho A_\sigma + \lambda t^2[A_\sigma, A_\rho]\right) \\ &= \partial_\mu \epsilon^{\mu\nu\sigma\rho}\left(A_\nu, \frac{1}{2}\partial_\sigma A_\rho - \frac{1}{2}\partial_\rho A_\sigma + \frac{\lambda}{3}[A_\sigma, A_\rho]\right).\end{aligned}$$

To finish, note that the second term's indices can be relabelled, yielding the final right hand side:

$$\mathsf{RHS} = \partial_\mu \epsilon^{\mu\nu\sigma\rho}\left(A_\nu, \partial_\sigma A_\rho + \frac{\lambda}{3}[A_\sigma, A_\rho]\right).$$

Hence we have the required result:

$$\frac{1}{4}\epsilon^{\mu\nu\sigma\rho}(F_{\mu\nu}, F_{\sigma\rho}) = \partial_\mu\left[\epsilon^{\mu\nu\sigma\rho}\left(A_\nu, \partial_\sigma A_\rho + \frac{1}{3}[A_\sigma, A_\rho]\right)\right].$$

(Note the question uses a different convention for our gauge couplings; taking $\lambda = 1$ gives agreement in this case.)

❊ **Comments:** The last part of the question shows us a very interesting result. The term $\epsilon^{\mu\nu\sigma\rho}(F_{\mu\nu}, F_{\sigma\rho})$ is a gauge-invariant term which could in principle be included in the Yang-Mills Lagrangian. However, in parts (iii) and (iv) we showed that this term is in fact a total derivative, so its inclusion will not affect the equations of motion, so it can be neglected.

This applies *provided* we are working with fields which decay to zero at spatial infinity. Otherwise, total derivatives in the Lagrangian can be impactful; this can allow us to develop interesting theories, which only depend on the boundary data for the fields, not the behaviour of the fields in between - these theories are *topological* in nature, and motivates the study of *topological quantum field theories*.

6. With notation as in the previous question, define a three-dimensional Lagrangian:

$$\mathcal{L} = \epsilon^{\mu\nu\rho}\left(\kappa_{ab}A^a_\mu \partial_\nu A^b_\rho + \frac{1}{3}f_{abc}A^a_\mu A^b_\nu A^c_\rho\right).$$

For a gauge transformation $\delta A^a_\mu = -\partial_\mu \lambda^a - f^a{}_{bc}A^b_\mu \lambda^c$, show that $\delta\mathcal{L} = -\partial_\mu(\epsilon^{\mu\nu\rho}\kappa_{ab}\lambda^a\partial_\nu A^b_\rho)$ so that:

$$\int d^3x\, \mathcal{L}$$

is invariant.

➾ **Solution:** Before beginning, it is useful to note that invariance $([X,Y],Z) = (-Y,[X,Z])$ implies that:

$$\kappa_{ab}f^b{}_{cd}X^aY^cZ^d = -\kappa_{cb}f^b{}_{ad}Y^cX^aZ^d \qquad \Rightarrow \qquad f_{acd} + f_{cad} = 0,$$

since the Killing form lowers group-theoretic indices. In particular, we see that the structure constants with lowered indices are completely antisymmetric.

With this fact, we can compute the infinitesimal variation directly:

$$\begin{aligned}
\delta\mathcal{L} &= \epsilon^{\mu\nu\rho}\kappa_{ab}\delta A^a_\mu\partial_\nu A^b_\rho + \epsilon^{\mu\nu\rho}\kappa_{ab}A^a_\mu\partial_\nu\delta A^b_\rho + \epsilon^{\mu\nu\rho}f_{abc}\delta A^a_\mu A^b_\nu A^c_\rho && \text{(antisymmetry in the third term)}\\
&= \epsilon^{\mu\nu\rho}\kappa_{ab}(-\partial_\mu\lambda^a - f^a{}_{cd}A^c_\mu\lambda^d)\partial_\nu A^b_\rho + \epsilon^{\mu\nu\rho}\kappa_{ab}A^a_\mu\partial_\nu\left(-\partial_\rho\lambda^b - f^b{}_{cd}A^c_\rho\lambda^d\right)\\
&\quad + \epsilon^{\mu\nu\rho}f_{abc}\left(-\partial_\mu\lambda^a - f^a{}_{de}A^d_\mu\lambda^e\right)A^b_\nu A^c_\rho\\
&= -\partial_\mu\left(\epsilon^{\mu\nu\rho}\kappa_{ab}\lambda^a\partial_\nu A^b_\rho\right) - \epsilon^{\mu\nu\rho}f_{bcd}A^c_\mu\lambda^d\partial_\nu A^b_\rho - \epsilon^{\mu\nu\rho}f_{acd}A^a_\mu\partial_\nu(A^c_\rho\lambda^d)\\
&\quad - \epsilon^{\mu\nu\rho}f_{abc}\partial_\mu\lambda^a A^b_\nu A^c_\rho - \epsilon^{\mu\nu\rho}f_{abc}f^a{}_{de}\lambda^e A^d_\mu A^b_\nu A^c_\rho\\
&= -\partial_\mu\left(\epsilon^{\mu\nu\rho}\kappa_{ab}\lambda^a\partial_\nu A^b_\rho\right) - \epsilon^{\mu\nu\rho}f_{bcd}A^c_\mu\lambda^d\partial_\nu A^b_\rho + \epsilon^{\mu\nu\rho}f_{bcd}A^c_\mu\partial_\nu A^b_\rho\lambda^d\\
&\quad + \epsilon^{\mu\nu\rho}f_{abc}A^b_\nu A^c_\rho\partial_\mu\lambda^a - \epsilon^{\mu\nu\rho}f_{abc}\partial_\mu\lambda^a A^b_\nu A^c_\rho - \epsilon^{\mu\nu\rho}f_{abc}f^a{}_{de}\lambda^e A^d_\mu A^b_\nu A^c_\rho && \text{(relabelling indices)}\\
&= -\partial_\mu\left(\epsilon^{\mu\nu\rho}\kappa_{ab}\lambda^a\partial_\nu A^b_\rho\right) - \epsilon^{\mu\nu\rho}f_{abc}f^a{}_{de}\lambda^e A^d_\mu A^b_\nu A^c_\rho.
\end{aligned}$$

It remains to show that the final term is zero. Rewriting things in index-free notation, we have:

$$\begin{aligned}
\epsilon^{\mu\nu\rho}f_{abc}f^a{}_{de}\lambda^e A^d_\mu A^b_\nu A^c_\rho &= \epsilon^{\mu\nu\rho}([A_\mu,\lambda],[A_\nu,A_\rho])\\
&= -\epsilon^{\mu\nu\rho}\left(\lambda,[A_\mu,[A_\nu,A_\rho]]\right) && \text{(invariance)}\\
&= 0,
\end{aligned}$$

by the Jacobi identity, $[A_\mu,[A_\nu,A_\rho]] + [A_\nu,[A_\rho,A_\mu]] + [A_\rho,[A_\mu,A_\nu]] = 0$, which gives:

$$\begin{aligned}
0 &= \epsilon^{\mu\nu\rho}[A_\mu,[A_\nu,A_\rho]] + \epsilon^{\mu\nu\rho}[A_\nu,[A_\rho,A_\mu]] + \epsilon^{\mu\nu\rho}[A_\rho,[A_\mu,A_\nu]]\\
&= 3\epsilon^{\mu\nu\rho}[A_\mu,[A_\nu,A_\rho]].
\end{aligned}$$