

I extracted this set of notes from the book

Group Theory in a Nutshell for Physicists by A. Zee

I skipped a lot of sections in the book because I was only trying to talk about the topics that I did include.

These were the topics that make sure you have seen certain ideas and their consequences.

If you read these notes, then you will learn a lot about important ideas that are not directly Group Theory but are related in important ways.

1 Symmetry and Groups

Symmetry and Transformation

Symmetry plays a central role in modern theoretical physics.

As the etymologist tells us, symmetry (“equal measure”) originates in geometry (“earth measure”).

We have a sense that an isosceles triangle is more symmetrical than an arbitrary triangle and that an equilateral triangle is more symmetrical than an isosceles triangle.

Going further, we feel that a pentagon is more symmetrical than a square, a hexagon more symmetrical than a pentagon, and an $(n + 1)$ -sided regular polygon is more symmetrical than an n -sided regular polygon.

And finally, a circle is more symmetrical than any regular polygon.

The n -sided regular polygon is left unchanged by rotations through any angle that is an integer multiple of $2\pi/n$, and there are n of these rotations.

The larger n is, the more such rotations there are.

This is why mathematicians and physicists feel that the hexagon is more symmetrical than a pentagon: $6 > 5$, QED.

To quantify this intuitive feeling, we should thus look at the set of transformations that leave the geometrical figure unchanged (that is, invariant).

For example, we can reflect the isosceles triangle across the median that divides it into equal parts (see figure 1a).

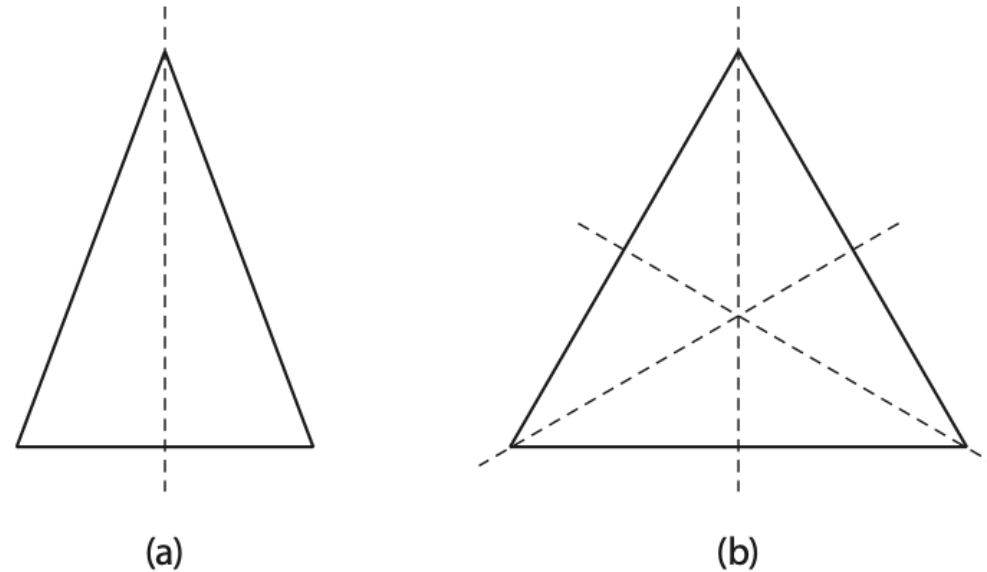


Figure 1

Call the reflection r ; then the set of transformations that leave the isosceles triangle invariant is given by $\{I, r\}$, where I denotes the identity transformation, that is, the transformation that does nothing.

A reflection followed by a reflection has the same effect as the identity transformation.

We write this statement as $r \cdot r = I$.

In contrast, the equilateral triangle is left invariant not only by reflection across any of its three medians (figure 1b) but also by rotation R_1 through $2\pi/3 = 120^\circ$ around its center, as well as rotation R_2 through $4\pi/3 = 240^\circ$.

The set of transformations that leave the equilateral triangle invariant is thus given by $\{I, r_1, r_2, r_3, R_1, R_2\}$.

That this set is larger than the set in the preceding paragraph quantifies the feeling that the equilateral triangle is more symmetrical than the isosceles triangle.

Note that $R_1 \cdot R_1 = R_2$ and that $R_1 \cdot R_2 = I$.

A circle is left invariant by an infinite number of transformations, namely, rotation $R(\theta)$ through any angle θ , and reflection across any straight line through its center.

The circle is more symmetrical than any regular polygon, such as the equilateral triangle.

Symmetry in physics

In physics we are interested in the symmetries enjoyed by a given physical system.

On a more abstract level, we are interested in the symmetries of the fundamental laws of physics.

One of the most revolutionary and astonishing discoveries in the history of physics is that objects do not fall down, but toward the center of the earth.

Newton's law of gravitation does not pick out a special direction: it is left invariant by rotations.

The history of theoretical physics has witnessed the discoveries of one unexpected symmetry after another.

Physics in the late twentieth century consists of the astonishing discovery that as we study Nature at ever deeper levels, Nature displays more and more symmetries.

Consider a set of transformations $T_1, T_2, \dots$ that leave the laws of physics invariant.

Let us first perform the transformation T_j , and then perform the transformation T_i .

The transformation that results from this sequence of two transformations is denoted by the “product” $T_i \cdot T_j$.

Evidently, if T_i and T_j leave the laws of physics invariant, then the transformation $T_i \cdot T_j$ also leaves the laws of physics invariant.

Here we label the transformations by a discrete index i .

In general, the index could also be continuous.

Indeed, the transformation could depend on a number of continuous parameters.

The classic example is a rotation $R(\theta, \varphi, \zeta)$, which can be completely characterized by three angles, as indicated.

For example, in one standard parametrization, the two angles θ and φ specify the unit vector describing the rotation axis, while the angle ζ specifies the angle through which we are supposed to rotate around that axis.

Groups

This discussion rather naturally invites us to abstract the concept of a group.

A group G consists of a set of entities $\{g_\alpha\}$ called group elements (or elements for short), which we could compose together (or more colloquially, multiply together).

Given any two elements g_α and g_β , the product $g_\alpha \cdot g_\beta$ is equal to another element, say, g_γ , in G .

In other words, $g_\alpha \cdot g_\beta = g_\gamma$.

Composition or multiplication is indicated by a dot (which I usually omit if there is no danger of confusion).

The set of all relations of the form $g_\alpha \cdot g_\beta = g_\gamma$ is called the multiplication table of the group.

Composition or multiplication (we will use the two words interchangeably) satisfies the following axioms:

1. Associativity: Composition is associative: $(g_\alpha \cdot g_\beta) \cdot g_\gamma = g_\alpha \cdot (g_\beta \cdot g_\gamma)$.
2. Existence of the identity: There exists a group element, known as the identity and denoted by I , such that $I \cdot g_\alpha = g_\alpha$ and $g_\alpha \cdot I = g_\alpha$.
3. Existence of the inverse: For every group element g_α , there exists a unique group element, known as the inverse of g_α and denoted by g_α^{-1} , such that $g_\alpha^{-1} \cdot g_\alpha = I$ and $g_\alpha \cdot g_\alpha^{-1} = I$.

A number of comments follow.

1. Composition is not required to commute. In general, $g_\alpha \cdot g_\beta$ is not equal to $g_\beta \cdot g_\alpha$. In this respect, the multiplication of group elements is, in general, like the multiplication of matrices but unlike that of ordinary numbers.

A group for which the composition rule is commutative is said to be abelian,[‡] and a group for which this is not true is said to be nonabelian.[§]

2. The right inverse and the left inverse are by definition the same. We can imagine mathematical structures for which this is not true, but then these structures are not groups. Recall (or read in the review of linear algebra) that this property holds for square matrices: provided that the inverse M^{-1} of a matrix M exists, we have $M^{-1}M = MM^{-1} = I$ with I the identity matrix.
3. It is often convenient to denote I by g_0 .
4. The label α that distinguishes the group element g_α may be discrete or continuous.
5. The set of elements may be finite (that is, $\{g_0, g_1, g_2, \dots, g_{n-1}\}$), in which case G is known as a finite group with n elements. (Our friend the jargon guy⁷ informs us that n is known as the order of the group.)

Mathematician of course can study groups on the abstract level without tying g_i to any physical transformation, but in some sense the axioms become clearer if we think of transformations in the back of our mind.

For example, $gI = Ig = g$ says that the net effect of first doing nothing and then doing something is the same as first doing something and then doing nothing, and the same as doing something.

Existence of the inverse says that the transformations of interest to physics can always be undone.

Examples of groups

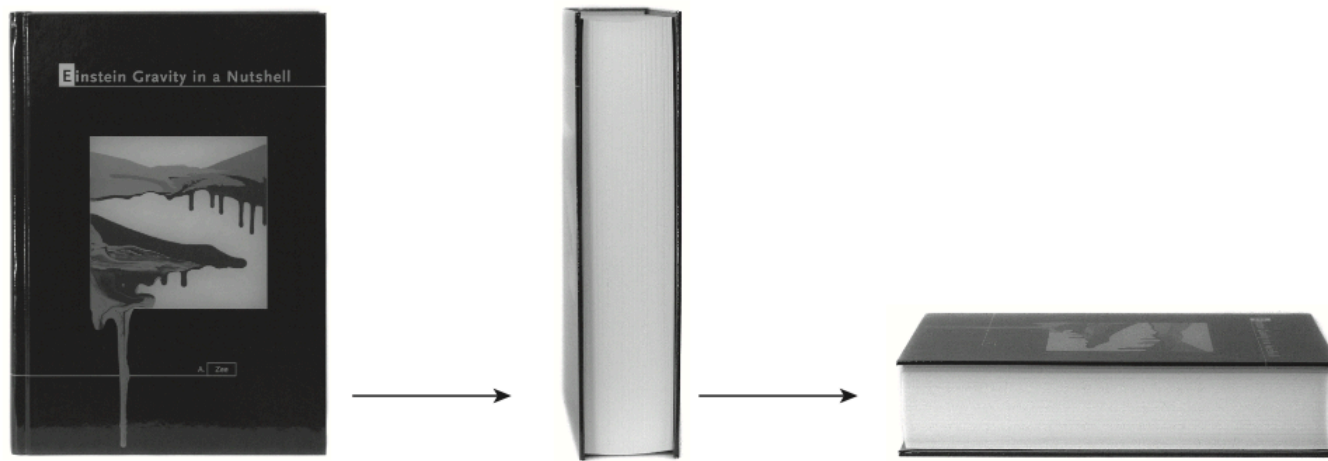
To gain a better understanding of what a group is, it is best to go through a bunch of examples.

For each of the following examples, it is easy to verify that the group axioms are satisfied.

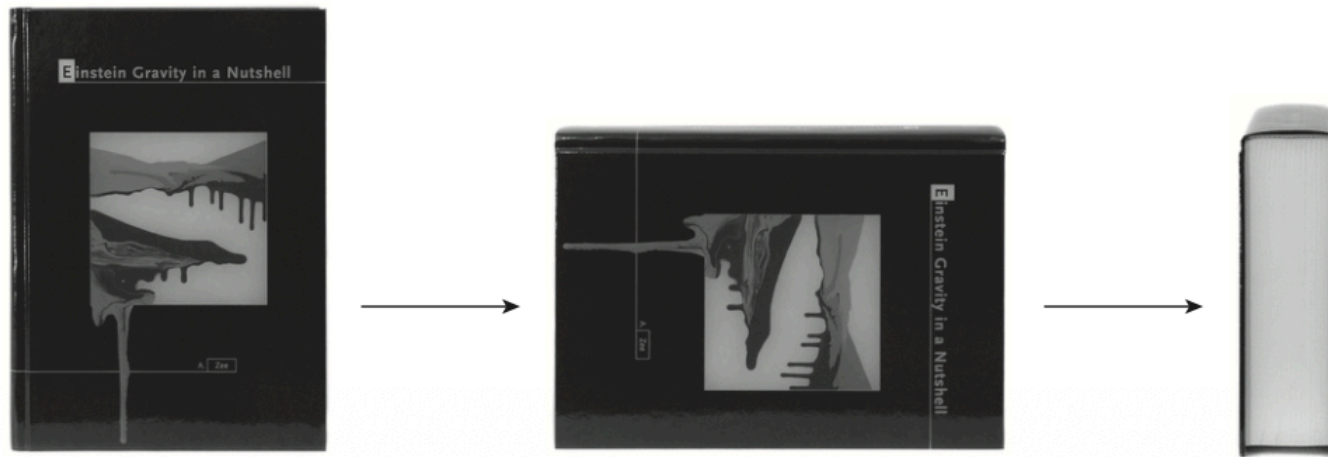
1. Rotations in 3-dimensional Euclidean space, as already mentioned, form the poster child of group theory and are almost indispensable in physics. Think of rotating a rigid object, such as a bust of Newton. After two rotations in succession, the bust, being rigid, has not been deformed in any way: it merely has a different orientation. Thus, the composition of two rotations is another rotation.

Rotations famously do not commute. See figure 2.

Descartes taught us that 3-dimensional Euclidean space could be thought of as a linear vector space, coordinatized with the help of three unit basis vectors $\vec{e}_x = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $\vec{e}_y = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, and $\vec{e}_z = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$, aligned along three orthogonal directions traditionally named x , y , and z . A rotation takes each basis vector into a linear combination of these three basis vectors, and is thus described by a 3-by-3 matrix. This group of rotations is called $SO(3)$. We shall discuss rotations in great detail in chapter I.3; suffice it to mention here that the determinant of a rotation matrix is equal to 1.



(a)



(b)

Figure 2



Figure 3

2. Rotations in 2-dimensional Euclidean space, namely a plane, form a group called $SO(2)$, consisting of the set of rotations around an axis perpendicular to the plane. Denote a rotation through angle ϕ by $R(\phi)$. Then $R(\phi_1)R(\phi_2) = R(\phi_1 + \phi_2) = R(\phi_2)R(\phi_1)$. These rotations commute. (See figure 3. I was surprised to discover that this bust of Dirac at St. John's College, Cambridge University, was not nailed down but could be rotated around the vertical axis. The photo depicts my attempt to give the bust a half-integral amount of angular momentum.)

3. The permutation group S_4 rearranges an ordered set of four objects, which we can name arbitrarily, for example, (A, B, C, D) or $(1, 2, 3, 4)$. An example would be a permutation that takes $1 \rightarrow 3, 2 \rightarrow 4, 3 \rightarrow 2$, and $4 \rightarrow 1$. As is well known, there are $4! = 24$ such permutations (since we have four choices for which number to take 1 into, three choices for which number to take 2 into, and two choices for which number to take 3 into). The permutation group S_n evidently has $n!$ elements. We discuss S_n in detail in chapter I.2.
4. Even permutations of four objects form the group A_4 . As is also well known, a given permutation can be characterized as either even or odd (we discuss this in more detail in chapter I.2). Half of the 24 permutations in S_4 are even, and half are odd. Thus, A_4 has 12 elements. The jargon guy tells us that A stands for “alternating.”
5. The two square roots of 1, $\{1, -1\}$, form the group Z_2 under ordinary multiplication.
6. Similarly, the three cube roots of 1 form the group $Z_3 = \{1, \omega, \omega^2\}$ with $\omega \equiv e^{2\pi i/3}$.

Chugging right along, we note that the four fourth roots of 1 form the group $Z_4 = \{1, i, -1, -i\}$, where famously (or infamously) $i = e^{i\pi/2}$.

More generally, the N N th roots of 1 form the group $Z_N = \{e^{i2\pi j/N} : j = 0, \dots, N-1\}$. The composition of group elements is defined by $e^{i2\pi j/N} e^{i2\pi k/N} = e^{i2\pi(j+k)/N}$.

Quick question: Does the set $\{1, i, -1\}$ form a group?

7. Complex numbers of magnitude 1, namely $e^{i\phi}$, form a group called $U(1)$, with $e^{i\phi_1} e^{i\phi_2} = e^{i(\phi_1+\phi_2)}$. Since $e^{i(\phi+2\pi)} = e^{i\phi}$, we can restrict ϕ to range from 0 to 2π . At the level of physicist rigor, we can think of $U(1)$ as the “continuum limit” of Z_N with $e^{i2\pi j/N} \rightarrow e^{i\phi}$ in the limit $N \rightarrow \infty$ and $j \rightarrow \infty$ with the ratio held fixed $2\pi j/N = \phi$.

8. The addition of integers mod N generates a group. For example, under addition mod 5 the set $\{0, 1, 2, 3, 4\}$ forms a group: $2 + 1 = 3$, $3 + 2 = 0$, $4 + 3 = 2$, and so on. The composition of the group elements is defined by $j \cdot k = j + k \bmod 5$. The identity element I is denoted by 0. The inverse of 2, for example, is 3, of 4 is 1, and so on. The group is clearly abelian.

Question: Have you seen this group before?

9. The addition of real numbers form a group, perhaps surprisingly. The group elements are denoted by a real number u and $u \cdot v \equiv u + v$, where the symbol $+$ is what an elementary school student would call “add.” You can easily check that the axioms are satisfied. The identity element is denoted by 0, and the inverse of the element u is the element $-u$.
10. The additive group of integers is obtained from the additive group of real numbers by restricting u and v in the preceding example to be integers of either sign, including 0.
11. As many readers know, in Einstein’s theory of special relativity,¹¹ the spacetime coordinates used by two observers in relative motion with velocity v along the x -direction (say) are related by the Lorentz transformation (with c the speed of light):

$$\begin{aligned} ct' &= \cosh \varphi \, ct + \sinh \varphi \, x \\ x' &= \sinh \varphi \, ct + \cosh \varphi \, x \\ y' &= y \\ z' &= z \end{aligned} \tag{1}$$

where the “boost angle” φ is determined by $\tanh \varphi = v/c$. (In other words, $\cosh \varphi = 1/\sqrt{1 - \frac{v^2}{c^2}}$, and $\sinh \varphi = \frac{v}{c}/\sqrt{1 - \frac{v^2}{c^2}}$.) Suppressing the y - and z -coordinates, we can describe the Lorentz transformation by

$$\begin{pmatrix} ct' \\ x' \end{pmatrix} = \begin{pmatrix} \cosh \varphi & \sinh \varphi \\ \sinh \varphi & \cosh \varphi \end{pmatrix} \begin{pmatrix} ct \\ x \end{pmatrix} \tag{2}$$

Physically, suppose a third observer is moving at a velocity defined by the boost angle φ_2 relative to the observer moving at a velocity defined by the boost angle φ_1 relative to the first observer. Then we expect the third observer to be moving at some velocity determined by φ_1 and φ_2 relative to the first observer. (All motion is restricted to be along the x -direction for simplicity.) This physical statement is expressed by the mathematical statement that the Lorentz transformations form a group:

$$\begin{pmatrix} \cosh \varphi_2 & \sinh \varphi_2 \\ \sinh \varphi_2 & \cosh \varphi_2 \end{pmatrix} \begin{pmatrix} \cosh \varphi_1 & \sinh \varphi_1 \\ \sinh \varphi_1 & \cosh \varphi_1 \end{pmatrix} = \begin{pmatrix} \cosh(\varphi_1 + \varphi_2) & \sinh(\varphi_1 + \varphi_2) \\ \sinh(\varphi_1 + \varphi_2) & \cosh(\varphi_1 + \varphi_2) \end{pmatrix} \quad (3)$$

The boost angles add.*

2. Consider the set of n -by- n matrices M with determinants equal to 1. They form a group under ordinary matrix multiplication, since as was shown in the review of linear algebra, the determinant of the product two matrices is equal to the product of the determinants of the two matrices: $\det(M_1 M_2) = \det(M_1) \det(M_2)$. Thus, $\det(M_1 M_2) = 1$ if $\det(M_1) = 1$ and $\det(M_2) = 1$: closure is satisfied. Since $\det M = 1 \neq 0$, the inverse M^{-1} exists. The group is known as $SL(n, R)$, the special linear group with real entries. If the entries are allowed to be complex, the group is called $SL(n, C)$. (Matrices with unit determinant are called special.)

From these examples, we see that groups can be classified according to whether they are finite or infinite, discrete or continuous.

Note that a discrete group can well be infinite, such as the additive group of integers.

Concept of subgroup

In group theory, many concepts are so natural that they practically suggest themselves, for example, the notion of a subgroup.

Given a set of entities $\{g_\alpha\}$ that form a group G , if a subset $\{h_\beta\}$ also form a group, call it H , then H is known as a subgroup of G and we write $H \subset G$.

Here are some examples.

1. $SO(2) \subset SO(3)$. This shows that, in the notation $\{g_\alpha\}$ and $\{h_\beta\}$ we just used, the index sets denoted by α and β can in general be quite different; here α consists of three angles and β of one angle.
2. $S_m \subset S_n$ for $m < n$. Permuting three objects is just like permuting five objects but keeping two of the five objects untouched. Thus, $S_3 \subset S_5$.
3. $A_n \subset S_n$.
4. $Z_2 \subset Z_4$, but $Z_2 \not\subset Z_5$.
5. $SO(3) \subset SL(3, R)$.

Cyclic subgroups

the

the

the

the

the

For a finite group G , pick some element g and keep multiplying it by itself.

In other words, consider the sequence $\{g, g^2 = gg, g^3 = g^2g, \dots\}$.

As long as the resulting product is not equal to the identity, we can keep going.

Since G is finite, the sequence must end at some point with $g^k = I$.

The set of elements $\{I, g, g^2, \dots, g^{k-1}\}$ forms a subgroup Z_k .

Thus, any finite group has a bunch of cyclic subgroups.

If k is equal to the number of elements in G , then the group G is in fact Z_k .

Lagrange's theorem

Lagrange proved the following theorem.

Let a group G with n elements have a subgroup H with m elements.

Then m is a factor of n .

In other words, n/m is an integer.

The proof is as follows.

List the elements of $H : \{h_1, h_2, \dots, h_m\}$.

Note: Since H forms a group, this list must contain I .

We do not list any element more than once; thus, $h_a \neq h_b$ for $a \neq b$.

Let $g_1 \in G$ but $\notin H$ (in other words, g_1 is an element of G outside H).

Consider the list $\{h_1g_1, h_2g_1, \dots, h_mg_1\}$, which we denote by $\{h_1, h_2, \dots, h_m\}g_1$ to save writing.

Note that this set of elements does not form a group.

Can you explain why not?

I claim that the elements on the list $\{h_1g_1, h_2g_1, \dots, h_mg_1\}$ are all different from one another.

Proof by contradiction: For $a \neq b$, $h_ag_1 = h_bg_1 \Rightarrow h_a = h_b$ upon multiplication from the right by $(g_1)^{-1}$ (which exists, since G is a group).

I also claim that none of the elements on this list are on the list $\{h_1, h_2, \dots, h_m\}$.

Proof: For some a and b , $h_ag_1 = hb \Rightarrow g_1 = h^{-1}_a h_b$, which contradicts the assumption that g_1 is not in H .

Note that H being a group is crucial here.

Next, pick an element g_2 of G not in the two previous lists, and form $\{h_1g_2, h_2g_2, \dots, h_mg_2\} = \{h_1, h_2, \dots, h_m\} g_2$.

I claim that these m elements are all distinct.

Again, this proof follows by contradiction, as follows:

Answer: For $a \neq b$, $h_ag_2 = h_bg_2 \Rightarrow h_a = h_b$.

I also claim that none of these elements are on the two previous lists.

Yes, the proof proceeds again easily by contradiction.

For example, $h_ag_2 = h_bg_1 \Rightarrow g_2 = h^{-1}_ah_bg_1 = h_cg_1$, since H is a group, but this would mean that g_2 is on the list $\{h_1, h_2, \dots, h_m\}g_1$, which is a contradiction.

We repeat this process.

After each step, we ask whether there is any element of G left that is not on the lists already constructed.

If yes, then we repeat the process and construct yet another list containing m distinct elements.

Eventually, there is no group element left (since G is a finite group).

We have constructed k lists, including the original list $\{h_1, h_2, \dots, h_m\}$, namely, $\{h_1, h_2, \dots, h_m\}g_j$ for $j = 0, 1, 2, \dots, k-1$ (writing I as g_0).

Therefore $n = mk$, that is, m is a factor of n . QED.

As a simple example of Lagrange's theorem, we can immediately state that Z_3 is a subgroup of Z_{12} but not of Z_{14} .

It also follows trivially that if p is prime, then Z_p does not have a nontrivial subgroup.

From this you can already sense the intimate relation between group theory and number theory.

Direct product of groups

Given two groups F and G (which can be continuous or discrete), whose elements we denote by f and g , respectively, we can define another group $H \equiv F \otimes G$, known as the direct product of F and G , consisting of the elements (f, g) .

If you like, you can think of the symbol (f, g) as some letter in a strange alphabet. T

he product of two elements (f, g) and (f', g') of H is given by $(f, g)(f', g') = (ff', gg')$.

The identity element of H is evidently given by (I, I) , since $(I, I)(f, g) = (If, Ig) = (f, g)$ and $(f, g)(I, I) = (fI, gI) = (f, g)$.

If we were insufferable pedants, we would write $I_H = (I_F, I_G)$, since the identity elements I_H, I_F, I_G of the three groups H, F, G are conceptually quite distinct.

What is the inverse of (f, g) ?

If F and G have m and n elements, respectively, how many elements does $F \otimes G$ have?

Evidently, the inverse of (f, g) is (f^{-1}, g^{-1}) , and $F \otimes G$ has mn elements.

Klein's Vierergruppe V

A simple example is given by $Z_2 \otimes Z_2$, consisting of the four elements: $I = (1, 1)$, $A = (-1, 1)$, $B = (1, -1)$, and $C = (-1, -1)$.

For example, we have $AB = (-1, -1) = C$.

Note that this group is to be distinguished from the group Z_4 consisting of the four elements $1, i, -1, -i$.

The square of any element in $Z_2 \otimes Z_2$ is equal to the identity, but this is not true of Z_4 . In particular, $i^2 = -1 \neq 1$.

Incidentally, $Z_2 \otimes Z_2$, also known as Klein's Vierergruppe ("4-group" in German) and denoted by V , played an important historical role in Klein's program.

Note that the elements of F , regarded as a subgroup of $F \otimes G$, are written as (f, I) .

Similarly, the elements of G are written as (I, g) .

Clearly, (f, I) and (I, g) commute.

The direct product would seem to be a rather “cheap” way of constructing larger groups out of smaller ones, but Nature appears to make use of this possibility.

The theory of the strong, weak, and electromagnetic interaction is based on the group $SU(3) \otimes SU(2) \otimes U(1)$.

A teeny bit of history: “A pleasant human flavor”

Historians of mathematics have debated about who deserves the coveted title of “the founder of group theory.”

Worthy contenders include Cauchy, Lagrange, Abel, Ruffini, and Galois.

Lagrange was certainly responsible for some of the early concepts, but the sentimental favorite has got to be Evariste Galois, what with the ultra romantic story of him feverishly writing down his mathematical ideas the night before a fatal duel at the tender age of 20.

Whether the duel was provoked because of the honor of a young woman named du Motel or because of Galois’s political beliefs (for which he had been jailed) is apparently still not settled.

In any case, he was the first to use the word “group.”

Nice choice.

To quote the mathematician G. A. Miller, it is silly to argue about who founded group theory anyway:

We are inclined to attribute the honor of starting a given big theory to an individual just as we are prone to ascribe fundamental theorems to particular men, who frequently have added only a small element to the development of the theorem. Hence the statement that a given individual founded a big theory should not generally be taken very seriously. It adds, however, a pleasant human flavor and awakens in us a noble sense of admiration and appreciation. It is also of value in giving a historical setting and brings into play a sense of the dynamic forces which have contributed to its development instead of presenting to us a cold static scene. Observations become more inspiring when they are permeated with a sense of development.¹⁵

While symmetry considerations have always been relevant for physics, group theory did not become indispensable for physics until the advent of quantum mechanics, for reasons to be explained in later.

Eugene Wigner, who received the Nobel Prize in 1963 largely for his use of group theory in physics, recalled the tremendous opposition to group theory among the older generation (including Einstein, who was 50 at the time) when he first started using it around 1929 or so.

Schrodinger told him that while group theory provided a nice derivation of some results in atomic spectroscopy, “surely no one will still be doing it this way in five years.”

Well, a far better theoretical physicist than a prophet!

But Wigner’s childhood friend John von Neumann, who helped him with group theory, reassured him, saying “Oh, these are old fogeys.

In five years, every student will learn group theory as a matter of course.”

Pauli coined the term “die Gruppenpest” (“that pesty group business”), which probably captured the mood at the time.

Remember that quantum mechanics was still freshly weird, and all this math might be too much for older people to absorb.

Multiplication table: The “once and only once rule”

A finite group with n elements can be characterized by its multiplication table, as shown here.

	$\dots$	g_j	$\dots$
$\vdots$	$\cdot$	$\cdot$	$\cdot$
g_i	$g_i g_j$		
$\vdots$			$\cdot$
$\vdots$			$\cdot$

We construct a square n -by- n table, writing the product $g_i g_j$ in the square in the i th row and the j th column:

A simple observation is that, because of the group properties, in each row any group element can appear once and only once.

To see this, suppose that in the i th row, the same group element appears twice, that is, $g_i g_j = g_i g_k$ for $j \neq k$.

Then multiplying by g_i^{-1} from the left, we obtain $g_j = g_k$, contrary to what was assumed.

It follows that each of the n elements must appear once to fill up the n slots in that row.

We might refer to this as the “once and only once rule.”

The same argument could be repeated with the word “row” replaced by “column,” of course.

For n small, all possible multiplication tables and hence all possible finite groups with n elements can readily be constructed.

Let us illustrate this for $n = 4$.

For pedagogical reasons, we will do this in two different ways, one laborious, the other “slick.”

Finite groups with four elements: The slow way

First, we proceed very slowly, by brute force.

Call the four elements I , A , B , and C .

- 1. By definition of the identity, the first row and first column can be filled in automatically:

	I	A	B	C
I	I	A	B	C
A	A			
B	B			
C	C			

2. We are to fill in the second row with I , B , and C . The first entry in that row is A^2 . There are two possible choices: choice (a): $A^2 = B$, or choice (b): $A^2 = I$. (You might think that there is a third choice, $A^2 = C$, but that is the same as choice (a) upon renaming the elements. What you call C I will call B .)

Let us now follow choice (a) and come back to choice (b) later.

3. The multiplication table now reads

	I	A	B	C
I	I	A	B	C
A	A	B	2	3
B	B	4	5	6
C	C			

where for your and my convenience I have numbered some of the boxes yet to be filled in.

4. We have to put C and I into boxes 2 and 3. But we cannot put C into box 3, since otherwise the fourth column will break the “once and only once rule”: C would appear twice:

	I	A	B	C
I	I	A	B	C
A	A	B	C	I
B	B	4	5	6
C	C			

5. Again by the “once and only once rule,” box 4 can only be C or I . The latter choice would mean $BA = I$ and hence $B = A^{-1}$, but we already know from the second row of the multiplication table that $AB = C \neq I$. Thus, box 4 can only be C . Hence box 5 is I , and 6 is A .
6. Finally, the last three blank entries in the fourth column are fixed uniquely by the “once and only once rule.” We obtain²²

	I	A	B	C
I	I	A	B	C
A	A	B	C	I
B	B	C	I	A
C	C	I	A	B

Now that we have the multiplication table, we know everything about the group, and we can ask: What group is this?

From the second row, we read off $A^2 = B$, $A^3 = AA^2 = AB = C$, $A^4 = AA^3 = AC = I$.

The group is Z_4 .

Interestingly, we don’t even have to finish constructing the entire multiplication table.

In this simple case, by the time we had filled in the second row, we could have quit.

The rest of the table, however, provides us with a lot of consistency checks to ensure that we have not messed up.

For example, from the last row, we have $CB = A$.

But we know from the second row that $B = A^2$ and $C = A^3$, and hence the statement $CB = A$ says that $A^3A^2 = A^5 = A$, showing that indeed $A^4 = I$.

We now go back to choice (b): $A^2 = I$, so that

	I	A	B	C
I	I	A	B	C
A	A	I	2	3
B	B	4	5	6
C	C	7	8	9

1. We are to fill boxes 2 and 3 with C and B . By the “once and only once rule” in the third and fourth columns, these boxes can only be C and B in that order.

2. By the same reasoning, we can only fill boxes 4 and 7 with C and B , respectively. We thus obtain

	I	A	B	C
I	I	A	B	C
A	A	I	C	B
B	B	C		
C	C	B		

3. Now it looks like we could fill in the four remaining empty boxes with either

I	A
A	I

or

A	I
I	A

But the two choices amount to the same thing. We simply rename B and C . Thus, we obtain

	I	A	B	C
I	I	A	B	C
A	A	I	C	B
B	B	C	I	A
C	C	B	A	I

Again, what group is this?

It is just $\mathbb{Z}_2 \otimes \mathbb{Z}_2$: $A^2 = I$, $B^2 = I$, $C = AB = BA$ (and hence also $C^2 = I$).

A quick way: Construct the cyclic subgroups

Here is an alternative to this laborious procedure of constructing the multiplication table step by step.

We use the earlier observation that in a finite group, if we keep multiplying an element by itself, we will reach the identity I .

Given a group G of four elements $\{I, A, B, C\}$, we keep multiplying A by itself.

If $A^4 = I$, then $G = Z_4$.

By Lagrange's theorem, the possibility $A^3 = I$ is not allowed.

If $A^2 = I$, then we multiply B by itself.

Either B^2 or B^4 equals I .

The latter is ruled out, so the only possibility is that $B^2 = I$, and $AB = BA = C$.

Then $G = Z_2 \otimes Z_2$, with the four elements represented by $(1, 1)$, $(1, -1)$, $(-1, 1)$, and $(-1, -1)$.

If you are energetic and driven, you could try to construct all possible finite groups with n elements, and see how large an n you could get to.

A quick hint: It's easy if n is prime.

Presentations

For large groups, writing down the multiplication table is clearly a losing proposition.

Instead, finite groups are defined by their properties, as in the examples listed above, or by their presentations, which list the elements (sometimes called generators) from which all other elements can be obtained by group multiplication, and the essential relations the generators satisfy.

Thus, in a self-evident notation, the groups Z_4 and $Z_2 \otimes Z_2$ are defined by their presentations as follows:

$$Z_4 : \langle A | A^4 = I \rangle \tag{4}$$

$$Z_2 \otimes Z_2 : \langle A, B | A^2 = B^2 = I, AB = BA \rangle \tag{5}$$

The two groups are clearly distinct.

In particular, Z_4 contains only one element that squares to I , namely A^2 .

Homomorphism and isomorphism

A map $f: G \rightarrow G'$ of a group G into the group G' is called a homomorphism if it preserves the multiplicative structure of G , that is, if $f(g_1)f(g_2) = f(g_1g_2)$.

Clearly, this requirement implies that $f(I) = I$ (more strictly speaking, the identity of G is mapped to the identity of G').

A homomorphism becomes an isomorphism if the map is one-to-one and onto.

Now we can answer the question posed earlier: the additive group of integers mod N is in fact isomorphic to Z_N .

For a more interesting example, consider $Z_2 \otimes Z_4$.

We use the additive notation here and thus write the elements as (n, m) and compose them according to $(n, m) \cdot (n', m') = (n + n' \bmod 2, m + m' \bmod 4)$.

We start with $(0, 0)$ and add $(1, 1)$ repeatedly: $(0, 0) + (1, 1) \xrightarrow{+(1,1)} (1, 1) \rightarrow (0, 2) \rightarrow (1, 3) \rightarrow (0, 4) = (0, 0)$; we get back to where we started.

Next, we start with $(0, 1)$ and again add $(1, 1)$ repeatedly: $(0, 1) + (1, 1) \xrightarrow{+(1,1)} (1, 2) \rightarrow (0, 3) \rightarrow (1, 0) \rightarrow (0, 1)$, getting back to where we started.

Thus we can depict $Z_2 \otimes Z_4$ by a rectangular 2-by-4 discrete lattice on a torus (see figure 4).

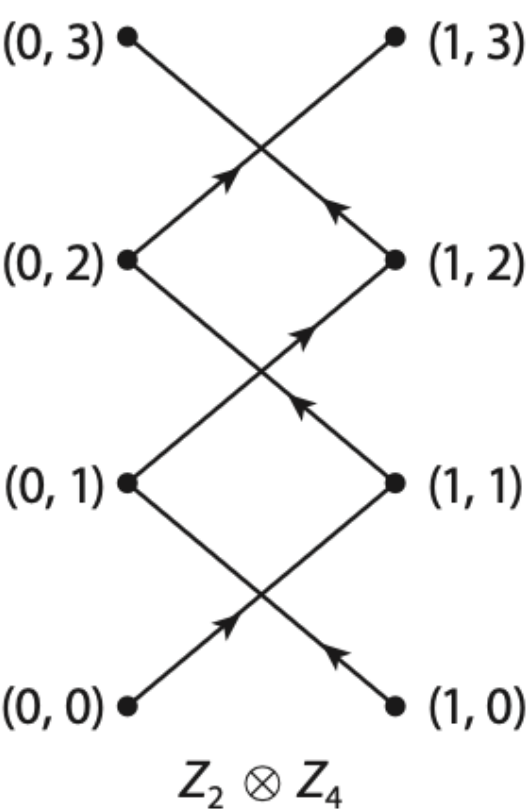


Figure 4

Now we come in for a bit of a surprise.

Consider $Z_2 \otimes Z_3$ consisting of (n, m) , which we compose by $(n + n' \bmod 2, m + m' \bmod 3)$.

Again, we start with $(0, 0)$ and add $(1, 1)$ repeatedly: $(0, 0) \rightarrow (1, 1) \xrightarrow{+(1,1)} (2, 2) = (0, 2) \rightarrow (1, 3) = (1, 0) \rightarrow (2, 1) = (0, 1) \rightarrow (2, 3) = (0, 0)$.

We are back where we started!

In the process, we cycled through all six elements of $Z_2 \otimes Z_3$.

We conclude that the six elements $(0, 0), (1, 1), (0, 2), (1, 0), (0, 1)$, and $(1, 2)$ describe Z_6 .

Thus, $Z_2 \otimes Z_3$ and Z_6 are isomorphic; they are literally the same group.

Note that this phenomenon, of a possible isomorphism between $Z_p \otimes Z_q$ and Z_{pq} , does not require p and q to be prime, only relatively prime.

As another example of isomorphism, the groups $SO(2)$ and $U(1)$ introduced earlier in the chapter are isomorphic.

The map $f: SO(2) \rightarrow U(1)$ is defined simply by $f(R(\phi)) = e^{i\phi}$

Rotations and the Notion of Lie Algebra

Two different approaches to rotations in the plane

If the permutation groups are the poster children for raising awareness of finite groups, then the rotation groups are surely the poster children for continuous groups.

Given that we were all born into 3-dimensional Euclidean space, it is hardly surprising that the rotation group plays a crucial role in physics.

Indeed, the very concept of a group was abstracted from the behavior of rotations.

My pedagogical strategy in this chapter is to first take something you know extremely well, namely, rotations in the plane, present it in a way possibly unfamiliar to you, and go through it slowly in great detail—“beating it to death,” so to speak.

And then we will move up to rotations in 3-dimensional space and beyond.

I start with two different approaches to rotations in the plane, the first based on trigonometry, the second based on invariance considerations.

Cartesian coordinates and trigonometry

As you know, from mechanics, we envisage either rotating the body we are studying or rotating the observer.

Here we consistently rotate the observer.

Put down Cartesian coordinate axes (see figure 1) so that a point P is labeled by two real numbers (x, y) .

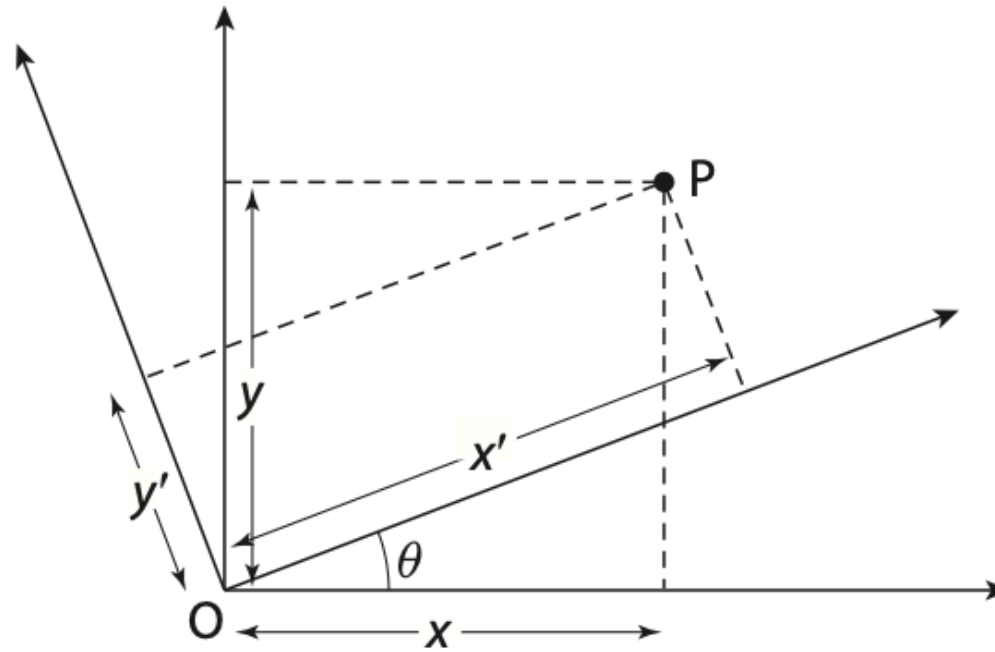


Figure 1

Suppose another observer (call him Mr. Prime) puts down coordinate axes rotated by angle θ with respect to the axes put down by the first observer (call her Ms. Unprime) but sharing the same origin O.

Elementary trigonometry tells us that the coordinates (x, y) and (x', y') assigned by the two observers to the same point P are related by

$$x' = \cos \theta x + \sin \theta y, \quad y' = -\sin \theta x + \cos \theta y \quad (1)$$

The distance from P to the origin O has to be the same for the two observers, of course.

According to Pythagoras, this requires $\sqrt{x'^2 + y'^2} = \sqrt{x^2 + y^2}$, which you could check using (1).

Introduce the column vectors $\vec{r} = \begin{pmatrix} x \\ y \end{pmatrix}$ and $\vec{r}' = \begin{pmatrix} x' \\ y' \end{pmatrix}$ and the rotation matrix

$$R(\theta) = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \quad (2)$$

Then we can write (1) more compactly as $\vec{r}' = R(\theta)\vec{r}$.

We have already used the word “vector” in linear algebra.

For the purpose of this chapter, a vector is a quantity (for example, the velocity of a particle in the plane) consisting of two real numbers, such that if Ms. Unprime represents it by $\vec{p} = \begin{pmatrix} p^1 \\ p^2 \end{pmatrix}$, then Mr. Prime will represent it by $\vec{p}' = R(\theta)\vec{p}$.

In short, a vector is something that transforms like the coordinates $\begin{pmatrix} x \\ y \end{pmatrix}$ under rotation.

The emphasis here is on the word “like”.

Invariance under linear transformations

Given two vectors $\vec{p} = \begin{pmatrix} p^1 \\ p^2 \end{pmatrix}$ and $\vec{q} = \begin{pmatrix} q^1 \\ q^2 \end{pmatrix}$, the scalar or dot product is defined by $\vec{p}^T \cdot \vec{q} = p_1 q_1 + p_2 q_2$.

Here the transposed vector $\vec{p}^T$ is the row vector (p_1, p_2) .

According to Pythagoras, the square of the length of $\vec{p}$ is given by $\vec{p}^2 \equiv \vec{p}^T \cdot \vec{p} = (p^1)^2 + (p^2)^2$. By definition, rotations leave the length of $\vec{p}$, and hence $\vec{p}^2$, invariant. In other words, if $\vec{p}' = R(\theta)\vec{p}$, then $\vec{p}'^2 = \vec{p}^2$. Since this works for any vector $\vec{p}$, including the case in which $\vec{p}$ happens to be composed of two arbitrary vectors $\vec{u}$ and $\vec{v}$ (namely, $\vec{p} = \vec{u} + \lambda\vec{v}$ with λ an arbitrary real number), and since $\vec{p}^2 = (\vec{u} + \lambda\vec{v})^2 = \vec{u}^2 + \lambda^2\vec{v}^2 + 2\lambda\vec{u}^T \cdot \vec{v}$, rotation also leaves the dot product between two arbitrary vectors invariant: $\vec{p}'^2 = \vec{p}^2$ (where $\vec{p}' = \vec{u}' + \lambda\vec{v}' = R\vec{p} = R\vec{u} + \lambda R\vec{v}$) implies that $\vec{u}'^T \cdot \vec{v}' = \vec{u}^T \cdot \vec{v}$, since λ can be varied arbitrarily.

Since $\vec{u}' = R\vec{u}$ (to unclutter things, we often suppress the θ dependence in $R(\theta)$) and so $\vec{u}'^T = \vec{u}^T R^T$, we now have $\vec{u}^T \cdot \vec{v} = \vec{u}'^T \cdot \vec{v}' = (\vec{u}^T R^T) \cdot (R\vec{v}) = \vec{u}^T (R^T R)\vec{v}$. As this holds for any two vectors $\vec{u}$ and $\vec{v}$, we must have the matrix equation

$$R^T R = I \tag{3}$$

This imposes a condition on R .

Matrices that satisfy (3) are called orthogonal.

As was already mentioned earlier, these orthogonal matrices generate the group $O(2)$, with the multiplication law $R(\theta_1)R(\theta_2) = R(\theta_1 + \theta_2)$.

As promised, we followed two approaches to studying rotations.

In the first approach, we used trigonometry, in the second, we insisted on linear transformations that leave the lengths of vectors unchanged.

At this point, we might pause to check that the two approaches are consistent by plugging in the explicit form of the rotation, given in (2), into the condition (3):

$$R(\theta)^T R(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (4)$$

Reflections

Recall from the linear algebra that the determinant of a product of matrices is equal to the product of the determinants: $\det M_1 M_2 = (\det M_1)(\det M_2)$, and that the determinant of the transpose of a matrix is the same as the determinant of that matrix: $\det M^T = \det M$.

Taking the determinant of (3), we obtain $(\det R)^2 = 1$, that is, $\det R = \pm 1$.

The determinant of an orthogonal matrix may be -1 as well as $+1$.

In other words, orthogonal matrices also include reflection matrices, such as $\mathcal{P} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, a reflection flipping the y-axis.

To focus on rotations, let us exclude reflections by imposing the condition (since $\det \mathcal{P} = -1$)

$$\det R = 1 \quad (5)$$

Matrices with unit determinant are called special.

Note that matrices of the form $\mathcal{P} R$ for any rotation R are also excluded by (5), since $\det(\mathcal{P} R) = \det \mathcal{P} \det R = (-1)(+1) = -1$.

In particular, a reflection flipping the x-axis $\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$, which is the product of $\mathcal{P}$ and a rotation through 90° , is also excluded.

We define a rotation as a matrix that is both orthogonal and special, that is, a matrix that satisfies both (3) and (5).

Thus, the rotation group of the plane consists of the set of all special orthogonal 2-by-2 matrices and is known as $SO(2)$.

Now, it turns out that it is good practice in physics and mathematics to turn equations and logical sequences around. Starting from (2), you can readily check (3) and (5).

Verify that, given (3) and (5), you can also get to (2).

Indeed, in the next section, we will learn how to do this in arbitrary dimensions, not just for the almost trivial 2-dimensional case.

Act a little bit at a time

The Norwegian physicist Marius Sophus Lie had the almost childishly obvious but brilliant idea that to rotate through, say, 29° , you could just as well rotate through a zillionth of a degree and repeat the process 29 zillion times.

To study rotations, it suffices to study rotation through infinitesimal angles.

Shades of Newton and Leibniz!

A rotation through a finite angle can always be obtained by performing infinitesimal rotations repeatedly.

As is typical with many profound statements in physics and mathematics, Lie's idea is astonishingly simple.

Replace the proverb “Never put off until tomorrow what you have to do today” by “Do what you have to do a little bit at a time.”

A rotation through an infinitesimal angle θ is almost the identity I , that is, no rotation at all, and so can be written as

$$R(\theta) \simeq I + A \tag{6}$$

Here A denotes some infinitesimal matrix of order θ .

The neglected terms in (6) are of order θ^2 and smaller.

Let us imagine Lie saying to himself, “Pretend that I slept through trigonometry class and I don't know anything about (2).

Instead, I will define rotations as the set of linear transformations on 2-component objects $\vec{u}' = R\vec{u}$ and $\vec{v}' = R\vec{v}$ that leave $\vec{u}^T \cdot \vec{v}$ invariant.

I will impose (3) $R^T R = I$ and derive (2).

But according to my brilliant idea, it suffices to solve this condition for rotations infinitesimally close to the identity.”

Following Lie, we plug $R \simeq I + A$ into $R^T R = I$.

Since by assumption A^2 , being of order θ^2 , can be neglected relative to A , we have

$$R^T R \simeq (I + A^T)(I + A) \simeq (I + A^T + A) = I \quad (7)$$

Thus, this requires $A^T = -A$, namely, that A must be antisymmetric.

But there is basically only one 2 x 2 antisymmetric matrix:

$$\mathcal{J} \equiv \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad (8)$$

In other words, the solution of $A^T = -A$ is $A = \theta \mathcal{J}$ for some real number θ , which, as we will discover shortly, is in fact the same as the angle θ we have been using.

Thus, rotations close to the identity have the form†

$$R = I + \theta \mathcal{J} + O(\theta^2) = \begin{pmatrix} 1 & \theta \\ -\theta & 1 \end{pmatrix} + O(\theta^2) \quad (9)$$

The antisymmetric matrix $\mathcal{J}$ is known as the generator of the rotation group.

We obtain, without knowing any trigonometry, that under an infinitesimal rotation, $x \rightarrow x' \simeq x + \theta y$, and $y \rightarrow y' = -\theta x + y$, which is of course consistent with (1).

We could also obtain this result by drawing an elementary geometrical figure involving infinitesimal angles.

Now recall the identity $e^x = \lim_{N \rightarrow \infty} (1 + x/N)^N$ (which you can easily prove by differentiating both sides).

Then, for a finite (that is, not infinitesimal) angle θ , we have

$$R(\theta) = \lim_{N \rightarrow \infty} \left(R\left(\frac{\theta}{N}\right) \right)^N = \lim_{N \rightarrow \infty} \left(1 + \frac{\theta \mathcal{J}}{N} \right)^N = e^{\theta \mathcal{J}} \quad (10)$$

The first equality represents Lie's profound idea: we cut up the given noninfinitesimal angle θ into N pieces so that θ/N is infinitesimal for N large enough and perform the infinitesimal rotation N times.

The second equality is just (9).

For the last equality, we use the identity just mentioned, which amounts to the definition of the exponential.

As an alternative but of course equivalent path to our result, simply assert that we have every right, to leading order, to write $R(\frac{\theta}{N}) = 1 + \frac{\theta \mathcal{J}}{N} \simeq e^{\frac{\theta \mathcal{J}}{N}}$.

Thus,

$$R(\theta) = \lim_{N \rightarrow \infty} \left(R\left(\frac{\theta}{N}\right) \right)^N = \lim_{N \rightarrow \infty} \left(e^{\frac{\theta \mathcal{J}}{N}} \right)^N = e^{\theta \mathcal{J}} \quad (11)$$

In calculus, we learned about the Taylor or power series.

Taylor said that if we gave him all the derivatives of a function $f(x)$ at $x = 0$ (say), he could construct the function.

In contrast, Lie said that, thanks to the multiplicative group structure, he only needs the first derivative of the group element $R(\theta)$ near the identity.

Indeed, we recognize that $\mathcal{J}$ is just $dR(\theta)/d\theta|_{\theta=0}$.

The reason that Lie needs so much less is of course that the group structure is highly restrictive.

Finally, we can check that the formula $R(\theta) = e^{\theta \mathcal{J}}$ reproduces (2) for any value of θ .

We simply note that $\mathcal{J}^2 = -I$ and separate the exponential series, using Taylor's idea, into even and odd powers of $\mathcal{J}$:

$$\begin{aligned}
e^{\theta \mathcal{J}} &= \sum_{n=0}^{\infty} \theta^n \mathcal{J}^n / n! = \left(\sum_{k=0}^{\infty} (-1)^k \theta^{2k} / (2k)! \right) I + \left(\sum_{k=0}^{\infty} (-1)^k \theta^{2k+1} / (2k+1)! \right) \mathcal{J} \\
&= \cos \theta \, I + \sin \theta \, \mathcal{J} = \cos \theta \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \sin \theta \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}
\end{aligned} \tag{12}$$

which is precisely $R(\theta)$ as given in (2).

Note that this works, because $\mathcal{J}$ plays the same role as i in Euler's identity $e^{i\theta} = \cos \theta + i \sin \theta$.

To summarize, the condition $R^T R = I$ determines the rotation matrix obtained previously by using trigonometry.

An old friend of the author's, Confusio, wanders by.

He looks mildly puzzled.

“How come we don't have to impose $\det R = 1$?”

Ah, that's because we are looking at R s that are continuously related to the identity I .

The reflection $\mathcal{P}$, with $\det \mathcal{P} = -1$, is manifestly not continuously related to the identity, with $\det I = +1$.

Two approaches to rotation

To summarize, there are two different approaches to rotation.

In the first approach, applying trigonometry to figure 1, we write down (1) and hence (2).

In the second approach, we specify what is to be left invariant by rotations and hence define rotations by the condition (3) that rotations must satisfy.

Lie then tells us that it suffices to solve (3) for infinitesimal rotations.

We could then build up rotations through finite angles by multiplying infinitesimal rotations together, thus arriving also at (2).

It might seem that the first approach is much more direct.

One writes down (2) and that's that.

The second approach appears more roundabout.

The point is that the second approach generalizes to higher dimensional spaces (and to other situations, for example see later on the unitary groups) much more readily than the first approach does, as we shall see presently.

Distance squared between neighboring points

Before we go on, let us take care of one technical detail. We assumed that Mr. Prime and Ms. Unprime set up their coordinate systems to share the same origin O. We now show that this condition is unnecessary if we consider two points P and Q (rather than one point, as in our discussion above) and study how the vector connecting P to Q transforms. Let Ms. Unprime assign the coordinates $\vec{r}_P = (x, y)$ and $\vec{r}_Q = (\tilde{x}, \tilde{y})$ to P and Q, respectively. Then Mr. Prime's coordinates $\vec{r}'_P = (x', y')$ for P and $\vec{r}'_Q = (\tilde{x}', \tilde{y}')$ for Q are then given by $\vec{r}'_P = R(\theta)\vec{r}_P$ and $\vec{r}'_Q = R(\theta)\vec{r}_Q$. Subtracting the first equation from the second, we have $(\vec{r}'_P - \vec{r}'_Q) = R(\theta)(\vec{r}_P - \vec{r}_Q)$. Defining $\Delta x = \tilde{x} - x$, $\Delta y = \tilde{y} - y$, and the corresponding primed quantities, we obtain $\begin{pmatrix} \Delta x' \\ \Delta y' \end{pmatrix} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix}$. Rotations leave the distance between the points P and Q unchanged: $(\Delta x')^2 + (\Delta y')^2 = (\Delta x)^2 + (\Delta y)^2$. You recognize of course that this is a lot of tedious verbiage stating the perfectly obvious, but I want to be precise here. Of course, the distance between any two points is left unchanged by rotations. (This also means that the distance between P and the origin is left unchanged by rotations; ditto for the distance between Q and the origin.)

Let us take the two points P and Q to be infinitesimally close to each other and replace the differences $\Delta x', \Delta x$, and so forth by differentials dx', dx , and so forth.

Indeed, 2-dimensional Euclidean space is defined by the distance squared between two nearby points: $ds^2 = dx^2 + dy^2$.

Rotations are defined as linear transformations $(x, y) \rightarrow (x', y')$ such that

$$dx^2 + dy^2 = dx'^2 + dy'^2 \quad (13)$$

The whole point is that this now makes no reference to the origin O (or to whether Mr. Prime and Ms. Unprime even share the same origin).

The column $d\vec{x} = \begin{pmatrix} dx^1 \\ dx^2 \end{pmatrix} \equiv \begin{pmatrix} dx \\ dy \end{pmatrix}$ is defined as the basic or ur-vector, the template for all other vectors.

To repeat, a vector is defined as something that transforms like $d\vec{x}$ under rotations.

From the plane to higher-dimensional space

The reader who has wrestled with Euler angles in a mechanics course knows that the analog of (2) for 3-dimensional space is already quite a mess. In contrast, Lie's approach allows us, as mentioned above, to immediately jump to N -dimensional Euclidean space, defined by specifying the distance squared between two nearby points as given by the obvious generalization of Pythagoras's theorem: $ds^2 = \sum_{i=1}^N (dx^i)^2 = (dx^1)^2 + (dx^2)^2 + \dots + (dx^N)^2$. Rotations are defined as linear transformations $d\vec{x}' = R d\vec{x}$ (with R an N -by- N matrix) that leave ds^2 unchanged.

The preceding discussion allows us to write this condition as $R^T R = I$. As before, we want to eliminate reflection and to focus on rotations by imposing the additional condition $\det R = 1$. The set of N -by- N matrices R that satisfy these two conditions forms the simple orthogonal group $SO(N)$, which is just a fancy way of saying the rotation group in N -dimensional space.

Earlier, we already talked about the rotation group as if it is self-evident that rotations form a group. Let us make it official by checking that $SO(N)$ satisfies all the group axioms. The product of two rotations is a rotation: $(R_1 R_2)^T (R_1 R_2) = (R_2^T R_1^T)(R_1 R_2) = R_2^T (R_1^T R_1) R_2 = R_2^T R_2 = I$ (is that slow enough?), and $\det(R_1 R_2) = \det R_1 \det R_2 = 1$. Matrix multiplication is associative. The condition $\det R = 1$ guarantees the existence of the inverse.

Lie in higher dimensions

The power of Lie now shines through when we want to work out rotations in higher-dimensional spaces.

All we have to do is satisfy the two conditions $R^T R = I$ and $\det R = 1$.

Lie shows us that the first condition, $R^T R = I$, is solved immediately by writing $R \simeq I + A$ and requiring $A = -A^T$, namely, that A be antisymmetric.

That's it.

We could be in a zillion-dimensional space, but still, the rotation group is fixed by requiring A to be antisymmetric.

But it is very easy to write down all possible antisymmetric $N \times N$ matrices!

For $N=2$, there is only one, namely, the $\mathcal{J}$ introduced earlier.

For $N=3$, there are basically three of them:

$$\mathcal{J}_x = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}, \quad \mathcal{J}_y = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \mathcal{J}_z = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (14)$$

Any 3-by-3 antisymmetric matrix can be written as $A = \theta_x J_x + \theta_y J_y + \theta_z J_z$, with three real numbers θ_x , θ_y , and θ_z .

The three 3-by-3 antisymmetric matrices $\mathcal{J}_x$, $\mathcal{J}_y$, $\mathcal{J}_z$ are known as generators.

They generate rotations, but are of course not to be confused with rotations, which are by definition 3 x 3 orthogonal matrices with determinant equal to 1.

One upshot of this whole discussion is that any 3-dimensional rotation (not necessarily infinitesimal) can be written as

$$R(\theta) = e^{\theta_x \mathcal{J}_x + \theta_y \mathcal{J}_y + \theta_z \mathcal{J}_z} = e^{\sum_i \theta_i \mathcal{J}_i} \quad (15)$$

(with $i = x, y, z$) and is thus characterized by three real numbers θ_x , θ_y , and θ_z .

As I said, those readers who have suffered through the rotation of a rigid body in a course on mechanics surely would appreciate the simplicity of studying the generators of infinitesimal rotations and then simply exponentiating them.

To mathematicians, physicists often appear to use weird notations.

There is not an i in sight, yet physicists are going to stick one in now.

If you have studied quantum mechanics, you know that the generators $\mathcal{J}$ of rotation studied here are related to angular momentum operators.

You would also know that in quantum mechanics observables are represented by hermitean operators or matrices.

In contrast, in our discussion, the $\mathcal{J}$ s come out naturally as real antisymmetric matrices and are thus antihermitean.

To make them hermitean, we multiply them by some multiples of the imaginary unit i .

Thus, define $\mathcal{J}_x \equiv -i\mathcal{J}_x$, $\mathcal{J}_y \equiv -i\mathcal{J}_y$, $\mathcal{J}_z \equiv -i\mathcal{J}_z$, and write a general rotation as

$$R(\theta) = e^{i\Sigma_j\theta_j J_j} = e^{i\vec{\theta}\cdot\vec{J}} \tag{16}$$

treating the three real numbers θ_j and the three matrices $\mathcal{J}_j$ as two 3-dimensional vectors.

Now you should write down the generators of rotations in 4-dimensional space.

At least count how many there are.

See how easy it is to study rotations in arbitrarily high dimensional space?

We will come back to this later in these notes.

Any student of physics knows that many physical situations exhibit spherical symmetry, in which case the rotation group $SO(3)$ plays a central role

Lie algebra

Earlier, we mentioned that, in general, rotations do not commute.

Following Lie, we could try to capture this essence of group multiplication by focusing on infinitesimal rotations.

Let $R \simeq I + A$ be an infinitesimal rotation.

For an arbitrary rotation R' , consider $RR'R^{-1} \simeq (I + A)R'(I - A) \simeq R' + AR' - R'A$ (where we have consistently ignored terms of order A^2).

If rotations commute, then $RR'R^{-1}$ would be equal to R' .

Thus, the extent to which this is not equal to R' measures the lack of commutativity.

Now, suppose R' is also an infinitesimal rotation $R' \simeq I + B$.

Then $RR'R^{-1} \simeq I + B + AB - BA$, which differs from $R' \simeq I + B$ by the matrix

$$[A, B] \equiv AB - BA, \quad (17)$$

known as the commutator between A and B (from linear algebra).

For $SO(3)$, for example, A is a linear combination of the J_i s, which we shall call the generators of the Lie algebra of $SO(3)$. Thus, we can write $A = i \sum_i \theta_i J_i$ and similarly $B = i \sum_j \theta'_j J_j$. Hence $[A, B] = i^2 \sum_{ij} \theta_i \theta'_j [J_i, J_j]$, and so it suffices to calculate the commutators $[J_i, J_j]$ once and for all.

Recall that for two matrices M_1 and M_2 , $(M_1 M_2)^T = M_2^T M_1^T$. Transposition reverses the order.* Thus, $([J_i, J_j])^T = -[J_i, J_j]$. In other words, the commutator $[J_i, J_j]$ is itself an antisymmetric 3-by-3 matrix and thus can be written as a linear combination of the J_k s:

$$[J_i, J_j] = i c_{ijk} J_k \quad (18)$$

The summation over k is implied by the repeated index summation convention.

The coefficients c_{ijk} in the linear combination, with a factor of i taken out explicitly, are real numbers.

Evidently, $c_{ijk} = -c_{jik}$.

By explicit computation using (14), we find

$$[J_x, J_y] = i J_z \quad (19)$$

You should work out the other commutators or argue by cyclic substitution $x \rightarrow y \rightarrow z \rightarrow x$ that

$$[J_y, J_z] = i J_x \quad (20)$$

$$[J_z, J_x] = i J_y \quad (21)$$

The three commutation relations, (19), (20), and (21), may be summarized by

$$[J_i, J_j] = i\epsilon_{ijk}J_k \quad (\text{sum over } k \text{ implied}) \quad (22)$$

We define the totally antisymmetric symbol ϵ_{ijk} by saying that it changes sign on the interchange of any pair of indices (and hence it vanishes when any two indices are equal) and by specifying that $\epsilon_{123} = 1$.

In other words, we found that $c_{ijk} = \epsilon_{ijk}$.

Without further apology, I will often commit minor notational abuses, such as jumping back and forth between x, y, z and $1, 2, 3$.

The statement is that the commutation relations (22) tell us about the multiplicative structure of infinitesimal rotations.

By exponentiating the generators J_i , as in (15), we then manage to capture the multiplicative structure of rotations not necessarily infinitesimal.

Note that the commutation relations fix the sign of J_i .

In other words, the transformation $J_i \rightarrow -J_i$, $i = 1, 2, 3$ does not leave (22) invariant.

Structure constants

the

the

the

the

the

Lie's great insight is that the preceding discussion holds for any group whose elements $g(\theta_1, \theta_2, \dots)$ are labeled by a set of continuous parameters such that $g(0, 0, \dots)$ is the identity I .

For example, the continuous parameters would be the angles θ_i , $i=1, 2, 3$ in the case of $SO(3)$.

For these groups, now known as Lie groups, this is what you do in four easy steps:

1. Expand the group elements around the identity by letting the continuous parameters go to zero: $g \simeq I + A$.
2. Write $A = i \sum_a \theta_a T_a$ as a linear combination of the generators T_a as determined by the nature of the group.
3. Pick two group elements near the identity: $g_1 \simeq I + A$ and $g_2 \simeq I + B$. Then $g_1 g_2 g_1^{-1} \simeq I + B + [A, I + B] \simeq I + B + [A, B]$. The commutator $[A, B]$ captures the essence of the group near the identity.
4. As in step 2, we can write $B = i \sum_b \theta'_b T_b$ as a linear combination of the generators T_b . Similarly, we can write $[A, B]$ as a linear combination of the generators T_c . (We know this because, for g_1 and g_2 near the identity, $g_1 g_2 g_1^{-1}$ is also near the identity.) Plugging in, we then arrive at the analog of (18) for any continuous group, namely, the commutation relations

$$[T_a, T_b] = i f_{abc} T_c \tag{23}$$

The commutator between any two generators can be written as a linear combination of the generators.

The commutator between any two generators can be written as a linear combination of the generators.

The commutation relations between the generators define a Lie algebra, with f_{abc} referred to as the structure constants of the algebra.

The structure constants determine the Lie algebra, which essentially determines the Lie group.

This brief introduction to Lie algebra at this stage is necessarily somewhat vague, but we will go into more details soon enough.

The key idea is that we can go a long way toward understanding a continuous group by studying its Lie algebra.

Note that, while a Lie group is characterized by multiplication, its Lie algebra is characterized by commutation.

Confusio said, “When I first studied group theory, I did not clearly distinguish between Lie group and Lie algebra. That they allow totally different operations did not sink in. I was multiplying the J_i s together and couldn’t make sense of what I got.”

Absolutely, it is crucial to keep in mind that Lie group and Lie algebra are mathematically rather different structures.

In a group, you multiply (or if one wants to pick nits, compose) two elements together to get another element.

In the corresponding algebra (assuming of course that the group is continuous), you take two elements, and you commute them to get another element of the algebra. (Again, to keep the nitpickers at bay, it may be perhaps better to say two members of the algebra, since we have spoken often of group elements.)

The rotation group offers a good example.

The members $\mathcal{J}$ of its algebra are real antisymmetric matrices, but if you multiply two real antisymmetric matrices together, you certainly do not get a real antisymmetric matrix.

The algebra does not close under multiplication, only under commutation.

Perhaps one confusing point for the beginner is that to calculate the commutator, one has to, in an intermediate step, multiply two real antisymmetric matrices together.

Speaking somewhat colloquially, one has to first get out of the algebra before one can get back in.

We give presently a more mathematical formulation.

A Lie algebra is a linear space spanned by linear combinations $\sum_i \theta_i \mathcal{J}_i$ of the generators. In contrast, it makes no sense, in the rotation group, to form linear combinations of rotations. Given two rotations R_1 and R_2 , the linear sum $R_s = c_1 R_1 + c_2 R_2$ certainly is not in general a rotation.

A mathematician might define a Lie algebra abstractly as a linear vector space V equipped with a map $f : V \otimes V \rightarrow V$ satisfying various properties, for instance, $f(A, B) = -f(B, A)$.

Historically, the relation between Lie group and Lie algebra was also hinted at by the Baker-Campbell-Hausdorff formula,⁶ stating that $e^A e^B = e^C$, with $C = A + B + \frac{1}{2}[A, B] + \frac{1}{12}([A, [A, B]] + [B, [B, A]]) + \dots$.

Rotations in higher-dimensional space

With your experience with (8) and (14), it is now a cinch for you to generalize and write down a complete set of antisymmetric $N \times N$ matrices.

Start with an $N \times N$ matrix with 0 everywhere.

Stick a 1 into the m th row and n th column; due to antisymmetry, you are obliged to put a (-1) into the n th row and m th column.

Call this antisymmetric matrix $\mathcal{T}_{(mn)}$.

We put the subscripts (mn) in parentheses to emphasize that (mn) labels the matrix.

They are not indices to tell us which element of the matrix we are talking about.

As explained before, physicists like Hermite a lot and throw in a $-i$ to define the hermitean matrices $J_{(mn)} = -i\mathcal{T}_{(mn)}$.

Explicitly,

$$(J_{(mn)})^{ij} = -i(\delta^{mi}\delta^{nj} - \delta^{mj}\delta^{ni}) \quad (24)$$

To repeat, in the symbol $(J_{(mn)})^{ij}$, which we will often write as $J_{(mn)}^{ij}$ for short, the indices i and j indicate, respectively, the row and column of the entry $(J_{(mn)})^{ij}$ of the matrix $J_{(mn)}$, while the indices m and n , which I put in parentheses for pedagogical clarity, indicate which matrix we are talking about. The first index m on $J_{(mn)}$ can take on N values, and then the second index n can take on only $(N - 1)$ values, since, evidently, $J_{(mm)} = 0$. Also, since $J_{(nm)} = -J_{(mn)}$, we require $m > n$ to avoid double counting. Thus, there are only $\frac{1}{2}N(N - 1)$ real antisymmetric N -by- N matrices $J_{(mn)}$. The Kronecker deltas in (24) merely say what we said in words in the preceding paragraph.

As before, an infinitesimal rotation is given by $R \simeq I + A$ with the most general A a linear combination of the $J_{(mn)}$ s: $A = i \sum_{m,n} \theta_{(mn)} J_{(mn)}$, where the antisymmetric coefficients $\theta_{(mn)} = -\theta_{(nm)}$ denote $\frac{1}{2}N(N - 1)$ generalized angles. (As a check, for $N = 2$ and 3 , $\frac{1}{2}N(N - 1)$ equals 1 and 3 , respectively.) The matrices $J_{(mn)}$ are the generators of the group $SO(N)$.

Rotation around an axis or rotation in a plane

Confusio speaks up: “All this makes sense, but I am nevertheless puzzled. We physics students are used to rotating around an axis,^ in particular, the x , y , and z axes. For $N = 3$, the case that I am most familiar with, the generators are labeled by only one index, to wit, J_x , J_y , and J_z . Strange that $J_{(mn)}$ is in general labeled by two indices.”

We should thank Confusio.

This notational peculiarity has indeed confused many students.

The explanation is simple.

Dear reader, please think: What does it mean to rotate around the fifth axis in, say, 8-dimensional Euclidean space?

It makes no sense!

However, it makes perfect sense to rotate in the (3-7) plane, for example: simply rotate the third axis and the seventh axis into each other.

For $N=3$, the indices m, n take on three values, and so we can write $J_x = J_{23}$, $J_y = J_{31}$, and $J_z = J_{12}$.

In 3-dimensional space, and only in 3-dimensional space, a plane is uniquely specified by the vector perpendicular to it.

Thus, a rotation commonly spoken of as a rotation around the z -axis is better thought of as a rotation in the (12)-plane, that is, the (xy)-plane.

The Lie algebra for $SO(N)$

Our next task is to work out the Lie algebra for $SO(N)$, namely, the commutators between the $J_{(mn)}$ s.

You could simply plug in (24) and chug away..... exercise!

But a more elegant approach is to work out $SO(4)$ as an inspiration for the general case. First, $[J_{(12)}, J_{(34)}] = 0$, as you might expect, since rotations in the (1-2) plane and in the (3-4) plane are like gangsters operating on different turfs. Next, we tackle $[J_{(23)}, J_{(31)}]$. Notice that the action takes place entirely in the $SO(3)$ subgroup of $SO(4)$, and so we already know the answer: $[J_{(23)}, J_{(31)}] = [J_x, J_y] = i J_z = i J_{(12)}$. These two examples, together with antisymmetry $J_{(mn)} = -J_{(nm)}$, in fact take care of all possible cases. In the commutator $[J_{(mn)}, J_{(pq)}]$, there are three possibilities for the index sets (mn) and (pq) : (i) they have no integer in common, (ii) they have one integer in common, or (iii) they have two integers in common. The commutator vanishes in cases (i) and (iii), for trivial (but different) reasons. In case (ii), suppose $m = p$ with no loss of generality, then the commutator is equal to $i J_{(nq)}$.

We obtain, for any N,

$$[J_{(mn)}, J_{(pq)}] = i(\delta_{mp} J_{(nq)} + \delta_{nq} J_{(mp)} - \delta_{np} J_{(mq)} - \delta_{mq} J_{(np)}) \quad (25)$$

This may look rather involved to the uninitiated, but in fact it simply states in mathematical symbols the last three sentences of the preceding paragraph.

First, on the right hand side, a linear combination of the J s (as required by the general argument above) is completely fixed by the first term by noting that the left hand side is antisymmetric under three separate interchanges: $m \leftrightarrow n$, $p \leftrightarrow q$, and $(mn) \leftrightarrow (pq)$.

Next, all those Kronecker deltas just say that if the two sets (mn) and (pq) have no integer in common, then the commutator vanishes.

If they do have an integer in common, simply “cross off” that integer.

For example, $[J_{(12)}, J_{(14)}] = iJ_{(24)}$ and $[J_{(23)}, J_{(31)}] = -iJ_{(21)} = iJ_{(12)}$.

Confusio looks a bit puzzled. “But (25) does not look much like the canonical definition (23) of a Lie algebra.”

We tell him that his confusion is merely notational.

The generators are labeled by the peculiar symbol (mn) made out of two letters from a certain alphabet and parentheses.

We might as well call it a , with $a = 1, 2, \dots, N/2(N-1)$, and write J_a for the generator $J_{(mn)}$.

For an explicit example, consider $SO(4)$ with its six generators $J_{(mn)}$.

Make a 1-to-1 map of the indices (mn) $\leftrightarrow a$ as follows: $(12) \leftrightarrow 1$, $(13) \leftrightarrow 2$, $(14) \leftrightarrow 3$, $(23) \leftrightarrow 4$, $(24) \leftrightarrow 5$, $(34) \leftrightarrow 6$.

How the $N/2(N-1)$ J_a s are ordered does not matter as long as the correspondence (mn) $\leftrightarrow a$ is 1-to-1.

The right hand of (25) is just a linear combination of J_c s, and hence (25) can indeed be written as $[T_a, T_b] = if_{abc}T_c$, as in (23).

We will come back to this point when we discuss the adjoint representation later.

2 Representing Group Elements by Matrices

Group elements may be represented by matrices.

Representation theory is full of delightful theorems and surprising results, such as the Great Orthogonality theorem.

A favorite saying: character is a function of class.

Indeed, some of us feel that constructing character tables is loads of fun.

The important concept of real, pseudoreal, and complex representations is studied, culminating in the construction of a reality checker.

Using the character table, you can count the number of square roots of any element in a finite group.

The elegant theorem stating that crystals with five-fold symmetry are impossible is proven.

After that, we relax and have fun with number theory, discussing various results associated with Euler, Fermat, Wilson, and Frobenius.

Representation Theory

What is a representation?

More than a century later, while representation theory is not exactly easy, it does not seem all that difficult, either.

The notion of representing group elements by matrices is both natural and intuitive.

Given a group, the idea is to associate each element g with a $d \otimes d$ matrix $D(g)$ such that

$$D(g_1)D(g_2) = D(g_1g_2) \tag{1}$$

for any two group elements g_1 and g_2 .

The matrix $D(g)$ is said to represent the element g , and the set of matrices $D(g)$ for all $g \in G$ is said to furnish or provide a representation of G .

The size of the matrices, d , is known as the dimension of the representation.

The requirement (1) says that the set of matrices $D(g)$ “reflects” or “mirrors” the multiplicative table of the group.

In words, the product g_1g_2 of two group elements g_1 and g_2 is represented by the product of the matrices representing g_1 and g_2 respectively.

To emphasize this all-important concept of representation, let us write (1) “graphically” as

$$\begin{array}{ccccccc} g_1 & \cdot & g_2 & = & g_1 \cdot g_2 & & \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & & \\ D(g_1) & \cdot & D(g_2) & = & D(g_1 \cdot g_2) & & \end{array} \tag{2}$$

Note that the symbol $\cdot$ denotes two distinct concepts: in the top row, the composition, or more colloquially, the multiplication, of two group elements; in the bottom row, the multiplication of two matrices. (As already mentioned earlier, we often omit the dot, as in (1), for example.)

All this will become clearer with the examples to be given here.

Consider S_4 , the permutation group of four objects. Think of the four objects as the four vectors* $v_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$, $v_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$, $v_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$, and $v_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$. Then we can represent the element (2413), which takes $2 \rightarrow 4$, $4 \rightarrow 1$, $1 \rightarrow 3$, and $3 \rightarrow 2$, by the 4-by-4 matrix $D(2413) = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$. By construction, $D(2413)v_2 = v_4$, $D(2413)v_4 = v_1$, and so on. The action of the matrix $D(2413)$ on the four vectors mirrors precisely the action of the permutation (2413) on the four objects labeled 1, 2, 3, and 4. Similarly, we have, for example, $D(34) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$.

According to what we have learned in chapter I.2, we have $(34)(2413) = (23)(14)$. Here, let us multiply the two matrices $D(34)$ and $D(2413)$ together. (Go ahead, do it!) We find $D(34)D(2413) = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$, which is precisely $D((23)(14))$, as expected. This verifies (1), at least in this particular instance. Exercise! Write down a few more matrices in this 4-dimensional representation of S_4 and multiply them.

NOTE: To lessen clutter, we have omitted the arrow on top of the vectors here.

I presume that you are not surprised that we have found a 4-dimensional representation of S_4 .

In summary, the group S_4 can be represented by 24 distinct 4×4 matrices.

Note that these are very special matrices, with 0 almost everywhere except for four 1s, with one single 1 in each column (and in each row).

All this should be fairly self-evident: what is the difference between four vectors labeled v_1, v_2, v_3, v_4 and four balls labeled 1, 2, 3, 4?

Group elements and the matrices that represent them

In our example, the matrix $D(34)$ represents the permutation (34), but physicists might say that $D(34)$ is essentially what they mean by (34).

In fact, physicists often confound group elements with the matrices that represent them.

For example, when a physicist thinks of a rotation, he or she typically has in mind a 3×3 matrix.

A mathematician, in contrast, might think of rotations as abstract entities living in some abstract space, defined entirely by how they multiply together.

In practice, many of the groups used in theoretical physics are defined by the matrices representing them, for example, the Lorentz group described earlier.

Very roughly speaking, the representation of a group is like a photograph or a map of the group, to the extent that it preserves the multiplicative structure of the group.

A photo or a map of a village is of course not the village itself, but it shows accurately how various buildings and geographical features are situated relative to one another.

On a practical level, it is much easier to tell a computer to multiply matrices together than to feed it the multiplication table of a group.

Note that (1) implies that $D(I) = I_d$.

Here I have carefully distinguished the I on the left (denoting that special group element called the identity) from the I_d on the right (denoting the $d \times d$ identity matrix).

The meaning of I_d depends on the particular representation being discussed.

In contrast, the other I is an abstract entity fixed once the group is specified.

More sloppily, I might have abused notation slightly and written the more mystifying but fairly self-evident $D(I) = I$.

To show that $D(I) = I_d$, observe that $D(I)D(g) = D(Ig) = D(g)$ for any g .

Multiply by $D(g^{-1})$ from the right, and we obtain $(D(I)D(g))D(g^{-1}) = D(I)(D(g)D(g^{-1}))$, which, due to (1), reduces to $D(I)D(I) = D(I)$.

Multiplying by $(D(I))^{-1}$, we obtain $D(I) = I_d$.

This also tells us that $D(g^{-1}) = D(g)^{-1}$, as we would expect, since the representation is supposed to mirror the multiplicative structure of the group.

Introduction to representation theory

Now that we know what a representation is, we can naturally think of many questions.

Does every group G have a representation?

How many representations does it have?

An infinite number, perhaps?

What are some general properties of representations?

How do we characterize these representations and distinguish among them?

How is your mathematical intuition?

Do you feel that, the more sophisticated a group, the more representations it ought to have?

But then, how would you measure the “sophistication” of a group?

Is it merely the number of elements?

Or, more intelligently, do you feel that sophistication would be measured more by the number of different types of elements?

Recall the notion of equivalence classes from earlier.

We have a partial answer to the first question.

We learned earlier that every finite group is isomorphic to a subgroup of S_n , and since S_n has a matrix representation, every finite group can be represented by matrices.

As for continuous groups, in the list of examples given earlier, almost all groups—the rotation groups and the Lorentz group, for example—are defined in terms of matrices, so a fortiori they can be represented by matrices.

An exception appears to be the additive group of real numbers.

How in the world could addition be represented by multiplication?

You smile, since you already know the answer.

Let $D(u)$ be the 1-dimensional matrix e^u : then $D(u)D(v) = D(u + v)$.

Actually, we do not even have to invoke the exponential function.

Consider the 2-dimensional matrix

$$D(u) = \begin{pmatrix} 1 & 0 \\ u & 1 \end{pmatrix} \tag{3}$$

Verify that $D(u)D(v) = D(u + v)$ and $D(0) = I$. (Note that 0 denotes the identity of the additive group.)

I bet you didn't know that addition could be represented by multiplying 2-by-2 matrices together.

Let me also ask you, has the group described by (3) ever appeared in physics?

For those of you who do not know the answer, just wait, and it will be revealed later.

Indeed, the Lorentz group mentioned earlier also defines a 2-dimensional representation of the additive group, since $D(\phi_1)D(\phi_2) = D(\phi_1 + \phi_2)$. (Recall that ϕ represents the boost angle.)

You may have realized that this representation of addition secretly also involves the exponential function.

So by the physicist's laughable standard of rigor, it certainly seems that all the groups you are likely to encounter in physics can be represented by matrices.

To answer the second question that introduced this section, namely, how many representations a group might have, we are first obliged to mention that, in representation theory, the trivial representation $D(g) = 1$, for every $g \in G$, also counts as a perfectly valid representation.

The basic requirement (1) of being a representation is certainly satisfied, since $D(g_1)D(g_2) = 1 \cdot 1 = 1 = D(g_1g_2)$.

Yes indeed, this representation is trivial, hence the name.

But as you will see, in the representation theory to be developed now and the next part, it is important to include it.

This is perhaps reminiscent of the introduction of the number 0 in the history of mathematics.

Here the notion of faithful versus unfaithful representations naturally suggests itself.

To use a more mathematical language, we say that a d -dimensional representation is a map of the group G into some subgroup of $GL(d, \mathbb{C})$.

The requirement (1) merely says that the map is homomorphic.

But if in addition the map is isomorphic, that is, one-to-one, then the representation is faithful.

Otherwise, it is unfaithful.

As already mentioned, many of the groups given earlier are defined in terms of matrices.

For example, the rotation group $SO(3)$ is defined by 3×3 orthogonal matrices, as discussed in detail earlier.

Naturally, these representations are known as defining or fundamental representations.

The defining representation of $SO(3)$ is 3-dimensional.

As another example, the defining representation of Z_N is 1-dimensional, namely, the element $e^{i2\pi j/N}$ is represented by itself for $j = 0, \dots, N-1$.

But interestingly, after some thought, we realize that we can also represent $e^{i2\pi j/N}$ by $e^{i2\pi kj/N}$ for some fixed integer k , which can take on any of the values $0, \dots, N-1$.

Check that this indeed furnishes a representation:

$$D(e^{i2\pi j_1/N})D(e^{i2\pi j_2/N}) = e^{i2\pi kj_1/N}e^{i2\pi kj_2/N} = e^{i2\pi k(j_1+j_2)/N} = D(e^{i2\pi (j_1+j_2)/N})$$

I am being extra pedantic here.

Thus, Z_N actually has N different 1-dimensional representations, labeled by the integer k .

What is the $k = 0$ representation? (Yes, it is the trivial representation.)

More specifically, the group Z_3 has not only the 1-dimensional representation $\{1, \omega, \omega^2\}$ (where $\omega \equiv e^{2\pi i/3}$ is the cube root of 1) but also the nontrivial 1-dimensional representation $\{1, \omega^2, \omega\}$.

Before reading on, you should verify that this indeed furnishes a representation.

For example, $\omega^2\omega^2 = \omega^4 = \omega$.

Wait, you just read that Z_3 has three different irreducible representations.

What is the third?

Character is a function of class

Now that we know that a given group can have many different representations, let us label the different representations by a superscript r , s , $\dots$, and write $D^{(r)}(g)$ for the matrix representing the element g in the representation r .

Given a representation $D^{(r)}(g)$, define the important concept of the character $\chi^{(r)}$ of the representation by $\chi^{(r)}(g) \equiv \text{tr } D^{(r)}(g)$.

The character, as the name suggests, helps characterize the representation.

Nominally, the character depends on r and g .

Recall from earlier, however, that the elements of a group can be divided up into equivalence classes. Two elements g_1 and g_2 are equivalent ($g_1 \sim g_2$) if there exists another element f such that

$$g_1 = f^{-1}g_2f \tag{4}$$

We then find

$$\begin{aligned} \chi^{(r)}(g_1) &= \text{tr } D^{(r)}(g_1) = \text{tr } D^{(r)}(f^{-1}g_2f) = \text{tr } D^{(r)}(f^{-1})D^{(r)}(g_2)D^{(r)}(f) \\ &= \text{tr } D^{(r)}(g_2)D^{(r)}(f)D^{(r)}(f^{-1}) = \text{tr } D^{(r)}(g_2)D^{(r)}(I) \\ &= \text{tr } D^{(r)}(g_2) = \chi^{(r)}(g_2) \end{aligned} \tag{5}$$

where in the third, fifth, and sixth equalities we used (1) and in the fourth equality we used the cyclicity of the trace.

In other words, if $g_1 \sim g_2$, then $\chi^{(r)}(g_1) = \chi^{(r)}(g_2)$.

Thus,

$$\chi^{(r)}(c) = \text{tr } D^{(r)}(g) \quad (\text{for } g \in c) \tag{6}$$

The trace on the right hand side does not depend on g as such, but only on the class that g belongs to.

All the elements of a given equivalence class have the same character.

As an example, we learned earlier that A_4 has four equivalence classes.

For a given representation r of A_4 , $\chi^{(r)}(c)$ is a function of c , a variable that takes on four values.

We can now proudly utter perhaps the most memorable statement in group theory:

“Character is a function of class.”

Equivalent representations

As for how many representations a group might have, we all agree, first of all, that two representations, $D(g)$ and $D'(g)$, are really the same representation (more formally, the two representations are equivalent) if they are related by a similarity transformation

$$D'(g) = S^{-1}D(g)S. \tag{7}$$

From linear algebra, $D(g)$ and $D'(g)$ are essentially the same matrix, merely written in two different bases, with the matrix S relating one set of basis vectors to the other set.

Put another way, given a representation $D(g)$, define $D'(g)$ by (7) with some S whose inverse exists.

Then $D'(g)$ is also a representation, since $D'(g_1)D'(g_2) = (S^{-1}D(g_1)S)(S^{-1}D(g_2)S) = S^{-1}D(g_1)D(g_2)S = S^{-1}D(g_1g_2)S = D'(g_1g_2)$.

Note that it is the same S for all g .

Think of it as follows.

Suppose we have found a representation of a group G , for example, the 4-dimensional representation of S_4 described earlier.

In other words, we list $24 = 4!$ 4×4 matrices $D(g)$ satisfying (1).

Some ugly dude could come along, choose some ugly 4×4 matrix S , and use (7) to produce a list of 24 ugly 4-by-4 matrices $D'(g)$ satisfying (1).

If he chooses a particularly messy S , the two sets of matrices $D'(g)$ and $D(g)$ could look very different.

If we are given two representations, how do we decide whether they are equivalent or not?

Taking the trace of (7) and once again using the cyclicity of the trace, we obtain

$$\chi'(c) = \text{tr } D'(g) = \text{tr } SD(g)S^{-1} = \text{tr } D(g)S^{-1}S = \text{tr } D(g) = \chi(c) \quad (8)$$

where g is a member of the class c .

Thus, if there exists some class c for which $\chi'(c) \neq \chi(c)$, we can conclude immediately that the two representations are in fact different.

What if $\chi'(c) = \chi(c)$ for all c ?

If this holds for only one or two c , physicists of the “black sheep school of thought” might still admit that it could be a coincidence, but for all c ?

Most “reasonable” theoretical physicists would say that it is strong circumstantial evidence indicating that the two representations are in fact the same.

Indeed, physicist intuition is right.

We will see in the next part that various theorems state that for two different representations r and s , the characters $\chi^{(r)}(c)$ and $\chi^{(s)}(c)$ are “more different than different”: they are orthogonal in some well-defined sense.

Reducible or irreducible representation

Now we come to the all-important notion of whether a given representation is reducible or irreducible.

For the sake of definiteness, focus on SO(3).

We have the trivial 1-dimensional representation $D^{(1)}(g) = 1$ and the 3-dimensional defining representation $D^{(3)}(g)$.

Are there other representations?

Think about it before reading on.

Can you give me an 8-dimensional representation?

“Sure,” you say, “you want an 8-dimensional representation for SO(3). I give you an 8-dimensional representation. Here it is:”

$$D(g) = \left(\begin{array}{c|c|c|c} D^{(1)}(g) & 0 & 0 & 0 \\ \hline 0 & D^{(1)}(g) & 0 & 0 \\ \hline 0 & 0 & D^{(3)}(g) & 0 \\ \hline 0 & 0 & 0 & D^{(3)}(g) \end{array} \right)$$

(9)

The vertical and horizontal lines are merely to guide the eye.

By the way, a matrix with this form is said to be block diagonal: it contains smaller matrices along its diagonal, with all other entries equal to zero.

Note that the symbol 0 in (9) carries several different meanings: it could denote a 1-by-1 matrix with its single entry equal to zero, or a 1-by-3 rectangular matrix with all its entries equal to zero, or a 3-by-1 rectangular matrix with all its entries equal to zero, or a 3-by-3 square matrix with all its entries equal to zero.

Each element g is represented by an 8-by-8 matrix $D(g)$ (8 since $1 + 1 + 3 + 3 = 8$).

Show that this is indeed a representation; that is, it satisfies (1).

Ah, you have stacked two copies of $D(1)(g)$ and two of $D(3)(g)$ together.

Indeed, by this cheap trick, you could construct representations with any dimension.

You and I, being reasonable people, should agree that $D(g)$ does not count as a “new” representation.

The representation $D(g)$ is known as reducible, and usually written as a direct sum of the representations it reduces into: in our example, $D(g) = D^{(1)}(g) \oplus D^{(1)}(g) \oplus D^{(3)}(g) \oplus D^{(3)}(g)$.

Representations that are not reducible are called irreducible.

Clearly, we should focus on irreducible, rather than reducible, representations.

It is clear as day that $D(g)$ in the form given above is reducible.

But that ugly dude mentioned earlier could come along again and present you with a set of matrices $D'(g) = S^{-1}D(g)S$.

If he chooses a particularly messy S , all those zeroes in (9) would get filled in, and we would have a hard time recognizing that $D'(g)$ is reducible.

Going back to the definition of the trivial representation $D^{(1)}(g) = 1$, you might have wondered why we used 1 and not the $k \otimes k$ identity matrix I_k for some arbitrary positive integer k .

The answer is that the representation would then be reducible unless $k = 1$.

The representation $D^{(1)}$ may be trivial, but it is not reducible.

One goal of representation theory is to develop criteria to determine whether a given representation is irreducible or not and to enumerate all possible irreducible representations.

Since every group has an infinity of reducible representations, the real question is to figure out how many irreducible representations it has.

Restriction to a subgroup

A representation of a group G clearly also furnishes a representation of any of its subgroups.

Denote the elements of the subgroup H by h .

If $D(g_1)D(g_2) = D(g_1g_2)$ for any two group elements g_1 and g_2 of G , then a fortiori $D(h_1)D(h_2) = D(h_1h_2)$ for any two group elements h_1 and h_2 of H .

We refer to this representation of H as the representation of G restricted to the subgroup H .

When restricted to a subgroup H , an irreducible representation of G will in general not be an irreducible representation of H .

It will, in all likelihood, decompose, or fall apart, into a bunch of irreducible representations of H .

The reason is clear: there may well exist a basis in which $D(h)$ is block diagonal (that is, has the form such as that shown in (9)) for all h in H , but there is no reason in general to expect that $D(g)$, for all g in G but not in H , would also be block diagonal.

Simply stated, there are, by definition, fewer elements in H than in G .

How an irreducible representation of a group G decomposes upon restriction of G to a subgroup H will be a dominant and recurring theme in these notes.

Unitary representations

The all-important unitarity theorem states that finite groups have unitary representations, that is to say, $D^\dagger(g)D(g) = I$ for all g and for all representations.

In practice, this theorem is a big help in finding representations of finite groups.

As a start, we can eliminate some proposed representations by merely checking if the listed matrices are unitary or not.

At this point, our friend Dr. Feeling strolls by.

“Let’s get an intuitive feel for this theorem,” he says helpfully.

Suppose the representation $D(g)$ is 1×1 , that is, merely a complex number $re^{i\theta}$.

Earlier, we showed that in a finite group, if we keep multiplying g by itself, eventually it has to come back to the identity: $g^k = I$ for some integer k .

But g^k is represented by $D(g^k) = D(g)^k = r^k e^{ik\theta}$.

No way this could get back to 1 if $r \neq 1$.

But if $r = 1$, then $D^\dagger(g)D(g) = e^{-i\theta} e^{i\theta} = 1$; that is, $D(g)$ is unitary.

This, in essence, is why the theorem must be true.

Proof of the unitarity theorem

The so-called rearrangement lemma says:

Given a function on the group elements, we have, for any $g' \in G$,

$$\sum_{g \in G} f(g) = \sum_{g \in G} f(g'g) = \sum_{g \in G} f(gg') \quad (10)$$

The three sums are actually the same sum; they differ only by having the terms rearranged.

We are now ready to prove the unitarity theorem.

Suppose that a given representation $\tilde{D}(g)$ is nonunitary.

Define

$$H = \sum_g \tilde{D}(g)^\dagger \tilde{D}(g) \quad (11)$$

where the sum runs over all elements $g \in G$.

We note that, for any g'

$$\begin{aligned} \tilde{D}(g')^\dagger H \tilde{D}(g') &= \sum_g \tilde{D}(g')^\dagger \tilde{D}(g)^\dagger \tilde{D}(g) \tilde{D}(g') = \sum_g (\tilde{D}(g) \tilde{D}(g'))^\dagger \tilde{D}(g) \tilde{D}(g') \\ &= \sum_g (\tilde{D}(gg'))^\dagger \tilde{D}(gg') = H \end{aligned} \quad (12)$$

The last equality holds because of the rearrangement lemma.

The matrix H is remarkably “invariant.”

Since H is hermitean, there exists a unitary matrix W such that $Q^2 = W^\dagger H W$ is diagonal and real.

We now show that in addition, the diagonal elements are not only real but also positive.

Hence the notation Q^2 : we can take the square root of the diagonal matrix Q^2 to obtain the diagonal and real matrix Q .

To show this, we invoke a theorem from linear algebra, that for any matrix M , the matrix $M^\dagger M$ has non-negative eigenvalues.

Let ψ be the column vector with 1 in the j th entry and 0 everywhere else.

Then

$$(\rho^2)_{jj} = \psi^\dagger \rho^2 \psi = \psi^\dagger W^\dagger H W \psi = \sum_g (\psi^\dagger W^\dagger) \tilde{D}(g)^\dagger \tilde{D}(g) (W \psi) = \sum_g \phi(g)^\dagger \phi(g) > 0 \quad (13)$$

Here we define $\phi(g) = \tilde{D}(g) W \psi$.

Thus, the matrix $\mathcal{Q}$ exists with the stated properties.

Define $D(g) \equiv \mathcal{Q} W^\dagger \tilde{D}(g) W \mathcal{Q}^{-1}$.

We now show that $D(g)$ is unitary.

Simply calculate: $D^\dagger(g) = \mathcal{Q}^{-1} W^\dagger \tilde{D}(g)^\dagger W \mathcal{Q}$, so that

$$\begin{aligned} D^\dagger(g) D(g) &= \rho^{-1} W^\dagger \tilde{D}(g)^\dagger W \rho^2 W^\dagger \tilde{D}(g) W \rho^{-1} \\ &= \rho^{-1} W^\dagger \tilde{D}(g)^\dagger H \tilde{D}(g) W \rho^{-1} \\ &= \rho^{-1} W^\dagger H W \rho^{-1} = \rho^{-1} \rho^2 \rho^{-1} = I \end{aligned} \quad (14)$$

where the third equality holds because of (12).

The unitarity theorem is proved.

It is instructive to see how the theorem works if $\tilde{D}(g)$ is already unitary.

Then $H = N(G)I$, with $N(G)$ the number of elements in the group, so that $W = I$, $Q = \sqrt{N(G)}I$ is proportional to the identity matrix, and hence $D(g) = Q \tilde{D}(g) Q^{-1} = \tilde{D}(g)$.

Note that in almost all the examples given thus far, the representation matrices are real rather than complex.

A real unitary matrix is orthogonal, of course; in other words, for $D(g)$ real, $D^T(g)D(g) = I$ for all g .

We will derive criteria for deciding whether a given irreducible representation is real or complex later

Compact versus non-compact

To physicists, the natural thing to do is to check whether the unitarity theorem holds for groups other than finite groups.

Well, it certainly works for $SO(2)$: $R(\theta)^T R(\theta) = I$, as was verified explicitly earlier essentially by definition.

Indeed, by the discussion there, it works for the rotation group in any dimension.

To a sloppy physicist, finite, infinite, what's the difference?

Assume everything converges unless proven otherwise.

Shoot first, and ask questions later.

So surely group representations have to be unitary in general.

The mathematician is aghast: you can't ignore rigor so blithely!

What about the Lorentz group with $L(\varphi) = \begin{pmatrix} \cosh \varphi & \sinh \varphi \\ \sinh \varphi & \cosh \varphi \end{pmatrix}$?

It is manifestly not true that $L(\varphi)^T L(\varphi)$ is equal to the identity.

In fact, nothing as fancy as the Lorentz group is needed for a counterexample; how about the representation for the additive group mentioned earlier: $D(x) = \begin{pmatrix} 1 & 0 \\ x & 1 \end{pmatrix}$?

Certainly, $D(x)^T D(x) = \begin{pmatrix} 1+x^2 & x \\ x & 1 \end{pmatrix} \neq I$.

Indeed, the claimed “physicist theorem” also fails for discrete groups if we restrict x to the integers.

So the issue is not a question of whether the group is discrete or continuous.

Rather the issue is whether the group is compact or not.

The notion of compactness arises naturally in the present context.

The sloppy physicist would say that the theorem that representations are unitary should also hold for continuous groups; the jump from finite group to continuous group just amounts to replacing the sum $\sum_g$ over the elements of a finite group, following Newton and Leibniz, by some kind of integral $\int d\mu(g)$ over the continuous group (with $d\mu(g)$ some kind of integration measure).

Thus, if the relevant integral $\int d\mu(g)(\dots)$ converges, the group is known to be compact, and the proof given above for finite groups formally goes through.

The representations of compact groups are unitary.

But what if the integral diverges?

Precisely, growls our mathematician friend, if an integral diverges, even physicists have to sit up and be on alert for possible danger.

We will come back later to a precise definition of compactness, but for the moment let us simply note that this discussion makes a lot of sense.

The rotation groups are parametrized by angles that typically are bounded between 0 and 2π , but for the Lorentz group the boost angle φ is not really an angle and ranges from $-\infty$ to $+\infty$. (Note that the issue is more intricate than this naive view. As mentioned earlier, we could also parametrize the Lorentz group using the more physical relative velocity v between inertial frames defined by $\tanh \varphi \equiv v$, with v ranging between -1 and $+1$.)

It really does matter what $\int d\mu(g)$ is.

Product representation

Stacking representations on top of one another as in (9) gives an easy but mostly uninteresting way of producing a larger representation, which we might call a direct sum representation, out of smaller representations.

The direct product of matrices discussed from linear algebra provides a more interesting way of constructing a larger representation out of smaller ones.

Given two representations, of dimension d_r and d_s , with representation matrices $D^{(r)}(g)$ and $D^{(s)}(g)$, respectively, we can define the direct product representation defined by the direct product matrices $D(g) = D^{(r)}(g) \otimes D^{(s)}(g)$, namely, the $d_r d_s \times d_r d_s$ matrix given by

$$D(g)_{a\alpha, b\beta} = D^{(r)}(g)_{ab} D^{(s)}(g)_{\alpha\beta} \quad (15)$$

We have intentionally used different letters to denote the indices on $D^{(r)}(g)$ and $D^{(s)}(g)$ to emphasize that they are entirely different beasts running over different ranges; in particular, $a, b = 1, \dots, d_r$, and $\alpha, \beta = 1, \dots, d_s$.

The rule for multiplying direct product matrices together (from linear algebra)

$$\begin{aligned} D(g)D(g') &= (D^{(r)}(g) \otimes D^{(s)}(g))(D^{(r)}(g') \otimes D^{(s)}(g')) \\ &= (D^{(r)}(g)D^{(r)}(g')) \otimes (D^{(s)}(g)D^{(s)}(g')) = D(gg') \end{aligned} \quad (16)$$

shows explicitly that the product representation is indeed a representation.

In general, however, there is no reason for this product representation to be irreducible.

We will learn to determine how the product representation reduces to a direct sum of irreducible representations.

The character of the direct product representation is easily calculated by setting the index $a\alpha$ equal to the index $b\beta$ in (15) and summing:

$$\chi(c) = \sum_{a\alpha} D(g)_{a\alpha, a\alpha} = \left(\sum_a D^{(r)}(g)_{aa} \right) \left(\sum_{\alpha} D^{(s)}(g)_{\alpha\alpha} \right) = \chi^{(r)}(c) \chi^{(s)}(c) \quad (17)$$

As usual, c denotes the class that the group element g belongs to.

The character of a direct product representation is the product of the characters of the representations that produce it.

This result nicely parallels the statement mentioned earlier that the character of a direct sum representation is the sum of the characters of the representations that produce it.

Notice that nothing in this discussion requires that the representations r and s be irreducible.

In physics, it is often useful to think in terms of the objects furnishing the representations.

Let ϕ_a , $a = 1, \dots, d_r$, denote the d_r objects that transform into linear combinations of one another and thus furnish the representation r .

Similarly, let ξ_α , $\alpha = 1, \dots, d_s$, denote the d_s objects that furnish the representation s .

Then, the $d_r d_s$ objects $\phi_a \xi_\alpha$ furnish the direct product representation $r \otimes s$.

As we shall see later, these abstract mathematical objects are actually realized in quantum mechanics as wave functions.

Finally, let me quote what Wigner said about the article that von Neumann gave him:

“Soon I was lost in the enchanting world of vectors and matrices, wave functions and operators. This reprint was my primary introduction to representation theory, and I was charmed by its beauty and clarity. I saved the article for many years out of a certain piety that these things create.”

I hope that you find it equally enchanting.

Schur's Lemma and the Great Orthogonality Theorem

Now we prove a number of elegant theorems, considered by many to be among the most beautiful in mathematics.

We will then use these theorems to determine the irreducible representations of various finite groups.

Those readers exposed to group theory for the first time might prefer to skip the detailed proofs and to merely absorb what these theorems say.

Schur's lemma

A crucial theorem in representation theory, known as Schur's lemma, states the following:

If $D(g)$ is an irreducible representation of a finite group G and if there is some matrix A such that $AD(g) = D(g)A$ for all g , then $A = \lambda I$ for some constant λ .

What does this mean?

If I give you a bunch of matrices $D_1, D_2, \dots, D_n$, the identity matrix I commutes with all these matrices, of course.

But it is also quite possible for you to find a matrix A , not the identity, that commutes with all n matrices.

The theorem says that you can't do this if the given matrices $D_1, D_2, \dots, D_n$ are not any old bunch of matrices you found hanging around the street corner, but the much-honored representation matrices furnishing an irreducible representation of a group.

To prove Schur's lemma, let's start with a small lemma to the lemma:

A can be taken to be hermitean with no loss of generality.

To see this, recall that $D(g)$ is unitary according to the "unitary theorem".

Take the hermitean conjugate of $AD(g) = D(g)A$ to obtain $D(g)^\dagger A^\dagger = A^\dagger D(g)^\dagger$.

Since $D(g)$ is unitary, we can write this as $D(g)^{-1}A^\dagger = A^\dagger D(g)^{-1}$, and hence $A^\dagger D(g) = D(g)A^\dagger$.

Adding and subtracting, we obtain $(A + A^\dagger)D(g) = D(g)(A + A^\dagger)$ and $i(A - A^\dagger)D(g) = D(g)i(A - A^\dagger)$.

The statement of Schur's lemma holds for the two hermitean matrices $(A + A^\dagger)$ and $i(A - A^\dagger)$.

Thus, we might as well focus on each of them, and rename the original matrix H to emphasize its hermiticity.

A useful notation from Linear Algebra

At this point, let us introduce a notation commonly used in several areas of physics.

Up to now, the indices carried by vectors and matrices have been written as subscripts.

My only justification is that it would be sort of confusing for beginners to see superscripts right from the start.

Anyway, I now decree that complex vectors carry an upper index: ψ^i .

Next, for each vector carrying an upper index ϕ^i , introduce a vector carrying a lower index by defining

$$\phi_i \equiv \phi^{i*} \quad (\text{not - no difference if everything real})$$

To be precise, the right hand side here is $(\phi^i)^*$, the complex conjugate of ϕ^i .

At the most pedestrian level, you could regard this as a cheap move merely to avoid writing the complex conjugation symbol $*$.

But then it follows that the scalar product of two complex vectors is given by

$$\phi^\dagger \psi = \phi_i \psi^i$$

The right hand side is shorthand for $\sum_i \phi_i \psi^i = \sum_i \phi^{i*} \psi^i$.

To Einstein's repeated index summation convention we can now add another rule: an upper index must be summed with a lower index.

We will never sum an upper index with an upper index, or sum a lower index with a lower index.

With this rule, the matrix M would then have to be written as M^i_j so that the vector $M\psi$ would be given by $(M\psi)^i = M^i_j \psi^j$.

In some sense, then, the upper index is a row index, and the lower index a column index.

This is consistent with the fact that in the scalar product above $\phi^\dagger \psi = \phi_i \psi^i$, the vector $\phi^\dagger$ is a row vector, that is, a 1 x n rectangular matrix, and the lower index labels the n different columns.

Proof of Schur's lemma

We want to prove that if $HD(g) = D(g)H$ for all g , then $H = \lambda I$ for some constant λ .

Since H is hermitean, it can be diagonalized: $H = W^\dagger H' W$ with H' diagonal and W some unitary matrix.

Transform to that basis: $D(g) = W^\dagger D'(g) W$.

The statement of the theorem $HD(g) = D(g)H$ becomes $(W^\dagger H' W)(W^\dagger D'(g) W) = (W^\dagger D'(g) W)(W^\dagger H' W)$, which becomes, upon multiplication by W from the left and $W^\dagger$ from the right, $H'D'(g) = D'(g)H'$.

Now drop the primes.

In the statement of the theorem, H can be taken, not only to be hermitean, but also to be diagonal.

Now using the prior discussion about upper and lower indices.

Now take the ij -component of the statement $HD(g) = D(g)H$ (using the upper and lower indices explained earlier but suspending the repeated index summation convention for the moment): $(HD(g))_{ij} = H_{i_i} D_{ij}(g) = (D(g)H)_{ij} = D_{ij}(g) H_{j_j}$, which implies that $(H_{i_i} - H_{j_j}) D_{ij}(g) = 0$.

Note that there are many equations here, as i, j , and g run over their ranges.

We are almost there.

For a given pair i, j , unless $D_{ij}(g) = 0$ for all g (note the emphasis on “all” here), we can conclude $H_{i_i} = H_{j_j}$.

We already know that H is diagonal; now we have shown that different diagonal elements are equal.

Taking all possible i, j , we conclude that H is proportional to the identity matrix.

This proves Schur's lemma.

The irreducibility of the representation is precisely to protect us against that “unless” clause in the preceding paragraph.

Suppose that the representation reduces to a direct sum of a 3-dimensional and a 7-dimensional representation:

$$D(g) = \left(\begin{array}{c|c} D^{(3)}(g) & 0 \\ \hline 0 & D^{(7)}(g) \end{array} \right).$$

Then the element of $D(g)$ in the second row and fifth column, for example, vanishes for all g .

In the proof, we could not show that $H^2_2 = H^5_5$.

Hence we cannot conclude that H is proportional to the identity matrix, only that

$$H = \left(\begin{array}{c|c} \mu I_3 & 0 \\ \hline 0 & \nu I_7 \end{array} \right)$$

is equal to two identity matrices I_3 and I_7 stacked together, with μ, ν two arbitrary real numbers.

To repeat, the stipulation that the representation is irreducible is crucial.

Otherwise, H could have the form just shown, for example, and is assuredly not proportional to the identity matrix.

The Great Orthogonality theorem

Now we are ready for the central theorem of representation theory: Given a d-dimensional irreducible representation $D(g)$ of a finite group G , we have

$$\sum_g D^\dagger(g)^i_j D(g)^k_l = \frac{N(G)}{d} \delta^i_l \delta^k_j \quad (1)$$

with $N(G)$ the number of group elements.

The heart of the theorem is the assertion that the sum on the left hand side is proportional to $\delta^i_l \delta^k_j$, which is either 0 or 1.

Dr. Feeling saunters by, muttering, “What this tells us is that when we sum over an entire group, any ‘orientational’ information is washed out.

This is analogous to a common situation in physics: a physical result, after angles are integrated or averaged over, cannot favor any particular direction.”

Note also that we are again using the upper and lower index convention appropriate for unitary matrices.

You need not remember the proportionality constant on the right hand side; that is easily determined by setting $j = k$ and summing over all d values that the index ranges over.

The left hand side becomes $\sum_g \delta^i_l = N(G) \delta^i_l$, which fixes the constant to be $N(G)/d$.

Here is the proof of the theorem.

Form the matrix $A = \sum_g D^\dagger(g) X D(g)$ for some arbitrary matrix X . Observe that, for any g , $D^\dagger(g) A D(g) = D^\dagger(g) (\sum_{g'} D^\dagger(g') X D(g')) D(g) = (\sum_{g'} D^\dagger(g'g) X D(g'g)) = A$ because of the group axioms. (We invoked the rearrangement lemma again.) By Schur's lemma, $A = \lambda I_d$. Trace this to obtain $\text{tr } A = \lambda d = \sum_g \text{tr } D^\dagger(g) X D(g) = \sum_g \text{tr } X = N(G) \text{tr } X$, which determines $\lambda = \frac{N(G)}{d} \text{tr } X$.

Thus far, the discussion is for any X . We now choose it to be 0 everywhere except for the entry $X^j_k = 1$ in the j th row and k th column (for some specific j and k , for example, 3 and 11), which is set equal to 1. Thus $\text{tr } X = \delta^k_j$. Now let us evaluate a specific entry A^i_l in the matrix A , using what we have just learned: $A^i_l = \sum_g (D^\dagger(g) X D(g))^i_l = \sum_g D^\dagger(g)^i_j D(g)^k_l = \lambda \delta^i_l = \frac{N(G)}{d} \delta^i_l \text{tr } X = \frac{N(G)}{d} \delta^i_l \delta^k_j$, which is precisely (1).

Clearly, (1) imposes a powerful constraint on the representation matrices, a constraint that we will exploit mercilessly to determine $D(g)$.

An important corollary is the following: If r and s are two inequivalent representations, then $\sum_g D^{(r)\dagger}(g)^{ij} D^{(s)}(g)^{kl} = 0$.

One quick nonrigorous “physicist type” argument is that if the two irreducible representations are inequivalent, then the indices i, j and the indices k, l live in entirely different spaces, and there is no way to write Kronecker deltas on the right hand side.

Strictly speaking, we should have written $\sum_g D^{(r)\dagger}(g)^{\mu\nu} D^{(s)}(g)^{kl} = 0$ to emphasize this point.

Specifically, if r and s have different dimensions, then Kronecker deltas do not even make sense.

With this corollary, we can then write (1) in the more complete form

$$\sum_g D^{(r)\dagger}(g)^i_j D^{(s)}(g)^k_l = \frac{N(G)}{d_r} \delta^{rs} \delta^i_l \delta^k_j \quad (2)$$

The Kronecker δ^{rs} is equal to 1 if the two irreducible representations are the same (that is, equivalent) and 0 if not.

Behold, the Great Orthogonality theorem!

Character orthogonality

The representation matrices $D^{(r)}(g)$ are of course basis dependent: we can always make a similarity transformation $D^{(r)}(g) \rightarrow S^{-1}D^{(r)}(g)S$.

Take the trace to get rid of the basis dependence.

The character $\chi^{(r)}(c) \equiv \text{tr } D^{(r)}(g)$ depends only on the class c that g belongs to, as discussed earlier.

Indeed, (2) contains so much information that we can afford to trace out some of it.

Set $i = j$ and $k = l$ and sum.

Invoke the celebrated saying “Character is a function of class!” and obtain $\sum_g (\chi^{(r)}(g))^* \chi^{(s)}(g) = \sum_c n_c (\chi^{(r)}(c))^* \chi^{(s)}(c) = N(G) \delta^{rs}$, with n_c denoting the number of elements belonging to class c .

We have thus derived a statement about the characters:

$$\sum_c n_c (\chi^{(r)}(c))^* \chi^{(s)}(c) = N(G) \delta^{rs} \quad (3)$$

This result will turn out to be enormously useful.

We will study plenty of examples of (3) later, but for the time being note that this works nicely for Z_N .

Indeed, the orthogonality in (3) for Z_N amounts essentially to the idea behind Fourier analysis.

For Z_N , the representations are labeled by an integer k that can take on the values $0, \dots, N - 1$.

The equivalence classes are labeled by an integer j that also can take on the values $0, \dots, N - 1$.

Since the irreducible representations are 1-dimensional, the characters are trivially determined to be $\chi^{(k)}(j) = e^{i2\pi kj/N}$.

Character orthogonality (3) then says $\sum_{j=0}^{N-1} e^{-i2\pi lj/N} e^{i2\pi kj/N} = \sum_{j=0}^{N-1} e^{i2\pi(k-l)j/N} = N \delta^{lk}$, which is pretty much the basis of Fourier's brilliant idea.

Character table

For a given finite group, we can construct its “character table” displaying $\chi^{(r)}(c)$.

Along the vertical axis, we list the different equivalence classes c , and along the horizontal axis, the different irreducible representations r .

To have something definite to wrap our minds around, I will give here, without derivation(do later), the character table for A_4 (with $\omega = e^{i2\pi/3}$, as was defined earlier):

A_4	n_c	c	1	1'	1''	3
	1	I	1	1	1	3
Z_2	3	(12)(34)	1	1	1	-1
Z_3	4	(123)	1	ω	ω^*	0
Z_3	4	(132)	1	ω^*	ω	0

(4)

To the left of the vertical line, the third column lists the four different equivalence classes of A_4 , as described earlier.

Each equivalence class is identified by a “typical” member: I , (12)(34), (123), (132).

The second column lists n_c , the number of elements belonging to each class: 1, 3, 4, 4.

As always, the identity belongs to a “class of one.”

The first column indicates the subgroup generated by each of these classes.

For example, $(123)^3 = I$, and so (123) generates Z_3 .

The top row lists (to the right of the vertical line) the four different irreducible representations of A_4 (as will be derived later).

They are named by their dimensions: three of them are 1-dimensional and known as $1, 1', 1''$, and one is 3-dimensional, known as 3 .

The irreducible representation 1 is just the trivial representation, representing every element of the group by the number 1 .

Thus, the first column in the table proper (just to the right of the vertical line) is fixed trivially.

The first row (again in the table proper; in the future this will be what we mean) gives $\chi^{(r)}(I) = d_r$, the dimension of the representation, since in the representation r the identity I , as explained earlier, is always represented by the $d_r \otimes d_r$ identity matrix.

One consequence follows immediately.

Let us first define two numbers characteristic of the group G : $N(C)$ = the number of equivalence classes and $N(R)$ = the number of irreducible representations. (For A_4 , they are both 4 .)

Consider, for each s , the array of the (in general) complex numbers $(n_c)^{1/2} \chi^{(s)}(c)$ as c ranges over its $N(C)$ possible values.

Regard this array as a vector in an $N(C)$ -dimensional complex vector space.

We are told by (3) that these vectors, altogether $N(R)$ of them, are orthogonal to one another.

In other words, the four columns in the table are orthogonal to one another.

This is known as column orthogonality.

But in an $N(C)$ -dimensional complex vector space, there are at most $N(C)$ such vectors.

We have just proved that

$$N(R) \leq N(C) \tag{5}$$

The number of irreducible representations of a finite group is bounded above by the number of classes.

This answers one of the questions raised earlier.

Intuitively, you might not be surprised.

We sense that the number of classes sort of measures how complicated the group is.

We will show later in this part that the inequality (5) is in fact an equality.

Indeed, the equality holds for Z_N , which as we noted earlier, has $N(C) = N$ (“everybody in his or her own class”) and $N(R) = N$. (OK, even physicists wouldn’t call this a proof.)

We can go back to the orthogonality theorem (2) before we took the trace and play the same game.

Clearly, we should expect to get more information.

So, consider, for each triplet (s, k, l) , the array of complex numbers $D^{(s)}(g)^{k_l}$ as g ranges over its $N(G)$ possible values.

Regard this array as a vector in an $N(G)$ -dimensional complex vector space.

Since, for each s , the indices k and l each range over d_s values, there are altogether $\sum_s d_s^2$ of these vectors.

We are told by (2) that they are orthogonal to one another.

Thus, reasoning as before, we obtain

$$\sum_s d_s^2 \leq N(G) \tag{6}$$

Intuitively, just as above, you might not be surprised that there is some kind of upper bound.

For a group of a certain “size” $N(G)$, the irreducible representations can’t be “too big.” (Reducible representations, in contrast, can be as big as you care to stack them.)

Later, we will show that the inequality (6) is actually also an equality.

A test for reducibility

Suppose a dubious looking character wants to sell you a used representation, swearing that it won’t fall apart on you.

Is there any way to tell if that’s true?

the

Given a representation, how can we test whether it is reducible or not?

Suppose it is reducible.

Picture the representation matrices as consisting of a stack of irreducible representations:

$$D(g) = \begin{pmatrix} \ddots & 0 & 0 & 0 & 0 \\ 0 & D^{(r)}(g) & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & D^{(s)}(g) & 0 \\ 0 & 0 & 0 & 0 & \ddots \end{pmatrix} \quad \text{for all } g \in G \quad (7)$$

In particular, the irreducible representation r with dimension d_r can appear n_r times possibly.

Then this representation has characters given by $\chi(c) = \sum_r n_r \chi^{(r)}(c)$.

According to (3)

$$\sum_c n_c \chi^*(c) \chi(c) = \sum_c n_c \sum_{r,s} n_r n_s \chi^{*(r)}(c) \chi^{(s)}(c) = N(G) \sum_{r,s} n_r n_s \delta^{rs} = N(G) \sum_r (n_r)^2 \quad (8)$$

This is a powerful result that we will see in action repeatedly later; here let's see what it tells us.

Given a representation, all we have to do is to take some traces to compute $\chi(c)$, then a quick sum over c yields the quantity $\sum_r (n_r)^2$.

If this quantity is equal to 1, we would know that one of the n_r s is equal to 1, with all the other n_r s equal to 0. (Since n_r is the number of times the irreducible representation r appears in the representation we are given, it has to be a non-negative integer.)

The given representation $D(g)$ contains the irreducible representation r once and only once.

In other words, $D(g)$ is irreducible; in fact, it is just $D^{(r)}(g)$.

In contrast, if the quantity $\sum_r (n_r)^2$ is larger than 1, then the representation we are given is in fact reducible.

Indeed, for smallish $\sum_r (n_r)^2 > 1$, we can immediately see the dimensions of the irreducible representations it breaks apart into.

For example, suppose $\sum_r (n_r)^2 = 3$; then the only possibility is for three different irreducible representations to each occur once.

We could go further.

Take the character of the representation you are being offered and contract it against the character of the representation r :

$$\sum_c n_c \chi^{*(r)}(c) \chi(c) = \sum_c n_c \sum_s n_s \chi^{*(r)}(c) \chi^{(s)}(c) = N(G) n_r \quad (9)$$

Thus, n_r is determined.

Not only can we tell whether the used representation we are being shown will fall apart, we even know how many pieces it will fall apart into.

Notational alert: In this game, many numbers appear.

It is easy to confound them.

Let's take stock: $N(C)$ is the number of equivalence classes, while n_c is the number of members in the class c ; $N(G)$ is the number of elements in the group; $N(R)$ is the number of irreducible representations the group has.

These numbers are all properties specific to the given group.

In contrast, n_r , the number of times the representation r appears in a possibly reducible representation we are given, merely reflects on that particular representation.

It is important to sharply distinguish these different concepts.

The characters of the regular representation are ridiculously easy to compute

Cayley's theorem: any finite group G with $N(G)$ elements is a subgroup of the permutation group $S_{N(G)}$, where $S_{N(G)}$ has a defining representation with dimension $N(G)$, any finite group G has an $N(G)$ -dimensional representation, known as the regular representation.

Let us apply the result of the previous section to the regular representation of a finite group G .

Note that the matrices furnishing the regular representation are rather large, with size given by $N(G)$. (For example, for A_4 they are 12 by 12.)

But these matrices are full of zeroes, as was shown for S_4 earlier.

Dr. Feeling wanders by, muttering, “With that huge factorial size and all those zeroes, I feel that the regular representation has to be massively reducible.”

With all those zeroes, the characters for the regular representation are ridiculously easy to compute: except for the identity matrix representing the identity element I , none of the other matrices have any diagonal entries by construction.

The characters all vanish, except for the character $\chi(I)$ of the identity (which is in a class by itself, and hence a class with a membership of 1).

The character $\chi(I)$ is equal to — as is always the case for any representation — the dimension of the representation, which is $N(G)$ for the regular representation by construction.

Let us apply the reducibility test (8) to the regular representation:

$$\sum_c n_c \chi^*(c) \chi(c) = (\chi(I))^2 = N(G)^2 = N(G) \sum_r (n_r)^2.$$

Hence $\sum_r (n_r)^2 = N(G)$, which for sure is larger than 1.

We thus learned that the regular representation is reducible.

Dr. Feeling’s intuition is on the mark.

Also, for the regular representation, (9) gives $\sum_c n_c \chi^{*(r)}(c) \chi(c) = \chi^{*(r)}(I) \chi(I) = d_r N(G) = N(G) n_r$, and thus

$$n_r = d_r \quad (10)$$

This remarkable result says that in the regular representation, each irreducible representation r appears as many times as its dimension d_r .

So, the matrices in the regular representation have the form

$$D^{\text{reg}}(g) = \left(\begin{array}{c|c|c|c|c} \ddots & 0 & 0 & 0 & 0 \\ \hline 0 & D^{(r)}(g) & 0 & 0 & 0 \\ \hline 0 & 0 & \ddots & 0 & 0 \\ \hline 0 & 0 & 0 & D^{(r)}(g) & 0 \\ \hline 0 & 0 & 0 & 0 & \ddots \end{array} \right) \quad (11)$$

where the $d_r \times d_r$ matrix $D^{(r)}(g)$ appears d_r times.

Since we know that $D^{\text{reg}}(g)$ is an $N(G) \times N(G)$ matrix, we learn that

$$\sum_r d_r^2 = N(G) \quad (12)$$

The inequality (6) is actually an equality, as promised.

The character table is square

At this point, we can go on and show that $N(C) \leq N(R)$, which together with (5) implies

$$N(C) = N(R) \tag{13}$$

Since the proof of the inequality just cited is a bit involved, I will skip it.

Knowing the structure of the group practically amounts to knowing how its elements fall into classes, and in particular, $N(C)$, the number of classes.

Then we know $N(R)$, the number of irreducible representations, which tells us how many terms there are in the sum $\sum_r d_r^2 = N(G)$.

For low $N(G)$, this is such a powerful constraint that the dimensions of the irreducible representations are often uniquely fixed.

For example, for A_4 with four equivalence classes, we have $1^2 + d_1^2 + d_2^2 + d_3^2 = 12$.

The unique solution is $1^2 + 1^2 + 1^2 + 3^2 = 12$.

We conclude that there are three inequivalent 1-dimensional irreducible representations and one 3-dimensional irreducible representation.

Knowing that the character table is square, we could invite ourselves to write $U^c_s = \sqrt{\frac{n_c}{N(G)}} \chi^{(s)}(c)$.

Then column orthogonality (3) takes on the suggestive form $\sum_c (U^c_r)^* U^c_s = \sum_c (U^\dagger)^r_c U^c_s = \delta_{rs}$

(where we have invented a new notation for the Kronecker delta, as we are free to do)

In other words, if we regard the character table suitably normalized as a matrix, that matrix is unitary.

But $U^\dagger U = I$ implies $U U^\dagger = I$, and hence $U^\dagger U = I$.

But written out, this says that different rows are orthogonal: $\sum_r \chi^{(r)}(c)^* \chi^{(r)}(c') = N(G) \delta_{cc'}/n_c$.

Note also that the orthonormality of the first row reproduces (12).

Dr. Feeling wanders by again. “What was your intuition regarding how many irreducible representations a group could have?” he asks, and then mutters, “Mine was that the more complicated a group, the more irreducible representations it would have. What (13) shows is that, not only is this feeling correct, but also it is the number of classes that measures how complicated a group is.”

Character Is a Function of Class

In this part I use the results of the preceding one to construct the character table of various finite groups.

I strongly recommend that you do not sit back and read this like a novel with lots of characters; the construction of character tables is a participatory sport.

Trust me, for the mathematically inclined, constructing character tables can be quite enjoyable.

Before we start, let us summarize some of the key results we have.

Dimensions of the irreducible representations:

$$\sum_r d_r^2 = N(G) \quad (1)$$

Column orthogonality:

$$\sum_c n_c (\chi^{(r)}(c))^* \chi^{(s)}(c) = N(G) \delta^{rs} \quad (2)$$

Row orthogonality:

$$\sum_r \chi^{(r)}(c)^* \chi^{(r)}(c') = \frac{N(G)}{n_c} \delta^{cc'} \quad (3)$$

The character table is square:

$$N(C) = N(R) \quad (4)$$

These results impose powerful constraints on the character table.

For good measure, I display the Great Orthogonality theorem from which these results were derived:

$$\sum_g D^{(r)\dagger}(g)_j^i D^{(s)}(g)_l^k = \frac{N(G)}{d_r} \delta^{rs} \delta_l^i \delta_j^k \quad (5)$$

We also derived from (2) two results giving the number of times n_r that the irreducible representation r appears in a given (possibly) reducible representation:

$$\sum_c n_c \chi^*(c) \chi(c) = N(G) \sum_r n_r^2 \quad (6)$$

and

$$\sum_c n_c \chi^{*(r)}(c) \chi(c) = N(G) n_r \quad (7)$$

As you will see, these two results are much used in this chapter.

The characters of A_3

When I was a student, after I worked my way through the proofs of some powerful theorems, I was often disappointed that those theorems were then applied to some pitifully trivial examples.

Well, for pedagogical clarity, my first example here is also rather trivial.

Consider A_3 , the group of even permutations of three objects.

It contains three elements: the identity I , the clockwise permutation $c = (123)$, and the anticlockwise permutation $a = (132)$.

The group is abelian, with c and a the inverse of each other.

As was noted in earlier, in an abelian group, every element is a single member of its own equivalence class.

There are thus three equivalence classes and hence, according to (4), three irreducible representations.

Then (1) can only be satisfied by $1^2 + 1^2 + 1^2 = 3$: there can only be three 1-dimensional irreducible representations; call them 1 , $1'$, and $1''$.

Since $c^3 = I$, and since the representation matrix, being 1×1 , is just a number, c can only be represented by 1 , $\omega \equiv e^{i2\pi/3}$, or $\omega^* = \omega^2$.

These three possibilities correspond to the three 1-dimensional irreducible representations, 1 , $1'$, and $1''$.

The number representing a is determined by $ca = I$.

The character table is thus fixed to be

$A_3 \quad n_c$			1	1'	1''
	1	I	1	1	1
Z_3	1	$c = (123)$	1	ω	ω^*
Z_3	1	$a = (132)$	1	ω^*	ω

(8)

Note that column and row orthogonality are satisfied, because $1 + \omega + \omega^* = 0$, namely, the three cube roots of unity sum to 0.

Cyclic groups

Many of the considerations for A_3 apply to any abelian finite group with N elements.

The N equivalence classes imply N irreducible representations, all 1-dimensional so that they satisfy $1^2 + 1^2 + \dots + 1^2 = N$.

This certainly makes sense, since 1×1 matrices (namely, numbers) all commute.

Furthermore, we know that all irreducible representations are unitary, and hence each irreducible representation must be simply some appropriate root of unity.

In fact, since A_3 is isomorphic to Z_3 , the discussion of A_3 here is just a special case of how Z_N can be represented.

As was already mentioned, the N 1-dimensional irreducible representations of Z_N are labeled by the integer $k = 0, 1, 2, \dots, N-1$.

The group element $e^{i2\pi j/N}$ is represented by $D^{(k)}(e^{i2\pi j/N}) = e^{i2\pi kj/N}$ (so that $k = 0$ is the trivial identity representation).

As was also remarked there, for Z_N , character orthogonality (2) gives

$$\sum_{j=0}^{N-1} e^{i2\pi(k-k')j/N} = N\delta_{kk'} \quad (9)$$

which is surely one of the most important identities in mathematics, science, and engineering: the identity that motivates Fourier series.

From A_3 to S_3

It is instructive to go from A_3 to S_3 with its $3! = 6$ elements.

As we shall see, S_3 is nonabelian.

At this point, our friend Confusio comes by and offers, “This game is easy; I could do it. Let me construct the character table of S_3 for you!”

Going from A_3 to S_3 , we add the elements (12) , (23) , and (31) , namely, the three transpositions or exchanges, which are the odd permutations not included in A_3 .

They clearly form an equivalence class by themselves.

“Four equivalence classes, and hence four irreducible representations,” Confusio mutters.

Since we always have the 1-dimensional trivial representation, the dimensions (call them a , b , and c) of the three nontrivial irreducible representations must satisfy $1^2 + a^2 + b^2 + c^2 = 6$.

But $1^2 + 1^2 + 1^2 + 2^2 = 7 > 6$, and $1^2 + 1^2 + 1^2 + 1^2 = 4 < 6$. “Oops!” exclaims Confusio.

Dear reader, can you see what’s wrong before reading on?

Confusio has unwittingly made a careless error.

You and I gently point out that, since (23) is now in the group, $c = (123)$ and $a = (132)$ become equivalent, and the two distinct classes they belong to in A_3 merge.

Thus, S_3 , just like A_3 , has only three equivalence classes.

There are only three irreducible representations, not four, and now things work out nicely:
 $1^2 + 1^2 + 2^2 = 6$.

Call the irreducible representations $1, \bar{1},$ and 2 , according to their dimensions.

Let us now construct the character table.

Note that, as always, the first column and first row are automatically filled in:

S_3	n_c		1	$\bar{1}$	2
	1	I	1	1	2
Z_3	2	(123), (132)	1	1	-1_x
Z_2	3	(12), (23), (31)	1	-1	0_y

(10)

Here $\bar{1}$ denotes the signature representation: it is ± 1 , according to whether the permutation is even or odd.

This explains why it does not appear in A_3 .

Indeed, when we restrict S_3 to A_3 , $\bar{1}$ becomes 1.

The peculiar notation in the third column after the vertical line, the characters for the representation 2, is because I want to show you how to start with the column $(2, x, y)$ and use the theorems we have learned to determine x and y .

There are many ways to do it, since the various orthogonality theorems actually over-determine a simple character table like this one.

One way to start is to observe that the characters for 1 and $\bar{1}$ differ only in the last row.

This immediately implies, by subtracting two orthogonality relations, that $y=0$. (In other words, take the difference of $\sum_c n_c (\chi^{(1)}(c))^* \chi^{(2)}(c) = 0$ and $\sum_c n_c (\chi^{(\bar{1})}(c))^* \chi^{(2)}(c) = 0$.)

Next, the orthonormality (weighted by n_c) of $\chi^{(2)}(c)$ (that is, $\sum_c n_c (\chi^{(2)}(c))^* \chi^{(2)}(c) = 6$) gives

$$1 \cdot 2^2 + 2 \cdot x^2 + 3 \cdot 0^2 = 6 = 4 + 2x^2 \Rightarrow x^2 = 1, \text{ and so } x = \pm 1.$$

Choosing the $+$ sign would contradict orthogonality with 1 and $\bar{1}$.

So we are forced to choose $x = -1$, and the table is completed.

Let us double check that the solution $x = -1$ satisfies orthogonality with the identity representation:

$$1 \cdot 1 \cdot 2 + 2 \cdot 1 \cdot (-1) + 3 \cdot 1 \cdot 0 = 0, \text{ indeed.}$$

From the character table to the representation matrices

Now that we have constructed the character table, we might want to exhibit the 2-dimensional representation matrices explicitly.

In fact, the character table, particularly for smaller groups, contains enough information to determine the actual representation matrices.

The identity I is represented by the 2×2 identity matrix, of course.

To exhibit representation matrices explicitly, we have to commit to a particular basis.

Let's go to a basis in which (123) and (132) are diagonal.

Invoking the theorem that these representation matrices must be unitary and the fact that these two elements generate a Z_3 subgroup, we fix that $(123) \sim \begin{pmatrix} \omega & 0 \\ 0 & \omega^* \end{pmatrix}$ and $(132) \sim \begin{pmatrix} \omega^* & 0 \\ 0 & \omega \end{pmatrix}$. (which is which is a matter of convention.)

The traces of these matrices are equal to $\omega + \omega^* = -1$, in agreement with the character table.

Next, what about (12) , (23) , and (31) ?

Confusio comes along and ventures an educated guess: they must be represented by the three Pauli matrices $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$, and $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

Sounds very plausible: they are traceless (in accordance with the character table) and unitary.

But, Confusio, the representation matrices must satisfy the multiplication table.

Let (12) be represented by $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

Then since (12)(123) = (12)(12)(23) = (23), (23) is represented by $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \omega & 0 \\ 0 & \omega^* \end{pmatrix} = \begin{pmatrix} 0 & \omega^* \\ \omega & 0 \end{pmatrix}$.

So, Confusio guessed wrong.

The three transpositions (12), (23), and (31) are represented by $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\begin{pmatrix} 0 & \omega^* \\ \omega & 0 \end{pmatrix}$, and $\begin{pmatrix} 0 & \omega \\ \omega^* & 0 \end{pmatrix}$.

Some simple checks: they are unitary, traceless, and generate Z_2 , since, for example,

$$\begin{pmatrix} 0 & \omega^* \\ \omega & 0 \end{pmatrix} \begin{pmatrix} 0 & \omega^* \\ \omega & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

We could check the multiplication table some more, for example, (23)(31) = (231) = (123):

$$\begin{pmatrix} 0 & \omega^* \\ \omega & 0 \end{pmatrix} \begin{pmatrix} 0 & \omega \\ \omega^* & 0 \end{pmatrix} = \begin{pmatrix} \omega & 0 \\ 0 & \omega^* \end{pmatrix}, \text{ indeed.}$$

Or row orthogonality, between the first two rows, for example: $1 \cdot 1 + 1 \cdot 1 + 2 \cdot (-1) = 1 + 1 - 2 = 0$, as required.

By the very presence of the 2-dimensional irreducible representation, S_3 , unlike its subgroup A_3 , is not abelian.

The Great Orthogonality theorem is so constraining that there is usually more than one way to arrive at the same conclusion.

From (5), we note that any nontrivial irreducible representation r must satisfy

$$\sum_g D^{(r)}(g)^i_j = 0 \tag{11}$$

From this we can see that Confusio's guess cannot be correct.

We have $D^{(2)}(I) + D^{(2)}(123) + D^{(2)}(132)$ and $D^{(2)}(12) + D^{(2)}(23) + D^{(2)}(31)$ separately summing to 0, so that (11) holds a fortiori, but the three Pauli matrices do not sum to 0.

Note that for the correct set of representation matrices, the even and odd permutations must separately sum to 0, because of orthogonality with the $\bar{1}$ as well as with the 1 representation.

Link between group theory and geometry: Fixed points

You may recall that S_3 is also the invariance group of the equilateral triangle.

The two elements (123) and (132) correspond to rotations through $2\pi/3$ and $4\pi/3$, respectively, and the three transpositions (12), (23), and (31) to reflections across the three medians.

If we label the three vertices of the triangle by a, b, c or $1, 2, 3$ they can be thought of as the objects being permuted by S_3 .

This remark provides a nice link between group theory and geometry.

Confusio asks, “You mean you paint 1, 2, 3 on the three vertices? Then the vertices would be distinguishable, and rotation through $2\pi/3$ would not leave the triangle invariant.”

No, Confusio.

The labeling is just to help us keep track of which vertex we are talking about.

The three vertices are to be treated as identical.

The triangle is meant to be a mathematical triangle in your mind's eye, not a physical triangle.

For instance, suppose you were to draw a triangle on a piece of lined paper which is blank on the other side.

Then an exchange, (12) say, would flip the piece of paper over, and you could tell that this is not quite the same triangle as before.

Earlier, I remarked that S_n has an n -dimensional representation, its defining or fundamental representation.

Since we surely do not have room for a 3-dimensional representation for S_3 , this defining representation, which we shall refer to as 3, must be reducible.

Now the orthogonality theorem leaps into action again.

Let's start by writing down the 3 characters for the 3:

3
0
1

Think of this as an extra column one could attach to the character table in (10).

The top entry is of course 3: the character of the identity is always just the dimension of the representation.

To understand the bottom entry, note that in the 3, an element like (12) is represented by $\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

The third basis vector is untouched: it's a "fixed point," to borrow the language of topological maps.

Thus, the trace (namely, the character) is just 1.

In contrast, (123) is represented by $\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$, which has no fixed point and hence has a null character

OK, any guesses as to what irreducible representations this 3 reduces to?

It could be $3 \rightarrow 1 + \bar{1} + \bar{1}$, or $3 \rightarrow 2 + 1$, and so on.

For a simple example like this, there are only so many possibilities.

We simply plug in (7).

Orthogonality of characters between 2 and 3 gives $1 \cdot 2 \cdot 3 + 2 \cdot (-1) \cdot 0 + 3 \cdot 0 \cdot 1 = 6 = 1(6)$, and between 1 and 3 gives $1 \cdot 1 \cdot 3 + 2 \cdot 1 \cdot 0 + 3 \cdot 1 \cdot 1 = 6 = 1(6)$.

Thus, 3 contains 2 once and 1 once; in other words, $3 \rightarrow 2 + 1$.

Indeed, the characters add up correctly: $\begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$

The reader might have recognized that this is the same problem as decomposing a vector that usually starts a review of linear algebra: let

$$\begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix} = x \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + y \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} + z \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix},$$

solve for x, y, z . (For ease of writing we use the “vector” notation instead of the squareboxes.)

Write this as a matrix equation $C \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix}$, where we regard the character table of S_3 in (10) as a matrix

$$C = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 1 & -1 \\ 1 & -1 & 0 \end{pmatrix}.$$

Use your knowledge of linear algebra to find C^{-1} and verify that $C^{-1} \begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix}$ gives the solution listed above.

For groups with large numbers of equivalence classes, this procedure could then be performed by a machine.

The class of inverses: Real and complex characters

Given the class c , we defined the class $\bar{c}$ to consist of the inverse of all elements belonging to c .

Since $D(g)$ is unitary, we have $D(g^{-1}) = (D(g))^{-1} = D(g)^\dagger$.

Tracing, we obtain

$$\chi(\bar{c}) = \chi(c)^* \tag{12}$$

One consequence of this is that if a class contains the inverses of its members, then $\bar{c} = c$, and $\chi(c)$ is real.

For example, the characters of S_3 are all real, but not those of its subgroup A_3 .

Various approaches for determining the character table

We might summarize the discussion above by enumerating the various approaches for determining the character table, at least in the case of a simple example like A_3 .

1. Method of algebraic insight

Recognize that A_3 is the same as Z_3 , and look up the Fourier representations of Z_3 .

$$\begin{array}{ccc} 1 & 1 & 1 \\ 1 & \omega & \omega^* \\ 1 & \omega^* & \omega \end{array} \tag{13}$$

(Here I am displaying only the heart of the character table.)

2. Method of geometrical insight

We realize that the three objects being permuted by A_3 can be thought of as the three vertices of an equilateral triangle. It follows that A_3 has a 2-dimensional representation consisting of $\{I, R, R^2\}$, with $R = \begin{pmatrix} c & s \\ -s & c \end{pmatrix}$, $c \equiv \cos(2\pi/3)$, and $s \equiv \sin(2\pi/3)$. But since the group is abelian, we know that there is no irreducible 2-dimensional representation, so it must be reducible.³ Diagonalize R . Since $\det R = 1$, the diagonalized form of R must have the form $\begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix}$. But $R^3 = I$; hence $e^{i\theta} = e^{2\pi i/3} = \omega$. (Indeed, by inspection we know that R is diagonalized by $S^{-1}RS$, with $S = \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix} / \sqrt{2}$.)

3. Method of exploiting the orthogonality theorems

$$\begin{array}{ccc} 1 & 1 & 1 \\ 1 & x & u \\ 1 & y & v \end{array} \tag{14}$$

Various orthogonality constraints, for example, $1 + x + y = 0$, $1 + |x|^2 + |y|^2 = 3 \implies |x| = |y| = 1$, and so on, quickly determine $x = v = \omega$, $y = u = \omega^*$.

The invariance group T of the tetrahedron and A_4

Plato particularly liked the tetrahedron and associated it with fire.

Let's find the characters of its invariance group, known as T .

We proceed using a mix of geometrical insight, appeal to the orthogonality theorems, and informed guessing.

By now, it doesn't take much brilliance to realize that if we mentally label the four vertices of the tetrahedron by 1, 2, 3, and 4, we see that T is isomorphic to A_4 , with $4!/2 = 12$ elements.

Recall that we already listed the equivalence classes of A_4 .

There are four of them, with representative members I , $(12)(34)$, (123) , and (132) , respectively.

Thus, there should be four irreducible representations.

The isomorphism with T immediately tells us that A_4 has a 3-dimensional representation.

Indeed, $T = A_4$, being the invariance group of a geometrical object that lives in 3-dimensional Euclidean space, is necessarily a subgroup of the rotation group and hence represented by 3×3 matrices.

Indeed, the orthogonality theorems assert that the dimensions of the irreducible representations satisfy $1^2 + d_1^2 + d_2^2 + d_3^2 = 12$.

Since $1^2 + 1^2 + 2^2 + 2^2 = 10 < 12$, $1^2 + 2^2 + 2^2 + 2^2 = 13 > 12$, and so on, with the only possible solution $1^2 + 1^2 + 1^2 + 3^2 = 12$, we can proclaim (even if we are not genetically endowed with a modicum of geometrical insight) that there must be a 3-dimensional representation.

We also learned that A_4 has two other 1-dimensional representations, traditionally known as $1'$ and $1''$, besides the trivial representation 1 .

Playing around a bit† with the orthogonality theorems, we soon manage to construct the character table

A_4	n_c		1	$1'$	$1''$	3
	1	I	1	1	1	3
Z_2	3	(12)(34)	1	1	1	−1
Z_3	4	(123)	1	ω	ω^*	0
Z_3	4	(132)	1	ω^*	ω	0

(15)

As always, we indicate for each class the number of members, a typical member, and the cyclic subgroup a member of that class would generate.

The first column (all 1s) and the first row (just the dimensions) are immediately fixed.

The two 1-dimensional representations $1'$ and $1''$ are easily fixed by virtue of their 1-dimensionality, unitarity, and their having to represent elements generating Z_3 .

Column and row orthogonality are of course so restrictive as to practically determine everything.

For example, the two bottom entries in the column for the irreducible representation 3 follow immediately from the orthogonality between 3 and 1, and between 1' and 1''.

You might think that the fourth, rather than the cube, root of unity would appear in the character table for A_4 .

One way of seeing the appearance of ω and ω^* is by restricting A_4 to A_3 .

We erase the second row (corresponding to (12)(34)) and the fourth column, get rid of a few entries, and necessarily should recover the known character table for A_3 .

Beginners are also sometimes surprised that A_4 does not have a 4-dimensional irreducible representation.

Exhibiting the 3 of A_4

Once again, we are seized by ambition and would like to determine the actual matrices in the 3-dimensional irreducible representation.

A tedious but straightforward way would be to center the tetrahedron and work out the rotation matrices that leave it unchanged.

Instead, we proceed by inspired and informed guessing.

Actually, the character table contains enough hints so that hardly any guesswork is needed.

First, note that the three elements (12)(34), (13)(24), and (14)(23) commute with one another.

Thus, the matrices representing them can be simultaneously diagonalized, according to a theorem of linear algebra.

Let's go to a basis in which they are diagonal.

The matrices each generate a Z_2 , and thus their diagonal entries can only be ± 1 .

Since A_4 consists of even permutations, these matrices have determinant = +1.

They can only be

$$r_1 \equiv \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad r_2 \equiv \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad r_3 \equiv \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (16)$$

That there are precisely three such matrices and that they have trace = -1, agreeing with the entry of -1 in the character table, provide strong evidence that we are on the right track.

What about (123)?

Since it generates Z_3 , it is clearly represented by $c \equiv \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$, which cyclically permutes the three basis vectors in the clockwise sense.

What about the other three guys in the same equivalence class as c ?

Well, they are $r_1 c r_1$, $r_2 c r_2$, and $r_3 c r_3$.

As a check, all six of these matrices have zero trace, in agreement with the character table.

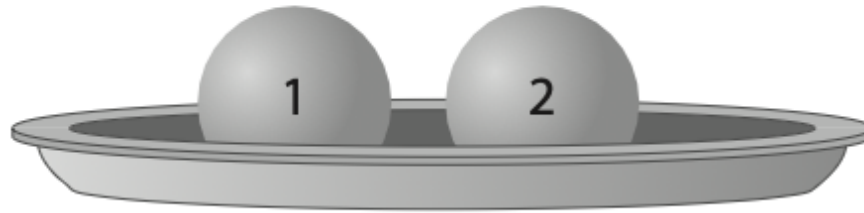


Figure 1

The tray method

For more practice with representations, I show you one way of constructing a larger (and hence reducible) representation of A_4 .

To be concrete, let us think of the four objects permuted by A_4 as four balls labeled 1, 2, 3, and 4.

Denote by $\langle 12 \rangle$ a tray into which we have put a 1-ball and a 2-ball (using an obvious terminology).

With the rule of not allowing the two balls in each tray to be of the same type, we can have $6 = 4 \cdot 3/2$ different types of trays, namely $\langle 12 \rangle$, $\langle 13 \rangle$, $\langle 14 \rangle$, $\langle 23 \rangle$, $\langle 24 \rangle$, and $\langle 34 \rangle$.

This furnishes a 6-dimensional representation of A_4 , each element of which takes a tray into some other type of tray.

See figure 1.

The characters of this representation, referred to as 6, are easy to write down: $\chi = 6$ for I as always (namely, the dimension of the representation); $\chi = 0$ for the class typified by (123), which has no fixed point (so that the 6-by-6 matrix representing (123) vanishes along the diagonal); similarly, $\chi = 0$ for the class typified by (132); and finally, $\chi = 2$ for the class typified by (12)(34), which leaves the tray <12> and the tray <34> unchanged (so that the 6-by-6 matrix representing (12)(34) has two 1s along the diagonal).

Thus, the characters for the 6 are given by $\begin{pmatrix} 6 \\ 2 \\ 0 \\ 0 \end{pmatrix}$.

First check: $1 \cdot 6^2 + 3 \cdot 2^2 + 0 + 0 = 36 + 12 = 48 = 4(12)$.

So the 6 reduces into four irreducible representations.

Orthogonality with 1: $1 \cdot 1 \cdot 6 + 3 \cdot 1 \cdot 2 + 0 + 0 = 6 + 6 = 12$, so 6 contains 1 once.

Similarly, 6 contains $1'$ once, and $1''$ once.

Orthogonality with 3: $1 \cdot 3 \cdot 6 + 3 \cdot (\mp 1) \cdot 2 + 0 + 0 = 18 - 6 = 12$, so 6 also contains 3 once.

We conclude that $6 \rightarrow 1 + 1' + 1'' + 3$ (and of course $6 = 1 + 1 + 1 + 3$ provides another check).

Indeed, $\begin{pmatrix} 6 \\ 2 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ \omega \\ \omega^* \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ \omega^* \\ \omega \end{pmatrix} + \begin{pmatrix} 3 \\ -1 \\ 0 \\ 0 \end{pmatrix}$.

Note that $1'$ and $1''$ are complex conjugate of each other, and thus they must occur in 6 with equal weight, since the representation matrices of 6 are manifestly real.

A geometrical fact

Before moving on, I mention one fact that seems strange at first sight.

Consider the element $(12)(34)$: can we really simultaneously exchange vertices 1 and 2, and vertices 3 and 4, of a solid tetrahedron by rotation?

Group theory assures us that yes, it is possible.

Denote by m_{12} the midpoint on the edge joining 1 and 2, and by m_{34} the midpoint on the edge joining 3 and 4, as shown in figure 2.

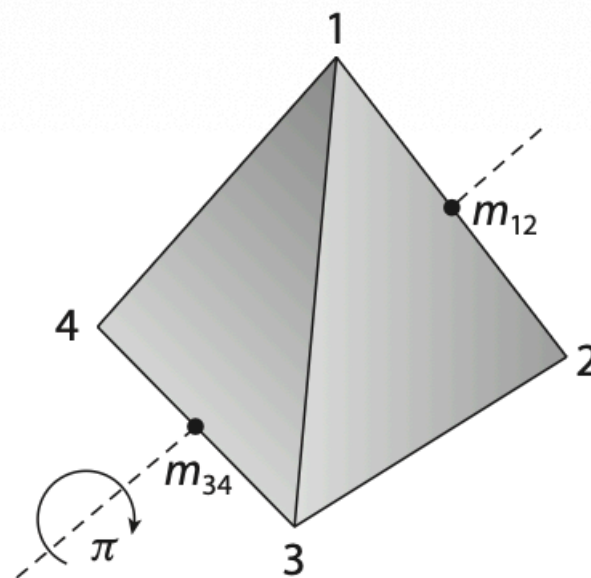


Figure 2

Imagine a line joining m_{12} and m_{34} .

Rotate the tetrahedron around this line through 180° .

You should be able to see that this exchanges vertices 1 and 2, and vertices 3 and 4.

Since there are three such lines joining midpoints to opposite midpoints, there are three 180° rotations, one around each line, corresponding to $(12)(34)$, $(13)(24)$, and $(14)(23)$.

Another way of seeing this is to first rotate the base of the tetrahedron through $2\pi/3$, then rotate again through $2\pi/3$ around a vertex on the base, as shown in figure 3.

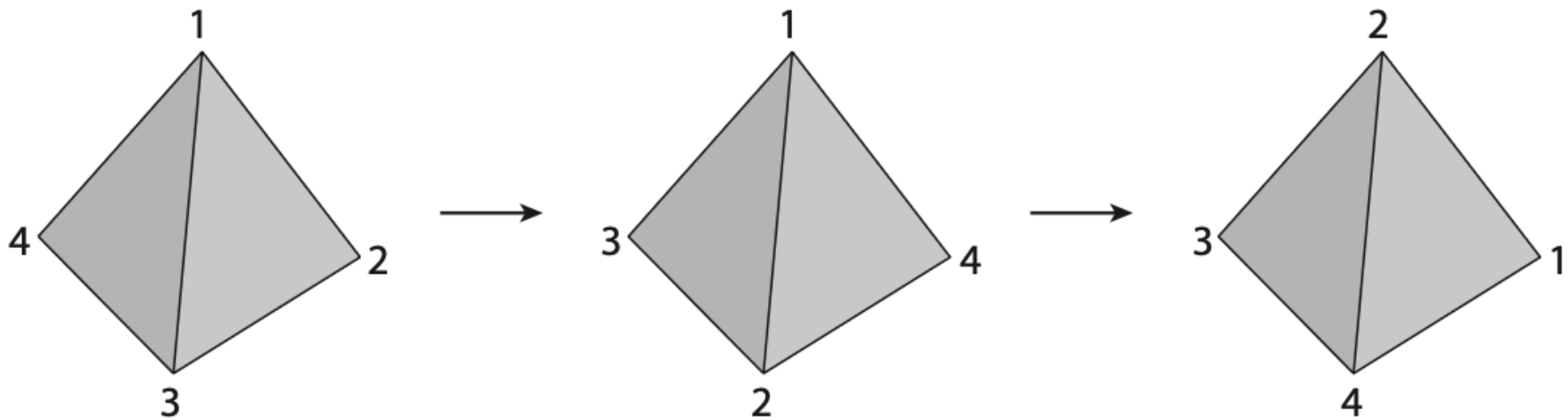


Figure 3

The net result is indeed the simultaneous exchanges of vertices 1 and 2 and of vertices 3 and 4.

Perhaps you realize that this is just $(12)(34) = (12)(23)(23)(34) = (123)(234)$ (recall the rules you learned earlier).

Fun with character tables

the the the the the the

It is instructive to go from A_4 to S_4 .

Going from four to five equivalence classes (with typical members I , $(12)(34)$, (123) , (12) , and (1234) , respectively), we now should have five irreducible representations, and the only solution for their dimensions is $1^2 + 1^2 + 2^2 + 3^2 + 3^2 = 24$.

Next, we determine the number of members in each equivalence class, namely, 1, 3, 8, 6, and 6 which adds up to 24, of course.

Note that as remarked earlier, upon $S_4 \rightarrow A_4$, the equivalence class to which (123) belongs splits up into two equivalence classes: $8 \rightarrow 4 + 4$.

I leave it to you to derive the character table. Here it is:

S_4	n_c		1	$\bar{1}$	2	3	$\bar{3}$
	1	I	1	1	2	3	3
Z_2	3	$(12)(34)$	1	1	2	-1	-1
Z_3	8	(123)	1	1	-1	0	0
Z_2	6	(12)	1	-1	0	1	-1
Z_4	6	(1234)	1	-1	0	-1	1

(17)

The last two rows pertain to the $6 + 6 = 12$ elements that do not belong to A_4 .

Perhaps the appearance of a 2-dimensional irreducible representation is a bit of surprise, now that we understand the presence of the 3-dimensional irreducible representations.

On restriction to A_4 , 2 breaks into $1' + 1''$.

By now, you should be able to construct the character table of just about any finite group that comes your way.

Personally, I rather enjoy constructing the character tables of (small) finite groups.

You might want to try a few.

Let us look at the invariance group of the square, called D_4 .

As was already mentioned, it consists of rotations R around its center, reflections r across the two medians joining the center of the square to the midpoints of its sides, and reflections d across the two diagonals.

See figure 4.

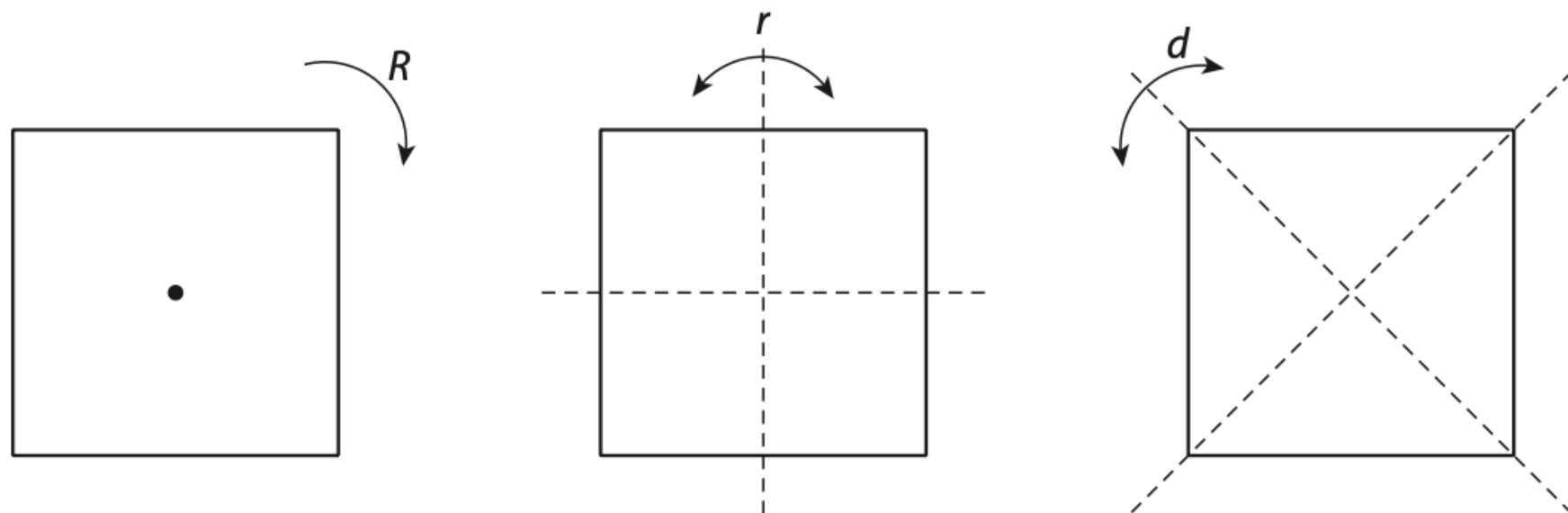


Figure 4

The corresponding 2-by-2 transformation matrices are

$$\begin{aligned} I &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ R &= \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad R^2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = -I, \quad R^3 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \\ r_x &= \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \quad r_y = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \\ d_1 &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad d_2 = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} \end{aligned} \tag{18}$$

I would like to pause and clear up a potential source of confusion.

In a group, minus the identity (that is, $-I$) does not exist and has no more meaning than, say, $3g$ or $-g$ for a group element g .

Groups are defined in the abstract by some composition law, which in everyday usage is often referred to as multiplication.

It does not make sense, in general, to multiply a group element g by some number.

Here, R represents a rotation through 90° , and R^2 , a rotation through 180° , happens to be represented by a matrix that is numerically (-1) times the identity matrix.

If you like, you can think of $-I$ as a convenient symbol for $\mathbb{R}^2$.

Or we should simply use some other symbol and not introduce the symbol $-I$ at all.

Similarly, it may be confusing for some texts to refer to r_y as $-r_x$.

Incidentally, to show that composition of group elements is noncommutative, the rotational group $SO(3)$ offers the canonical example, as was given earlier, but D_4 provides an even simpler example.

Under $r_y R$, the northwest corner of the square goes to the southeast, but under $R r_y$, the northwest stays put.

See figure 5.

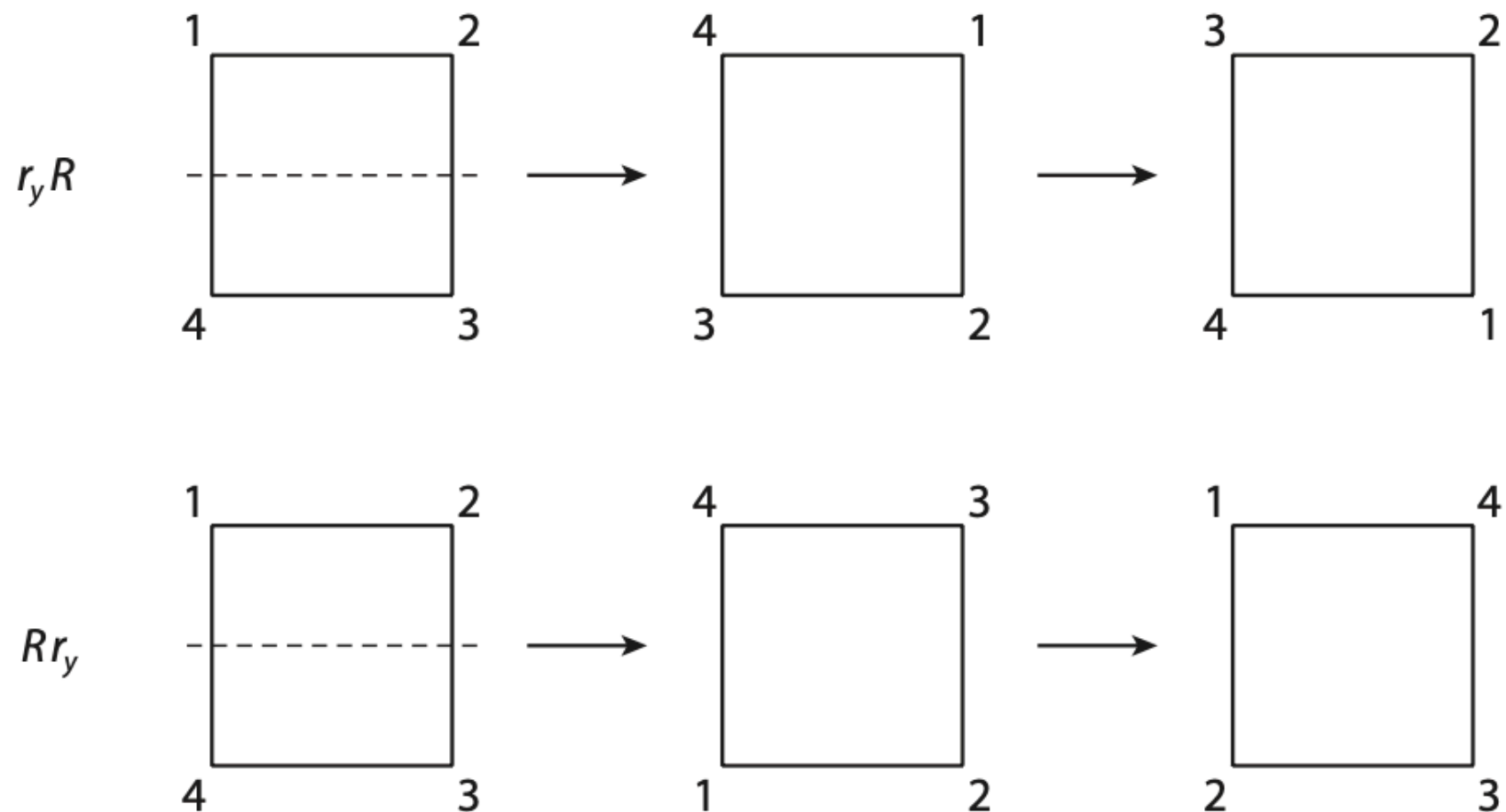


Figure 5

The eight elements of D separate into five equivalence classes: $\{I\}$, $\{-I\}$, $\{R, R^3\}$, $\{r_x, r_y\}$, and $\{d_1, d_2\}$. (Verify this!)

Thus, there are five irreducible representations.

Furthermore, $8 = 1^2 + 1^2 + 1^2 + 1^2 + 2^2$ is the only way to decompose 8.

We already know about the 2-dimensional representation, the defining representation.

Count them: four 1-dimensional representations!

Thus, we immediately have most of the character table for D_4 filled in already:

D_4	n_c		1	1'	1''	1'''	2
	1	I	1	1	1	1	2
Z_2	1	$-I$	1	1	1	1	-2
Z_4	2	R, R^3	1	a	d	g	0
Z_2	2	r_x, r_y	1	b	e	h	0
Z_2	2	d_1, d_2	1	c	f	i	0

(19)

Row orthogonality between first and second rows: $1^2 + 1^2 + 1^2 + 1^2 + 2 \cdot (-2) = 0$, check.

Column orthogonality between first and fifth columns: $1 \cdot 2 + 1 \cdot (-2) + 0 + 0 + 0 = 0$, check.

I leave it for you to fill in the entries purposely left undetermined in the table.

Incidentally, we can present this group as

$$D_4 : \langle R, r | R^4 = r^2 = I \rangle \tag{20}$$

Thus, the eight elements are $\{I, R, R^2, R^3, r, rR, rR^2, rR^3\}$.

The group D_4 is a particular example of D_n , known as the dihedral group and defined as the invariance group of the n -sided regular polygon.

It so happens that $D_3 = S_3$, but in general D_n is not isomorphic to some permutation group.

(Just count the number of elements: D_4 has 8 elements, but S_4 already has 24 elements. Also, number the vertices of the square by 1, 2, 3, and 4. No way that (12) can be effected by a rotation or a reflection.)

The quaternionic group

Earlier, I introduced the quaternionic group Q with the eight elements 1, -1, i , $-i$, j , $-j$, k , and $-k$, with the multiplication table determined by the rules $i^2 = j^2 = k^2 = -1$ and $ij = -ji = k$, $jk = -kj = i$, and $ki = -ik = j$.

You should construct the character table before reading on.

As always, 1 and -1, since they commute with everybody else, form two exclusive equivalence classes with a single member each.

Similarly, j and $-j$.

And similarly, k and $-k$.

To check this explicitly, compute $(-j)ij = -j k = -i$, and so indeed, $i \sim -i$.

We can then cyclically permute this computation.

Thus, there are five equivalence classes altogether, which means that there are five irreducible representations.

The dimensions of the four nontrivial irreducible representations are determined by $1^2 + a^2 + b^2 + c^2 + d^2 = 8$, which has the unique solution $a = b = c = 1$ and $d = 2$.

Again, I leave it to you to fill in the rest:

Q	n_c		1	1'	1''	1'''	2
	1	1	1	1	1	1	2
Z_2	1	-1	1	1	1	1	-2
Z_4	2	$i, -i$	1	1	-1	-1	0
Z_4	2	$j, -j$	1	-1	1	-1	0
Z_4	2	$k, -k$	1	-1	-1	1	0

(21)

Here is a sketch of how the construction might proceed.

After filling in the first column and the first row automatically, we proceed to the second row.

For the 1-dimensional representations $1'$, $1''$, $1'''$ the entries can only be ± 1 , since the corresponding cyclic group is Z_2 .

Row orthonormality (3) can only work if the entry in the last column is ± 2 .

Row orthogonality (3) between the first two rows can then only work with the entries shown.

Column orthonormality (2) for the last column $2^2 + (-2)^2 + |a|^2 + |b|^2 + |c|^2 = 8$ implies the unique solution $a = b = c = 0$. (The notation here is self-explanatory.)

Row orthonormality (3) of the third row $1^2 + |x|^2 + |y|^2 + |z|^2 + 0^2 = 8/2 = 4$ only works if $|x| = |y| = |z| = 1$.

The cyclic group associated with the third row also requires $x^4 = y^4 = z^4 = 1$.

Row orthogonality (3) between the first and third row then fixes one of the x , y , z to be 1, with the other two equal to -1. (It doesn't matter which is which, since $1'$, $1''$, and $1'''$ are labeled arbitrarily.)

Democracy between i , j , and k then fixes the pattern shown.

Did you notice that the character tables for Q and for D_4 are the same?

This answers the question about whether a character table uniquely characterizes a group.

3. Group Theory in a Quantum World

Group theory has a somewhat limited impact on classical mechanics, but it really blossomed with the coming of the quantum era.

Wigner is rightfully honored for introducing group theory into quantum physics.

You will learn that symmetry, degeneracy, and representation theory are beautifully intertwined.

Modern physics is almost unthinkable without the Lagrangian and the Hamiltonian, much zippier than the equations of motion.

Group theory rules all of them.

Quantum Mechanics and Group Theory: Parity, Bloch's Theorem, and the Brillouin Zone

Quantum mechanics is linear

In my study of physics (and of a bit of mathematics), I am often astonished by how the collective mind of the physicists and the collective mind of the mathematicians would converge.

Indeed, Eugene Wigner, who won the Nobel Prize for introducing group theory into quantum mechanics, wrote an influential essay on the unreasonable effectiveness of mathematics in physics.

He described mathematics as “a wonderful gift we neither understand nor deserve.”

In physics, representation theory came alive with the advent of quantum mechanics. ^

In quantum mechanics, the states of a physical system, be it a molecule or a field or a string, are described by a wave function Ψ , evolving in time according to Schrodinger equation

$$i\hbar \frac{\partial}{\partial t} \Psi = H \Psi \quad (1)$$

Here $\hbar$ denotes Planck's constant, which we already encountered.

Writing $\Psi(t) = \psi e^{-i \mathcal{E} t / \hbar}$, where ψ depends on the coordinates but not time, we obtain the eigenvalue equation

$$H \psi = \mathcal{E} \psi \quad (2)$$

At the most elementary level, the Hamiltonian H is given by a linear differential operator, for instance, for a particle moving in a potential in 1-dimension, $H = -\frac{1}{2m} \frac{d^2}{dx^2} + V(x)$.

In other situations, for a particle with spin in a magnetic field $\vec{B}$, for example, H is given by $H = \vec{B} \cdot \vec{S}$, with $\vec{S}$ the spin operator.

At the physicist's level of rigor, and for the purposes here, we can simply think of any linear operator, including H , as a matrix, albeit often an infinite dimensional one.

Degeneracy was a mystery

The Hamiltonian H has eigenfunctions ψ^α with eigenvalues $\mathcal{E}^\alpha$ and with the index α possibly taking on an infinite number of values:

$$H\psi^\alpha = \mathcal{E}^\alpha \psi^\alpha \tag{3}$$

(Here we temporarily suspend the repeated index summation convention.)

In general, $\mathcal{E}^\alpha \neq \mathcal{E}^\beta$ for $\alpha \neq \beta$.

But often, among all these eigenfunctions ψ^α , there may exist a subset (call them ψ^a) with $a = 1, \dots, d$, having exactly the same eigenvalue, call it E :

$$H\psi^a = E\psi^a, \quad \text{with } a = 1, \dots, d \tag{4}$$

Assume that these eigenfunctions are finite in number, so that $d < \infty$.

The set of eigenfunctions ψ^a is said to exhibit a d -fold degeneracy.

The important point is that E does not depend on a .

In the early days of quantum physics, the existence of degeneracy was enormously puzzling.

When you diagonalize a matrix, the eigenvalues will in general be distinct.

It became clear that there must be a reason behind the degeneracy. (Also, do not forget the heroic experimental task of unraveling the degeneracy, given that in atomic spectroscopy, experimentalists could only measure the energy difference $|E^\alpha - E^\beta|$, so that degeneracy resulted in many lines superposed on top of one another. The situation was only clarified after tremendous experimental efforts.)

Symmetry implies degeneracy

Eventually, it was realized that degeneracy has a natural explanation.

In quantum mechanics, transformations are realized as unitary operators T .

Suppose a set of transformations leave H invariant, so that

$$T^\dagger H T = H \tag{5}$$

Then, as explained earlier, these transformations form a group G : if the transformations T and T' each leave H invariant, then the transformation $T \cdot T'$, which consists of the transformation T' followed by the transformation T , clearly leaves H invariant.

The T s form a symmetry group.

For example, G could be the rotation group $SO(3)$.

Since T is unitary, $T^\dagger = T^{-1}$, we can write (5) as

$$H T = T H \tag{6}$$

Given that T leaves H unchanged, its action on ψ^a produces an eigenstate of H with the same energy E .

More explicitly, if $H \psi^a = E \psi^a$, then, according to (6),

$$H(T\psi^a) = HT\psi^a = TH\psi^a = TE\psi^a = E(T\psi^a) \quad (7)$$

Since $\psi'^a = T\psi^a$ is an eigenstate of H with energy E , it has to be a linear combination of the ψ^a s:
 $\psi^a \rightarrow \psi'^a = (D(T))^{ab}\psi^b$.

If the transformation T_1 is followed by the transformation T_2 , so that we have effectively the transformation T_2T_1 , then $D(T_2T_1) = D(T_2)D(T_1)$.

The crucial fact about quantum mechanics is of course that, unlike classical mechanics, it is linear, and linear superpositions of states are also acceptable states.

Consider the linear vector space spanned by the d degenerate eigenstates, that is, the set of all linear combinations of the ψ^a s, namely $\sum_a c_a \psi_a$ for c_a a bunch of numbers.

On this space, the Hamiltonian H is just a d -by- d identity matrix multiplied by E .

From degeneracy to group representations

Given this entire setup, you can now see that the d degenerate eigenstates ψ^a furnish a d -dimensional irreducible representation of the group G .

If we know, or could guess, what G is, this also determines what the possible values of d are, since group theory fixes the possible irreducible representations of G .

This turns out to be of great importance in the development of quantum physics—especially of particle physics.

Conversely, given experimental information on the degeneracy, we can turn the logic around and try to determine what G might be.

In particular, G has to have at least one d -dimensional irreducible representation.

Thus, in quantum physics

$G \implies \text{degeneracy} \quad \text{and} \quad G \longleftarrow \text{degeneracy}$
$d = \text{degrees of degeneracy} = \text{dimension of irreducible representation}$

The word “degenerate” has a rather negative connotation in everyday usage, but in the history of quantum physics it is regarded very favorably, as the condition can shed some light on the underlying symmetry group.

For example, among the subnuclear particles discovered in the 1950s were eight baryons (“heavy particles,” that is, heavy compared to the electron; these include the proton and the neutron) approximately equal in mass.

Particle physicists raced each other to find a group with an 8-dimensional irreducible representation (more later).

Connection to Schur’s lemma: What group theory can and cannot tell you

The reader can now see that the derivation of Schur’s lemma earlier parallels exactly the discussion here.

The H that commutes with all the $d \times d$ matrices $D(g)$ in an irreducible representation of a group G corresponds to the Hamiltonian here.

Schur tells us that in the d -dimensional subspace, the Hamiltonian H is simply equal to some constant times the identity, which is precisely what we mean when we say that d energy levels are degenerate.

Dr. Feeling wanders by. “In hindsight, it’s almost obvious. If a bunch of states get transformed into one another by some symmetry group, which by definition leaves the Hamiltonian unchanged, you would think that these states all have the same energy. In a sense, the Hamiltonian cannot tell these states apart.”

Indeed, Schur’s lemma is even more powerful if a collection of states form a reducible representation of the symmetry group.

In that case, in some basis, the representation matrices are block diagonal:

$$D(g) = \left(\begin{array}{c|c|c|c} \ddots & 0 & 0 & 0 \\ \hline 0 & D^{(r)}(g) & 0 & 0 \\ \hline 0 & 0 & D^{(s)}(g) & 0 \\ \hline 0 & 0 & 0 & \ddots \end{array} \right) \quad \text{for all } g \in G \tag{8}$$

Then we know that the Hamiltonian has the form (as discussed earlier)

$$H = \begin{pmatrix} \ddots & 0 & 0 & 0 \\ 0 & E^{(r)} I & 0 & 0 \\ 0 & 0 & E^{(s)} I & 0 \\ 0 & 0 & 0 & \ddots \end{pmatrix} \quad (9)$$

In many cases, this enables us to diagonalize the Hamiltonian with almost no work.

We know that the energy levels break up into a collection of degenerate states ($d^{(r)}$ states with energy $E^{(r)}$, $d^{(s)}$ states with energy $E^{(s)}$, and so on), with group theory telling us what the numbers $d^{(r)}$ and $d^{(s)}$ are.

But of course group theory cannot tell you what the $E^{(r)}$ s are.

Think of it this way.

Change the Hamiltonian without changing the symmetry group.

For example, take the Schrodinger equation with a central force potential $V(r)$.

Imagine changing $V(r)$ smoothly.

The energy levels would move up and down, but as long as V remains a function of r , rather than of x , y , and z separately, then the degeneracy in the energy spectrum due to rotational symmetry cannot change.

Parity

After this very brief set up, our first example is almost laughably simple, involving a laughably simple group.

Consider a particle moving in a potential in one dimension and suppose $V(x) = V(-x)$.

Then the Hamiltonian $H = -\frac{1}{2m} \frac{d^2}{dx^2} + V(x)$ is left invariant by the symmetry group Z_2 , consisting of the identity and the reflection $r : x \rightarrow -x$.

Given an eigenfunction $\psi(x)$, the reflection acting on it gives $r\psi(x) = \psi(-x)$, which, according to the discussion above, has the same energy as $\psi(x)$.

Group theory tells us that Z_2 has only two irreducible representations, in one of which $r = 1$, and in the other $r = -1$ (more precisely, r is represented by 1 and by -1 in the two representations, respectively).

In one case, $\psi(x) = \psi(-x)$, and in the other, $\psi(x) = -\psi(-x)$.

In conclusion, if $V(x) = V(-x)$, then the solutions of the Schrodinger equation may be chosen to be either even or odd, something you probably knew from day one of your exposure to quantum mechanics.

Here we have a classic example of using a sledgehammer to crack open a peanut.

Bloch's theorem and Brillouin zone

Our next application, significantly less trivial, is of fundamental importance to solid state physics.

Again, a particle moves in a potential $V(x)$ in one dimension, but now suppose $V(x + a) = V(x)$ is periodic but otherwise unspecified.

The symmetry group consists of the elements $\{ \dots, T^{-1}, I, T, T^2, \dots \}$, where T denotes translation by the lattice spacing a : $x \rightarrow Tx = x + a$.

Since the group is abelian, it can only have 1-dimensional irreducible representations.

Hence $\psi'(x) \equiv T\psi(x) = \zeta\psi(x)$, so that T is represented by a complex number ζ .

But since wave functions are normalized, $\int dx |\psi'(x)|^2 = \int dx |\psi(x)|^2$, we have $|\zeta| = 1$.

Thus, we can write $\zeta = e^{ika}$ with k real (and with dimension of an inverse length).

Since $e^{i(ka+2\pi)} = e^{ika}$, we may restrict k to the range

$$-\frac{\pi}{a} \leq k \leq \frac{\pi}{a} \tag{10}$$

This range is the simplest example of a Brillouin zone.

For a 1-dimensional lattice consisting of N sites, we impose periodic boundary conditions and thus require $T^N = I$.

The symmetry group is just the familiar cyclic group Z_N .

The condition $e^{iNka} = 1$ thus implies that $k = (2\pi/Na)j$ with j an integer.

For N macroscopically large, the separation Δk between neighboring values of j , of order $2\pi/Na$, is infinitesimal, and so we might as well treat k as a continuous variable ranging from $-\pi/a$ to π/a .

It is convenient and conventional to write

$$\psi(x) = e^{ikx} u(x) \tag{11}$$

with $u(x + a) = u(x)$.

This statement is known as Bloch's theorem. (Of course, any $\psi(x)$ can be written in the form (11); the real content is the condition $u(x + a) = u(x)$.)

Note that this is a general statement completely independent of the detailed form of $V(x)$.

Given a specific $V(x)$, the procedure would be to plug (11) into the Schrodinger equation and solve for $u(x)$ for the allowed energy eigenvalues, which of course would depend on k and thus can be written as $E_n(k)$.

As k ranges over its allowed range, $E_n(k)$ would vary, sweeping out various energy bands labeled by the index n .

While both of these applications of group theory to quantum mechanics are important, they both involve abelian groups with their relatively simple structure.

The most celebrated application of group theory during the early years of quantum mechanics, as is well known, is to atomic spectroscopy and involves the nonabelian rotation group $SO(3)$, but before discussing this, we will have to wait until we work out its irreducible representations.

Symmetry in the Laws of Physics: Lagrangian and Hamiltonian

Physics is where the action is

We talked about the invariance of physical laws as a motivation for group theory earlier.

At the level of the equation of motion, what we meant is clear: the left and right hand sides have to transform in the same way under the symmetry transformation.

In particular, in the case of rotational invariance, force and acceleration both have to transform as a vector for $\vec{F} = m\vec{a}$ to make sense.

Otherwise, Newton's force law would depend on how we orient the Cartesian coordinates.

More generally, in the equations of motion of physics, every term has to transform in the same way; the equation is said to be covariant.

Here we go over alternative formulations in which the manifestation of symmetry is made even simpler and more transparent.

We begin with a lightning review of the Lagrangian, and later of the Hamiltonian, for those readers unfamiliar or rusty with these two concepts from classical mechanics.

The Lagrangian will be needed later while the Hamiltonian has already been used.

For a particle moving in one dimension in a potential, Newton with his $ma = F$ tells us that

$$m \frac{d^2 q}{dt^2} = -V'(q) \quad (1)$$

with q the position of the particle.

Lagrange and others discovered later that this could be formulated as an action principle, with the action defined by the integral

$$S(q) \equiv \int dt L\left(\frac{dq}{dt}, q\right) = \int dt \left\{ \frac{1}{2} m \left(\frac{dq}{dt}\right)^2 - V(q) \right\} \quad (2)$$

In the simplest example of a particle moving in one dimension, the Lagrangian is given by its kinetic energy minus its potential energy

$$L = \frac{1}{2} m \left(\frac{dq}{dt}\right)^2 - V(q) \quad (3)$$

What does the action principle say?

The action $S(q)$ is a functional of the path $q(t)$.

With each path, we assign a real number, namely, the action $S(q)$ evaluated for that particular path $q(t)$.

We are then instructed to vary $S(q)$ subject to the condition that the initial and final positions $q(t_i)=q_i$ and $q(t_f)=q_f$ are to be held fixed.

The particular path $q(t)$ that extremizes the action $S(q)$ satisfies Newton's law.

Often, the more profound a truth is in theoretical physics, the simpler the demonstration is.

Here, too, the demonstration is almost laughably simple.

We vary the action:

$$\delta S(q) = \int dt \delta L \left(\frac{dq}{dt}, q \right) = \int dt \left(\frac{\delta L}{\delta \frac{dq}{dt}} \delta \frac{dq}{dt} + \frac{\delta L}{\delta q} \delta q \right) \quad (4)$$

Using $\delta \frac{dq}{dt} = \frac{d(q+\delta q)}{dt} - \frac{dq}{dt} = \frac{d\delta q}{dt}$, integrate the first term in (4) by parts to obtain

$$\delta S = \int dt \left(\frac{d}{dt} \left(\frac{\delta L}{\delta \frac{dq}{dt}} \right) - \frac{\delta L}{\delta q} \right) \delta q(t) \quad (5)$$

The boundary terms vanish, because the initial and final positions are held fixed.

The action is extremized when δS vanishes, which implies, since $\delta q(t)$ is arbitrary, the Euler-Lagrange equation

$$\frac{d}{dt} \left(\frac{\delta L}{\delta \frac{dq}{dt}} \right) - \frac{\delta L}{\delta q} = 0 \quad (6)$$

To be precise and pedantic, I stress that the notation means the following. We pretend that $L(a, b)$ is an ordinary function of two variables a and b . By $\frac{\delta L}{\delta \frac{dq}{dt}}$ we mean $\frac{\partial L(a, b)}{\partial a}$ with a subsequently set equal to $\frac{dq}{dt}$ and b to $q(t)$. Similarly, by $\frac{\delta L}{\delta q}$ we mean $\frac{\partial L(a, b)}{\partial b}$ with a subsequently set equal to $\frac{dq}{dt}$ and b to $q(t)$.

Switching from Leibniz's notation to Newton's notation $\dot{f}(x) = \frac{df}{dx}$, we can write the Euler-Lagrange equation (5) in the elegantly⁴ compact form

$$\frac{\delta \dot{L}}{\delta \dot{q}} = \frac{\delta L}{\delta q} \quad (7)$$

This equation, suitably generalized to quantum fields, underlies all known dynamics in the universe (we will come back to this eventually, in parts VII and VIII).

Applied to the Lagrangian (3), we obtain (1): $\frac{\delta L}{\delta \frac{dq}{dt}} = m \frac{dq}{dt}$ and $\frac{\delta L}{\delta q} = -V'(q)$, and thus (6) states that $\frac{d}{dt}(m \frac{dq}{dt}) + V'(q) = 0$, namely, $ma = F$.

Note that the minus sign in the definition of the Lagrangian is absolutely necessary. For the harmonic oscillator, $V(q) = \frac{1}{2}kq^2$, and the equation of motion (6) reads $m\ddot{q} = -kq$, which is immediately solved by* $q \propto e^{-i\omega t}$ with $\omega = \sqrt{k/m}$.

Local versus global

Newton's equation of motion is described as “local” in time: it tells us what is going to happen in the next instant.

In contrast, the action principle is “global”: one integrates over various possible trajectories and chooses the best one.

While the two formulations are mathematically entirely equivalent, the action principle offers numerous advantages over the equation of motion approach.

Perhaps the most important is that the action leads directly to an understanding of quantum mechanics via the so-called Dirac-Feynman path integral, but that is another story for another time.

Historically, the action principle was inspired by Fermat's least time principle for light, which, by the way, goes all the way back to Heron of Alexandria (circa 65 A.D.).

The least time principle has a strongly teleological flavor—that light, and particularly daylight, somehow knows how to save time—a flavor totally distasteful to the postrational palate.

In contrast, during Fermat's time, there was lots of quasitheological talk about Divine Providence and Harmonious Nature, so no one questioned that light would be guided to follow the most prudent path.

After the success of the least time principle for light, physicists naturally wanted to find a similar principle for material particles.

Richard Feynman recalled that when he first learned of the action principle, he was blown away.

Indeed, the action principle underlies some of Feynman's deepest contributions to theoretical physics, as was alluded to above.

The laws of physics and symmetry

At the start of this part, I mentioned that a law of physics enjoys a symmetry if the different terms in the equation of motion transform in the same way.

In the least action formulation, the presence of a symmetry can be stated even more simply: the action S has to be invariant under the symmetry transformation; that is, S does not change.

In other words, S must transform like a singlet.

Put yet another way, S belongs to the trivial representation of the symmetry group.

Again, the reason is clear: otherwise, the action S associated with each path would depend on how we orient the Cartesian coordinates.

All these statements about the action apply, in most cases, to the Lagrangian as well: L has to be invariant under the symmetry transformation.

The simplest example is that of a particle moving in a spherically symmetric potential.

The Lagrangian is easy to write down, since we know that out of a vector $\vec{q}$ there is only one invariant we can form, namely, the scalar product $\vec{q} \cdot \vec{q}$.

Thus, L must have the form (compare with (3)) $L = \frac{1}{2}m(\frac{d\vec{q}}{dt} \cdot \frac{d\vec{q}}{dt}) - V(|\vec{q}|)$, where $|\vec{q}| = (\vec{q} \cdot \vec{q})^{\frac{1}{2}}$ denotes the length of the vector $\vec{q}$.

The rotation group is almost too familiar to be of much pedagogical value in this context.

In general, the relevant symmetry group might be much less familiar than the rotation group.

For example, we could have four equal masses connected by six identical springs to form a tetrahedron.

The symmetry group would then be our old friend A_4 from earlier.

The Hamiltonian

After Lagrange invented the Lagrangian, Hamilton invented the Hamiltonian.

Given a Lagrangian $L(\dot{q}, q)$, define the momentum by $p = \frac{\delta L}{\delta \dot{q}}$ and the Hamiltonian by

$$H(p, q) = p\dot{q} - L(\dot{q}, q) \quad (8)$$

where it is understood that $\dot{q}$ on the right hand side is to be eliminated in favor of p .

Let us illustrate this procedure by a simple example. Given the Lagrangian $L(\dot{q}, q) = \frac{1}{2}m\dot{q}^2 - V(q)$ in (3), we have $p = m\dot{q}$, which is precisely what we normally mean by momentum. The Hamiltonian is then given by

$$H(p, q) = p\dot{q} - L(\dot{q}, q) = p\dot{q} - \frac{1}{2}m\dot{q}^2 + V(q) = \frac{p^2}{2m} + V(q) \quad (9)$$

where in the last step we wrote $\dot{q} = p/m$. You should recognize the final expression as the total energy, namely, the sum of the kinetic energy $\frac{p^2}{2m}$ and the potential energy $V(q)$. The Hamiltonian represents the total energy of the system.

Given a Hamiltonian $H(p, q)$, Hamilton postulated the equations⁸

$$\dot{q} = \frac{\partial H}{\partial p}, \quad \dot{p} = -\frac{\partial H}{\partial q} \quad (10)$$

For the Hamiltonian in (9), these equations read $\dot{q} = \frac{p}{m}$ and $\dot{p} = -\frac{\partial V}{\partial q}$. Plugging the first equation into the second, we obtain $m\ddot{q} = -V'(q)$, thus recovering the correct equation of motion.

The role of formalism in theoretical physics

Practical minded theoretical physicists would question the utility of introducing the Lagrangian and the Hamiltonian.

These alternative formulations merely rewrite Newton's equation of motion, they might snort.

Indeed, the theoretical physics literature is full of papers that many practitioners dismiss as mere formalism.

Most of them do deserve to be consigned to the dustbins of history.

But in a few cases, in particular the Lagrangian and the Hamiltonian (otherwise we would hardly be talking about them here), the formalism turns out to be of central importance and points the way toward further developments.

Later, we will see that from the Lagrangian, it is but a short skip and hop to quantum field theory.

As for the Hamiltonian, it is well known that quantum mechanics was born when Heisenberg promoted q and p , and hence H , to operators. (Indeed, we already invoked this fact earlier.)

We will discuss this further later.

Take home message: Invariance easier than covariance

Given a Lagrangian, we have to vary it to obtain the equation of motion.

For normal everyday use, we might as well stick with the equation of motion.

But when symmetry comes to the fore and assumes the leading role, it is significantly easier to render the Lagrangian invariant than to render the equation of motion covariant.

For the familiar rotation group, the advantage is minimal; however, when we come to more involved symmetries, finding the correct Lagrangian is usually much easier than finding the correct equation of motion.

Indeed, as an example, when Einstein searched for his theory of gravity, he opted for the equation of motion.

He could have saved himself a considerable amount of travail if he had determined the relevant Lagrangian instead.

Symmetry and conservation

No free lunch in physics!

What I mean to say is that conservation laws abound in physics: energy is conserved, momentum is conserved, angular momentum is conserved, the number of quarks in the universe is conserved, the list goes on and on.

Emmy Noether, arguably the deepest woman physicist who ever lived, had the profound insight that each conservation law is associated with a continuous symmetry of the Lagrangian.

For example, angular momentum conservation corresponds to invariance of the action under the rotation group $SO(3)$.

Once again, group theory reigns supreme.

Indeed, each conserved quantity corresponds to a generator of the Lie algebra.

I cannot go into further details here, but mention the Noether relation between conservation and symmetry for future reference.

4. Tensor, Covering, and Manifold

No need to be tense about tensors at all.

Thanks to index democracy, tensors furnish a clever device for generating ever larger representations.

The $SO(N)$ groups, which generalize the ordinary rotation group, are used to illustrate this claim.

Another clever trick: introduce raising and lowering operators in the $SO(3)$ algebra.

You can then figure out the allowed representations by climbing up and down a ladder and requiring it to end at both ends.

Creation and annihilation operators represent another clever trick.

You learn how to multiply two representations together and decompose the product.

The method of tensors is extended to $SU(N)$, with the important consequence that the indices now have to live on two floors.

From $SU(N)$ we specialize to $SU(2)$ and learn about its many peculiarities.

In particular, it double covers $SO(3)$.

You will meet the mysterious spinor that turns out to be crucial for modern physics.

The behavior of electron spin under time reversal leads to a two-fold degeneracy.

You will learn to integrate over continuous groups, some of which have interesting topologies.

The familiar sphere appears not as a group manifold, but as a coset manifold.

The symplectic groups and their algebras are studied, with some interesting isomorphisms with more familiar algebras.

You will learn that quantum field theory hardly deserves its frightening reputation among students; it is but a short hop from Lagrangian mechanics.

Finally, in four interludes, we sample some interesting topics, such as crystal field splitting and special functions such as the one belonging to Bessel.

Tensors and Representations of the Rotation Groups $SO(N)$

Representing the rotation groups

By now, you know how to represent a finite group.

How about representing the continuous groups, such as $SO(N)$, the rotation group in N -dimensional Euclidean space?

We defined $SO(N)$, earlier, as the group of all $N \times N$ matrices R satisfying

$$R^T R = I \tag{1}$$

and

$$\det R = 1 \tag{2}$$

We might as well do things more generally in N -dimensional space; it costs us practically nothing except for allowing the relevant matrix indices to run all the way up to N instead of 3.

But to help those readers seeing this for the first time focus, I will often specialize to $N = 3$.

Henceforth, I will jump back and forth between N and 3 without further ado.

We will see later that $SO(3)$ has some special features not shared by $SO(N)$.

So, we already know from (1) and (2) the N -dimensional defining or fundamental representation (as well as the trivial 1-dimensional representation, of course).

The elements of the $SO(N)$ are represented, by definition, by the $N \times N$ matrices transforming the N unit basis vectors $\vec{e}_1, \vec{e}_2, \dots, \vec{e}_N$ into one another.

More precisely, the N -dimensional irreducible representation is furnished by a vector.

Let me stress again that a vector is defined by how it transforms under a rotation:

$$V^i \rightarrow V'^i = R^{ij} V^j \tag{3}$$

with $i, j = 1, 2, \dots, N$.

Several questions and some flying guesses

Several questions naturally suggest themselves (as earlier).

How many irreducible representations does the rotation group have?

How are they to be characterized?

What are their dimensions?

If you are encountering this material for the first time, the answers may be far from self-evident.

Indeed, if I were to tell you that the rotation group for good old 3-dimensional Euclidean space has a 5-dimensional irreducible representation, it may very well come as a surprise to you, as it did to me.

What are these five beasts living in 3-dimensional space?

Our understanding of the representation theory of finite groups offers some suggestive answers to the questions posed above.

As indicated earlier, the theorems we proved for finite groups depend crucially on the rearrangement identity and on the meaningfulness of the sum over group elements.

Our intuition suggests that, with some suitable modifications, they should continue to hold.

Indeed, as proud physicists, let's take a flying guess at the representations of $SO(2)$.

Consider the finite group Z_N , which we may think of as generated by the rotation through angle $2\pi/N$.

Earlier, we learned that it has N 1-dimensional irreducible representations, labeled by an integer $k=0, \dots, N-1$.

The group element $e^{i2\pi j/N}$ is represented by $D^{(k)}(e^{i2\pi j/N}) = e^{i2\pi kj/N}$.

Well, as $N \rightarrow \infty$, the finite group Z_N should turn into the continuous group $SO(2)$.

So, $SO(2)$ should have an infinite number of irreducible representations corresponding to $k=0, \dots, \infty$, in which a rotation through angle θ is represented by $e^{ik\theta}$.

Ha! The fly-by-night guess works: $D^{(k)}(\theta)D^{(k)}(\theta') = e^{ik\theta} e^{ik\theta'} = e^{ik(\theta + \theta')} = D^{(k)}(\theta + \theta')$ and $D^{(k)}(2\pi) = 1$.
(For more on this point, see later in this part.)

What do you expect?

We are self-confessed sloppy physicists in flagrant disregard for mathematical decorum, and we take limits whenever we like.

What about $SO(3)$, a larger group containing $SO(2)$ as a subgroup?

Anybody with the absolute minimum amount of mathematical sense can see that the irreducible representations of a group cannot necessarily be “lifted” into the irreducible representations of a larger group containing that group as a subgroup.

But still, let’s make another flying guess: $SO(3)$ also has an infinite number of irreducible representations labeled by an integer.

That also turns out to be true.

Constructing the irreducible representations of $SO(N)$

Both physicists and mathematicians, for reasons mysterious to the unsuspecting guy or gal in the street, would like to construct larger or higher-dimensional irreducible representations.

Earlier, in an attempt to construct larger representations of $SO(3)$, we, with laughable naivete, simply stacked smaller representations together.

The resulting representation is, by definition, reducible.

No good.

To obtain larger irreducible representations, we have to do something slightly more sophisticated: we have to generalize the concept of a vector by considering objects carrying more indices.

Imagine a collection of mathematical entities T^{ij} carrying two indices, with $i, j = 1, 2, \dots, N$ in N -dimensional space.

If the T_{ij} s transform under rotations according to

$$T^{ij} \rightarrow T'^{ij} = R^{ik} R^{jl} T^{kl} \quad (4)$$

then we say that T transforms like a tensor, and hence is a tensor. (Here we are using the Einstein summation convention: The right hand side actually means $\sum_{k=1}^N \sum_{l=1}^N R^{ik} R^{jl} T^{kl}$, that is, a double sum of N^2 terms.)

Indeed, we see that we are just generalizing the transformation law of a vector.

Fear of tensors

Allow me to digress.

In my experience teaching, invariably, a couple of students are confused by the notion of tensors.

The very word “tensor” apparently makes them tense.

Compounded and intertwined with their fear of tensors, the unfortunates mentioned above are also unaccountably afraid of indices.

Dear reader, if you are not one of these unfortunates, so much the better for you!

You can zip through this part.

But to allay the nameless fear of the tensorphobe,

I will go slow and be specific.

Think of the tensor T^{ij} as a collection of N^2 mathematical entities that transform into linear combinations of one another. Let us list T^{ij} explicitly for $N = 3$. There are $3^2 = 9$ of them: T^{11} , T^{12} , T^{13} , T^{21} , T^{22} , T^{23} , T^{31} , T^{32} , and T^{33} . That's it, nine objects that transform into linear combinations of one another. For example, (4) says that $T'^{21} = R^{2k} R^{1l} T^{kl} = R^{21} R^{11} T^{11} + R^{21} R^{12} T^{12} + R^{21} R^{13} T^{13} + R^{22} R^{11} T^{21} + R^{22} R^{12} T^{22} + R^{22} R^{13} T^{23} + R^{23} R^{11} T^{31} + R^{23} R^{12} T^{32} + R^{23} R^{13} T^{33}$. This shows explicitly, as if there were any doubt to begin with, that T'^{21} is given by a particular linear combination of the nine objects. That's all: the tensor T^{ij} consists of nine objects that transform into linear combinations of themselves under rotations.

We could generalize further and define 3-indexed tensors, 4-indexed tensors, and so forth by transformation laws, such as $W_{ijn} \rightarrow W'_{ij n} = R_{ik} R_{jl} R_{nm} W_{klm}$.

Here we focus on 2-indexed tensors, and if we say “tensor” without any qualifier we often, but not always, mean a 2-indexed tensor. (With this definition, we might say that a vector is a 1-indexed tensor and a scalar is a 0-indexed tensor, but this usage is not common.)

A scalar transforms as a tensor with no index at all, namely, $S \rightarrow S' = S$; in other words, a scalar does not transform.

The scalar furnishes the 1-dimensional trivial representation.

Representation theory

Mentally arrange the nine objects T_{ij} in a column $\begin{pmatrix} T^{11} \\ T^{12} \\ \vdots \\ T^{33} \end{pmatrix}$.

The linear transformation on the nine objects can then be represented by a 9 x 9 matrix $D(R)$ acting on this column.

For every rotation, specified by a 3 x 3 matrix R , we can thus associate a 9 x 9 matrix $D(R)$ transforming the nine objects T_{ij} linearly among themselves.

It is fairly clear that $D(R)$ gives a 9-dimensional representation of $SO(3)$, but for the benefit of the abecedarian, let us verify this explicitly.

Transform by the rotation R_1 followed by the rotation R_2 .

Using (4) twice, we obtain

$$\begin{aligned} T^{ij} &\rightarrow T'^{ij} = R_1^{ik} R_1^{jl} T^{kl} \\ &\rightarrow T''^{ij} = R_2^{ik} R_2^{jl} T'^{kl} = R_2^{ik} R_1^{km} R_2^{jl} R_1^{ln} T^{mn} = (R_2 R_1)^{im} (R_2 R_1)^{jn} T^{mn} \end{aligned} \quad (5)$$

Thus, indeed, $D(R_2)D(R_1) = D(R_2 R_1)$.

You see that the key is that each of the two indices on T_{ij} transforms independently, that is, in parallel without interfering with each other.

They live in the same household but do not talk to each other.

Another way to think of this (but one that is potentially confusing to some students) is that (4) mandates that a tensor T_{ij} transforms as if it were equal to the product of two vectors $V^i W^j$, but in general, it isn't.

Indeed, if T_{ij} were actually equal to $V^i W^j$, then $T_{11}/T_{21} = T_{12}/T_{22} = \dots = T_{1N}/T_{2N} = V^1/V^2$, and the N^2 components of the tensor would not be all independent.

The tensor T furnishes a 9-dimensional representation of the rotation group $SO(3)$.

Reducible or irreducible?

But is this 9-dimensional representation reducible or not?

Of these nine entities T_{ij} that transform into one another, is there a subset among them that only transform into one another?

A secret in-club, as it were.

A moment's thought reveals that there is indeed an in-club.

Consider $A_{ij} \equiv T_{ij} - T_{ji}$.

Under a rotation,

$$\begin{aligned} A^{ij} \rightarrow A'^{ij} &\equiv T'^{ij} - T'^{ji} = R^{ik} R^{jl} T^{kl} - R^{jk} R^{il} T^{kl} \\ &= R^{ik} R^{jl} T^{kl} - R^{jl} R^{ik} T^{lk} = R^{ik} R^{jl} (T^{kl} - T^{lk}) = R^{ik} R^{jl} A^{kl} \end{aligned} \tag{6}$$

I have again gone painfully slowly here, but it is obvious, isn't it?

We just verified in (6) that A^{ij} transforms like a tensor and is thus a tensor.

Furthermore, this tensor changes sign on interchange of its two indices ($A^{ij} = -A^{ji}$) and so is said to be antisymmetric.

The transformation law (4) treats the two indices democratically, without favoring one over the other, as was pointed out already, and thus preserves the antisymmetric character of a tensor: if $A^{ij} = -A^{ji}$, then $A^{ij} = -A^{ji}$ also.

Again, the key is that each of the two indices on T^{ij} transforms independently.

How many objects are here?

Let us count.

The index i in A^{ij} can take on N values; for each of these values, the index j can take on only $N - 1$ values (since the N diagonal elements $A^{ii} = -A^{ii} = 0$ for $i = 1, 2, \dots, N$, repeated index summation suspended momentarily here); but to avoid double counting (since $A^{ij} = -A^{ji}$), we should divide by 2.

Hence, the number of independent components in A is equal to $1/2 N(N - 1)$.

For example, for $N = 3$ we have the three objects A^{12} , A^{23} , and A^{31} .

The attentive reader will recall that we did the same counting earlier.

Clearly, the same goes for the symmetric combination $S_{ij} \equiv T_{ij} + T_{ji}$.

You can verify as a trivial exercise that $S_{ij} \rightarrow S'_{ij} = R_{ik}R_{jl}S_{kl}$. A

tensor S_{ij} that does not change sign on interchange of its two indices ($S_{ij} = S_{ji}$) is said to be symmetric.

In addition to the components S_{ij} with $i \neq j$, S also has N diagonal components, namely, $S^{11}, S^{22}, \dots, S^{NN}$.

Thus, the number of independent components in S is equal to $\frac{1}{2} N(N - 1) + N = \frac{1}{2} N(N + 1)$.

This is a long-winded way of saying that the symmetric tensor S has more components than the antisymmetric tensor A , but I have encountered confusion here among beginning students also.

For $N=3$, the number of components in A and S are $\frac{1}{2} \cdot 3 \cdot 2 = 3$ and $\frac{1}{2} \cdot 3 \cdot 4 = 6$, respectively.

For $N=4$, the number of components in A and S are 6 and 10, respectively.

Thus, in a suitable basis, the 9×9 matrix referred to above actually breaks up into a 3×3 block and a 6×6 block.

The 9-dimensional representation is reducible.

But we are not done yet.

The 6-dimensional representation is also reducible.

To see this, note that

$$S^{ii} \rightarrow S'^{ii} = R^{ik} R^{il} S^{kl} = (R^T)^{ki} R^{il} S^{kl} = (R^{-1})^{ki} R^{il} S^{kl} = \delta^{kl} S^{kl} = S^{kk} \quad (7)$$

In the third equality, we used the O in SO(N). (Here we are using repeated index summation: the indices i and k are both summed over.)

In other words, the linear combination $S_{11} + S^{22} + \dots + S^{NN}$, the trace of S, transforms into itself; that is, it does not transform at all.

It is a loner forming an in-club of one.

The 6 x 6 matrix describing the linear transformation of the six objects S^{ij} breaks up into a 1 x 1 block and a 5 x 5 block.

Again, for the sake of the beginning student, let us work out explicitly the five objects that furnish the representation 5 of SO(3).

First define a traceless symmetric tensor $\tilde{S}$ for SO(N) by

$$\tilde{S}^{ij} = S^{ij} - \delta^{ij} (S^{kk} / N) \quad (8)$$

(The repeated index k is summed over.) Explicitly, $\tilde{S}^{ii} = S^{ii} - N(S^{kk} / N) = 0$, and $\tilde{S}$ is traceless. Specialize to $N = 3$. Now we have only five objects, namely, $\tilde{S}^{11}$, $\tilde{S}^{22}$, $\tilde{S}^{12}$, $\tilde{S}^{13}$, and $\tilde{S}^{23}$. We do not count $\tilde{S}^{33}$ separately, since it is equal to $-(\tilde{S}^{11} + \tilde{S}^{22})$.

Under an $SO(3)$ rotation, these five objects transform into linear combinations of one another, as just explained. Let us be specific: the object $\tilde{S}^{13}$, for example, transforms into $\tilde{S}'^{13} = R^{1k} R^{3l} \tilde{S}^{kl} = R^{11} R^{31} \tilde{S}^{11} + R^{11} R^{32} \tilde{S}^{12} + R^{11} R^{33} \tilde{S}^{13} + R^{12} R^{31} \tilde{S}^{21} + R^{12} R^{32} \tilde{S}^{22} + R^{12} R^{33} \tilde{S}^{23} + R^{13} R^{31} \tilde{S}^{31} + R^{13} R^{32} \tilde{S}^{32} + R^{13} R^{33} \tilde{S}^{33} = (R^{11} R^{31} - R^{13} R^{33}) \tilde{S}^{11} + (R^{11} R^{32} + R^{12} R^{31}) \tilde{S}^{12} + (R^{11} R^{33} + R^{13} R^{31}) \tilde{S}^{13} + (R^{12} R^{32} - R^{13} R^{33}) \tilde{S}^{22} + (R^{12} R^{33} + R^{13} R^{32}) \tilde{S}^{23}$, where in the last equality we used $\tilde{S}^{ij} = \tilde{S}^{ji}$ and $\tilde{S}^{33} = -(\tilde{S}^{11} + \tilde{S}^{22})$. Indeed, $\tilde{S}^{13}$ transforms into a linear combination of $\tilde{S}^{11}$, $\tilde{S}^{22}$, $\tilde{S}^{12}$, $\tilde{S}^{13}$, and $\tilde{S}^{23}$.

To summarize, if instead of the basis consisting of the nine entities T^{ij} , we use the basis consisting of the three entities A^{ij} , the single entity S^{kk} (remember repeated index summation!), and the five entities $\tilde{S}^{ij}$, the 9×9 matrix $D(R)$ breaks up into a 3×3 matrix, a 1×1 matrix, and a 5×5 matrix “stacked on top of one another.”

This is represented schematically as

$$S^{-1}D(R)S = \left(\begin{array}{c|c|c} (3\text{-by-3 block}) & 0 & 0 \\ \hline 0 & (1\text{-by-1 block}) & 0 \\ \hline 0 & 0 & (5\text{-by-5 block}) \end{array} \right) \quad (9)$$

Note that once we choose the new basis, this decomposition holds true for all rotations.

In other words, there exists a similarity transformation S that block diagonalizes $D(R)$ for all R .

We say that in $SO(3)$, $9 = 5 \oplus 3 \oplus 1$.

More generally, the N^2 -dimensional representation furnished by a general 2-indexed tensor decomposes into a $1/2 N (N - 1)$ -dimensional representation, a $(1/2 N (N + 1) - 1)$ -dimensional representation, and a 1-dimensional representation.

For example, in $SO(4)$, $16 = 9 \oplus 6 \oplus 1$; in $SO(5)$, $25 = 14 \oplus 10 \oplus 1$.

You might have noticed that in this entire discussion we never had to write out R explicitly in terms of the three rotation angles and how the five objects $\tilde{S}^{11}, \dots, \tilde{S}^{23}$ transform into one another in terms of these angles.

It is only the counting that matters.

You might regard that as the difference between mathematics and arithmetic.

An advanced tidbit

Since I am going so slowly in this part, I am worried about boring the more advanced readers.

Here is a tidbit for them.

With some “minor” modifications, both the 6-dimensional and the 10-dimensional representations of $SO(4)$ play a glorious role in theoretical physics.

Can you guess what they are?

A collection of six objects, and a collection of ten objects.

I offer you a hint.

Had Maxwell and Einstein known some group theory, the development of theoretical physics might have been accelerated.

Invariant symbols

At this point, let us formalize what we just did.

The discussion, to be given for general N , will end up underlining why $N = 3$ is special.

First, go back to (7) and write it as $S'^{ii} = \delta^{ij} S'^{ij} = (\delta^{ij} R^{ik} R^{jl}) S^{kl} = \delta^{kl} S^{kl} = S^{kk}$.

Thus, the crucial property we need to show that the trace does not transform (in other words, that it furnishes the 1-dimensional trivial representation) is embedded in the third equality here, namely

$$\delta^{ij} R^{ik} R^{jl} = \delta^{kl} \tag{10}$$

This is just (1), and so, we have used the O in $SO(N)$.

What about the S in $SO(N)$, namely, the requirement that $\det R = 1$?

Linear algebra says that the determinant can be written in terms of the antisymmetric symbol $\epsilon^{ijk\dots n}$.

In N -dimensional space, the antisymmetric symbol carries N indices and is defined by its two properties:

$$\varepsilon^{\dots l \dots m \dots} = -\varepsilon^{\dots m \dots l \dots} \quad \text{and} \quad \varepsilon^{12 \dots N} = 1 \quad (11)$$

In other words, the antisymmetric symbol ε flips sign on the interchange of any pair of indices.

It follows that ε vanishes when two indices are equal. (Note that the second property listed is just normalization.)

Since each index can take on only values $1, 2, \dots, N$, the antisymmetric symbol for N -dimensional space must carry N indices, as already noted.

For example, for $N = 2$, $\varepsilon^{12} = -\varepsilon^{21} = 1$, with all other components vanishing.

For $N = 3$, $\varepsilon^{123} = \varepsilon^{231} = \varepsilon^{312} = -\varepsilon^{213} = -\varepsilon^{132} = -\varepsilon^{321} = 1$, with all other components vanishing.

The determinant is defined, for any matrix R , by

$$\varepsilon^{ijk \dots n} R^{ip} R^{jq} R^{kr} \dots R^{ns} = \varepsilon^{pqr \dots s} \det R \quad (12)$$

For R a rotation, $\det R = 1$, and hence

$$\varepsilon^{ijk \dots n} R^{ip} R^{jq} R^{kr} \dots R^{ns} = \varepsilon^{pqr \dots s} \quad (13)$$

Colloquially speaking, when a bunch of rotation matrices encounter the antisymmetric symbol, they poof into thin air.

Inspecting (10) and (13) (namely, the O and the S in $SO(N)$, respectively), we see that δ_{ij} and $\epsilon^{ij k \dots n}$ can be thought of as invariant symbols: when acted on by rotation matrices in the manner shown, they turn into themselves. (We refer to δ_{ij} and $\epsilon^{ij k \dots n}$ as symbols rather than tensors, because they are merely a collection of 0s and 1s.)

We mention in passing that there are various identities involving these symbols.

One trivial but important identity is that if U_{ij} is a symmetric tensor and V_{ij} an antisymmetric tensor, then $U_{ij} V_{ij} = 0$.

To see this, note that $U_{ij} V_{ij} = -U_{ji} V_{ji} = -U_{ij} V_{ij} = 0$, since something equal to its own negative has to vanish.

Note that the second equality follows from relabeling the dummy summation indices.

In particular, when the antisymmetric symbol is contracted with any symmetric tensor, the result vanishes.

For example, $\epsilon^{ij k \dots n} U_{ij} = 0$.

Dual tensors

In light of this discussion, given an antisymmetric tensor A_{ij} , we can define another antisymmetric tensor $B^{k \dots n} = \epsilon^{ij k \dots n} A_{ij}$ carrying $N - 2$ indices.

Because of (11), the tensor $B^{k \dots n}$ manifestly flips sign on exchange of any pair of indices.

Let us verify that it is in fact a tensor.

Under a rotation, $B^{k\dots n} \rightarrow \varepsilon^{ij k\dots n} R^i{}_p R^j{}_q A^{pq}$.

But multiplying (13) by a bunch of R^T s carrying appropriate indices, we have $\varepsilon^{ij k\dots n} R^i{}_p R^j{}_q = \varepsilon^{pqr\dots s} R^{kr} \dots R^{ns}$.

Thus,

$$B^{k\dots n} \rightarrow \varepsilon^{ijk\dots n} R^i{}_p R^j{}_q A^{pq} = \varepsilon^{pqr\dots s} R^{kr} \dots R^{ns} A^{pq} = R^{kr} \dots R^{ns} B^{r\dots s} \quad (14)$$

precisely how a tensor carrying $N - 2$ indices should transform.

The tensors A and B are said to be dual to each other.

In particular, for $N = 3$, $B^k = \varepsilon^{ij k} A^{ij}$ carries $3 - 2 = 1$ index and transforms like a vector.

Thus, in the preceding discussion, when we discovered that the 9-dimensional reducible representation decomposes into $5 \oplus 3 \oplus 1$, the 3 is not a new irreducible representation, but just the good old vector or defining representation. (The 1 is of course just the trivial representation.)

The one new irreducible representation that we have discovered is the 5.

This result is far from trivial.

A priori, if you have never heard of any of this, you might be quite surprised, as I said earlier, that 3-dimensional rotations could transform 5 objects exclusively into linear combinations of themselves.

For $N = 4$, a 2-indexed antisymmetric tensor is dual to another 2-indexed antisymmetric tensor, since $4 - 2 = 2$.

This “peculiar” fact has also played an important role in theoretical physics.

Here is a tidbit for the more advanced reader, related to the tidbit given earlier: the electric and magnetic fields are dual to each other.

We will come back to this later.

Constructing larger irreducible representations of $SO(N)$

We are now able to construct a large class of irreducible representations of $SO(N)$, known (not surprisingly) as the tensor representations, each furnished by a tensor $T^{ij\dots m}$, transforming by definition according to $T^{ij\dots n} \rightarrow T'^{ij\dots n} = R^{ik}R^{jl} \dots R^{nm}T^{kl\dots m}$.

As was explained earlier, since the transformation treats each index on T democratically, T' will have whatever symmetry properties T has.

For example, suppose T is symmetric in its first three indices, antisymmetric in its next four indices, symmetric in its next two indices, and so on; then T' will be the same.

Some people spend some fractions of their lives classifying and studying all possible such patterns, but in the spirit of the Feynman “shut up and calculate” school of physics, we will restrain ourselves from doing so.

The language of Young tableaux was invented to keep track of these patterns.

Fortunately, Nature is kind to us, and mostly we will have to deal only with smallish groups, such as $SO(3)$ and $SO(4)$.

Even in particle physics, where largish groups like $SO(10)$ and $SO(18)$ occasionally pop up, we only have to consider tensors with no more than a few indices.

My professor in graduate school told me not to bother learning Young tableaux, and I am giving you the same advice here.

If you do run across a tensor that is neither totally symmetric nor totally antisymmetric, such as the example given above, you can usually deal with it on a case-by-case basis.

It is rare that you will need a full-blown treatment using Young tableaux.

I have to omit some topics, and Young tableaux is one of those that I choose to omit.

Contraction of indices

When we set two indices on a tensor equal and sum, as in (7), we say that we contract the two indices.

To see how this works in general, take a general tensor transforming like $T^{ij\dots np} \rightarrow T'^{ij\dots np} = R^{ik}R^{jl} \dots R^{nm}R^{pq} T^{kl\dots mq}$.

Take any two indices, say j and n , and contract them.

Then

$$T^{ij\dots jp} \rightarrow T'^{ij\dots jp} = R^{ik}R^{jl} \dots R^{jm}R^{pq}T^{kl\dots mq} = R^{ik} \dots R^{pq}T^{kl\dots lq} \quad (15)$$

since $R^{jl}R^{jm} = \delta^{lm}$.

In other words, $T^{ij\dots jp}$ transforms like a tensor $T^{i\dots p}$ with two fewer indices; the contracted indices j and n have disappeared, knocking each other off, so to speak.

You see that (7) is just a special case of this: S^{ii} transforms just as if it has no index.

Confusio says, “Yes, I get it. Even though the letter j appears in $T^{ij\dots jp}$, this actually stands for $T^{i1\dots 1p} + T^{i2\dots 2p} + \dots + T^{iN\dots Np}$.”

Why SO(3) is special

As we saw just now, 3 is special because $3 - 2 = 1$: a pair of antisymmetric indices can always be traded for a single index.

For $SO(3)$, we claim that we need do business only with totally symmetric traceless tensors carrying j indices, with j an arbitrary positive integer, that is, a tensor $S^{i_1 i_2 \dots i_j}$ that remains unchanged on the interchange of any pair of indices and that vanishes when any two indices are contracted.

The claim will be proved inductively in j .

We have already seen that the claims hold for $j = 2$: the antisymmetric 2-indexed tensor is equivalent to a 1-indexed tensor.

Let us now move on to a 3-indexed tensor T^{ijk} and ask what new irreducible representation it contains.

As before, we could symmetrize and antisymmetrize in the first two indices and decompose the tensor into the symmetric combination $T^{\{ij\}k} \equiv (T^{ijk} + T^{jik})$ and the antisymmetric combination $T^{[ij]k} \equiv (T^{ijk} - T^{jik})$.

The standard notation $\{ij\}$ and $[ij]$ indicates that the tensor is symmetric or antisymmetric, respectively, in the bracketed indices.

We don't care about the antisymmetric combination $T^{[ij]k}$, because we know that secretly it is just a 2-indexed tensor $B^{lk} \equiv \epsilon^{ijl} T^{[ij]k}$, and we have already disposed of all 2-indexed tensors.

Our attack is inductive, as I said.

As for the symmetric combination $T^{\{ij\}k}$, we can now proceed to make it symmetric in all three indices by brute force.

Explicitly, write $3T^{\{ij\}k} = (T^{\{ij\}k} + T^{\{jk\}i} + T^{\{ki\}j}) + (T^{\{ij\}k} - T^{\{jk\}i}) + (T^{\{ij\}k} - T^{\{ki\}j})$.

Verify that the expression in the first parenthesis is completely symmetric in all three indices; indeed, it is just the sum over the $6 = 3!$ permutations of the three indices carried by T^{ijk} .

The expressions in the other two round parentheses are antisymmetric in k_i and k_j , respectively; we can thus multiply them by ϵ^{kil} and ϵ^{kjl} , respectively, turning them into 2-indexed tensors, which we drop down the inductive ladder.

Thus, the only thing new is a 3-indexed tensor S^{ijk} totally symmetric in all three of its indices.

Furthermore, as in our preceding discussion, we can subtract out its trace, so that the resulting tensor $\tilde{S}^{ijk}$ is traceless; that is, so that $\delta_{ij} \tilde{S}^{ijk} = 0$. (Note that, due to the complete permutation symmetry in the indices, it does not matter which pair of indices the Kronecker delta is contracted with.)

Thus, the new object is a totally symmetric traceless 3-indexed tensor $\tilde{S}^{ijk}$. (We drop the tilde without further ceremony.)

The claim is thus proved.

Dimension of the irreducible representations of $SO(3)$

Let us count the number of independent components contained in $S^{i_1 i_2 \cdots i_j}$, which gives the dimension of the irreducible representation furnished by $S^{i_1 i_2 \cdots i_j}$ and labeled by the integer j .

We will do this in three baby steps.

First, suppose that the indices can take on only two values, 1 and 2.

Then the independent components are $S^{22\dots 2}, S^{22\dots 21}, S^{22\dots 211}, \dots, S^{11\dots 1}$.

Since the number of 1s can go from 0 to j , we count $j + 1$ possibilities here.

Second, allow the indices to take on the value 3: then the possibilities are $S^{33\dots 3xx\dots x}$ (that is, among the indices are k xs, where x stands for either 1 or 2, with k ranging from 0 to j , and $j - k$ 3s).

Thus, the total number is determined by using Gauss's summation formula:

$$\sum_{k=0}^j (k + 1) = \frac{1}{2}j(j + 1) + (j + 1) = \frac{1}{2}(j + 1)(j + 2) \quad (16)$$

As an interim check, we have $1/2 \cdot 2 \cdot 3 = 3$ for $j = 1$, and $1/2 \cdot 3 \cdot 4 = 6$ for $j = 2$.

But we are not done yet: we have to impose the traceless condition:

The left hand side here is a totally symmetric tensor carrying $(j - 2)$ indices, which according to (16), has $1/2 (j - 2 + 1)(j - 2 + 2) = 1/2 (j - 1)j$ components.

Therefore, setting these to zero amounts to imposing $1/2 j (j - 1)$ conditions.

So, finally, the dimension of the irreducible representation j is

$$d = \frac{1}{2}(j + 1)(j + 2) - \frac{1}{2}j(j - 1) = \frac{1}{2}(j^2 + 3j + 2 - j^2 + j) = 2j + 1 \quad (17)$$

As a check, note that, for $j = 0, 1, 2, 3, \dots$, $d = 1, 3, 5, 7, \dots$, with the first three numbers confirming our earlier discussion.

We see that the dimension d goes up only linearly with j .

In contrast, an unrestricted tensor carrying j indices, with each index allowed to take on three values, will have 3^j components.

The formula $d = 2j + 1$ is famous in the history of quantum mechanics and atomic physics.

In fact, it had already appeared in classical physics in connection with spherical harmonics, to be discussed later.

The tensors of SO(2)

From $N = 3$ let us descend to $N = 2$.

Note that the antisymmetric symbol ε^{ij} now carries only two indices.

Suppose a tensor $T^{\dots i \dots j \dots}$ carrying m indices is antisymmetric in the pair of indices i and j .

We can contract it with ε_{ij} to obtain a tensor $\varepsilon_{ij} T^{\dots i \dots j \dots}$ carrying $m - 2$ indices.

Consequently, in our inductive construction, at each step we can immediately proceed to considering only totally symmetric tensors $S^{i_1 i_2 \dots i_j}$.

In the preceding paragraph, we already determined that there are $j + 1$ of these.

But we have not yet imposed the traceless condition $\delta^{i_1 i_2} S^{i_1 i_2 \cdots i_j} = 0$.

Arguing as before, we see that the left hand side of this condition is a symmetric tensor with $j - 2$ indices, and hence these amount to $j - 2 + 1$ conditions.

Hence the dimensions of the irreducible representations are $(j + 1) - (j - 2 + 1) = 2$.

All of them are 2-dimensional!

Indeed, a moment's thought reveals what the representation matrices are

$$D^{(j)}(\theta) = \begin{pmatrix} \cos j\theta & \sin j\theta \\ -\sin j\theta & \cos j\theta \end{pmatrix} \quad (18)$$

In particular, $j = 1$ corresponds to the defining or fundamental representation.

Polar decomposition

Now Confusio yells indignantly, “What’s going on? You said at the beginning of this part that the irreducible representations of $SO(2)$ are 1-dimensional, like the irreducible representations of Z_N . But here the irreducible representations are 2-dimensional!”

Excellent!

That Confusio is getting more attentive by the day.

So, what is going on?

Well, the representation given in (18) is in fact reducible.

Consider the unitary transformation

$$\begin{aligned} U^\dagger D^{(j)}(\theta) U &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix}^\dagger \begin{pmatrix} \cos j\theta & \sin j\theta \\ -\sin j\theta & \cos j\theta \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix} \\ &= \frac{1}{2} \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix} \begin{pmatrix} e^{ij\theta} & e^{-ij\theta} \\ ie^{ij\theta} & -ie^{-ij\theta} \end{pmatrix} \\ &= \begin{pmatrix} e^{ij\theta} & 0 \\ 0 & e^{-ij\theta} \end{pmatrix} \end{aligned} \tag{19}$$

We have shown explicitly that the 2-dimensional representation $D^{(j)}(\theta)$ reduces to two 1-dimensional representations $e^{ij\theta}$ and $e^{-ij\theta}$.

We recognize the two columns $\begin{pmatrix} 1 \\ i \end{pmatrix}$ and $\begin{pmatrix} 1 \\ -i \end{pmatrix}$ in the unitary matrix U as eigenvectors of $D^{(j)}(\theta)$ with eigenvalue equal to $e^{ij\theta}$ and $e^{-ij\theta}$, respectively.

This corresponds to going from Cartesian coordinates (x, y) to complex coordinates $(z, z^*) = (x + iy, x - iy)$.

In physics, for example in the study of electromagnetic waves, this corresponds to going from transverse to circular polarizations, by writing $\vec{E}_x \pm i\vec{E}_y$, and so on.

It is also instructive to see how this works explicitly with the Cartesian tensors we have been using. Consider $j = 2$, with $S^{11} = -S^{22}$, $S^{12} = S^{21}$.

The polar decomposition corresponds to writing $S^{++} = S^{1+i2, 1+i2}$, $S^{--} = S^{1-i2, 1-i2}$.

In fact, this conversation confirms the isomorphism between $SO(2)$ and $U(1)$, noted earlier.

The 1-to-1 map is just $R(\theta) \leftrightarrow e^{i\theta}$.

Rotations in higher-dimensional space

The preceding discussion underlines an important fact.

Among the rotation groups $SO(N)$, the two cases $N = 3$ and $N = 2$ that are most familiar to physicists are in fact rather special.

Crucially, their antisymmetric symbols carry three and two indices, respectively.

To see what goes “wrong” when we go to higher N , consider a tensor T^{hijkl} that is symmetric in the three indices hij and antisymmetric in the two indices kl .

For $N = 3$, contracting with ϵ^{klm} reduces the number of indices from five to four, a possibility we exploited.

We need consider only totally symmetric tensors.

In contrast, for $N = 4$, contracting with the antisymmetric symbol gives $\varepsilon^{klmn}T^{ijkl}$, a tensor with the same number of indices, namely, five.

Thus, for $N > 3$, we would have to confront, in general, tensors with complicated symmetry patterns on interchanges of indices.

A comment to forestall any potential confusion.

The word “symmetry” is necessarily overused in group theory.

In discussing tensors, we often talk about whether a tensor is symmetric or antisymmetric under interchange of its indices.

The group involved is some permutation group, which has little or nothing to do with the group that the tensor is furnishing a representation of.

Self-dual and antiself-dual

The rotation group $SO(2n)$ in even-dimensional space enjoys an additional feature, that of self-dual and antiself-dual tensors.

Consider the antisymmetric tensor with n indices $A^{i_1 i_2 \cdots i_n}$ with $2n(2n-1) \cdots (n+1)/n! = (2n)!/(n!)^2$ components.

From our discussion, you would think that it furnishes an irreducible representation.

What could possibly reduce it?

Construct the tensor $B^{i_1 i_2 \dots i_n} \equiv \frac{1}{n!} \varepsilon^{i_1 i_2 \dots i_n i_{n+1} i_{n+2} \dots i_{2n}} A^{i_{n+1} i_{n+2} \dots i_{2n}}$ dual to A . Then A is dual to B , that is, $A^{i_{n+1} i_{n+2} \dots i_{2n}} = \frac{1}{n!} \varepsilon^{i_1 i_2 \dots i_n i_{n+1} i_{n+2} \dots i_{2n}} B^{i_1 i_2 \dots i_n}$.

It follows that the two tensors $T_{\pm}^{i_1 i_2 \dots i_n} \equiv (A^{i_1 i_2 \dots i_n} \pm B^{i_1 i_2 \dots i_n})$ are self-dual and antiself-dual, respectively. Schematically, $\varepsilon T_{\pm} \sim \varepsilon(A \pm B) \sim \varepsilon A \pm \varepsilon B \sim B \pm A \sim \pm(A \pm B) \sim \pm T_{\pm}$. Thus, $T_{+} \sim \varepsilon T_{+}$ is dual to itself, while $T_{-} \sim -\varepsilon T_{-}$ is dual to minus itself.

Clearly, under an $SO(2n)$ transformation, T_{+} transforms into a linear combination of T_{+} , while T_{-} transforms into a linear combination of T_{-} .

The two tensors correspond to two irreducible representations with dimension $(2n)!/(2(n!)^2)$, not $(2n)!/(n!)^2$.

For example, for $SO(6)$, the dimension of the self-dual and antiself-dual representation is equal to $6 \cdot 5 \cdot 4 / (2 \cdot 3 \cdot 2) = 20/2 = 10$.

Restriction to a subgroup

Consider an irreducible representation of some group G .

If we restrict ourselves to a subgroup $H \subset G$, that irreducible representation will in general break up into several irreducible representations of H .

This makes sense, since we have fewer transformations to take the components of that representation into one another.

This is best explained by some examples.

Let $G = \text{SO}(4)$, with the defining or vector representation consisting of the components of the 4-vector $V_i, i = 1, 2, 3, 4$.

Consider the subgroup $H = \text{SO}(3)$ consisting of those elements of $\text{SO}(4)$ that leave V^4 alone.

In other words, the subgroup $\text{SO}(3)$ rotates only (V^1, V^2, V^3) into one another.

The four objects V_i split into two sets: (V^1, V^2, V^3) and V^4 .

We write this as $4 \rightarrow 3 \oplus 1$: the 4-dimensional vector representation of $\text{SO}(4)$ breaks into a 3-dimensional representation and a 1-dimensional representation of $\text{SO}(3)$.

Later we will discuss in detail the Lorentz group, which at this point I simply say is, roughly speaking, just $\text{SO}(4)$ suitably modified.

The subgroup $\text{SO}(3)$ is the good old rotation group.

Then the statement $4 \rightarrow 3 \oplus 1$ simply states that spacetime breaks up into space plus time in nonrelativistic physics.

How does the 6-dimensional irreducible representation of $\text{SO}(4)$ furnished by the antisymmetric tensor A_{ij} break up?

We simply enumerate: A^{14}, A^{24}, A^{34} and A^{12}, A^{23}, A^{31} .

In other words, $6 \rightarrow 3 \oplus 3$.

As we shall see later, this corresponds to the electromagnetic field breaking into the electric and the magnetic fields.

How about the $1/2 \cdot 4 \cdot 5 - 1 = 9$ -dimensional irreducible representation of $SO(4)$ furnished by the symmetric traceless tensor S^{ij} ?

Again, we simply list the nine objects.

How do they break up?

Evidently, it is useful to introduce the indices $a, b = 1, 2, 3$.

First, we have S^{44} , which furnishes the 1-dimensional representation of $SO(3)$.

Next, S^{a4} , $a = 1, 2, 3$, which furnishes the 3-dimensional representation of $SO(3)$.

Finally, we form the symmetric traceless tensor $\tilde{S}^{ab} = S^{ab} - 1/3 \delta^{ab} S^{cc}$ (with c summed over 1, 2, 3; see (8)), which furnishes the 5-dimensional representation of $SO(3)$.

Note that the trace S^{cc} we are taking out here is equal to $-S^{44}$, which has already been counted.

Thus, $9 \rightarrow 5 \oplus 3 \oplus 1$.

The adjoint representation and the Jacobi identity

As promised, we now discuss in more detail the adjoint representation briefly mentioned earlier.

Perhaps the best way to motivate the adjoint is to go back to the treatment of $\text{SO}(3)$ given before.

We commute the three matrices

$$J_x = -i \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}, \quad J_y = -i \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad J_z = -i \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (20)$$

representing the Lie algebra of $\text{SO}(3)$ and obtain the commutation relation

$$[J_a, J_b] = i \varepsilon_{abc} J_c \quad (21)$$

with $a, b, c = x, y, z$.

In physics, objects carrying three indices, such as the totally antisymmetric symbol ε_{abc} , are notoriously awkward to handle.

In contrast, objects carrying two indices can be regarded as matrices that can be naturally multiplied together.

But now let us make a seemingly naive observation.

Let us fix a , by setting $a = x$, for example.

Then ε_{1bc} carries two indices bc , and we can think of it as a matrix, with its entries in the b th row and c th column given by ε_{abc} .

But the sharp-eyed reader will recognize that this is precisely the first matrix listed in (20).

Indeed, the three matrices J_a are precisely equal to $(J_a)_{bc} = -i\varepsilon_{abc}$.

Here $(J_a)_{bc}$ denotes the entry in the b th row and c th column of the matrix J_a .

Is this an amazing coincidence or what?

The three matrices representing the generators in the defining representation are given by the structure constants.

Is this a general result?

The answer is more complicated than a simple yes or no.

First, no, it can't be general.

A moment's thought reminds us that for $SO(N)$, the defining or fundamental representation is N -dimensional, while the indices on the structure constants f^{abc} range over $N(N-1)/2$ values.

It is only for $N = 3$ that $N = N(N-1)/2$.

But then, yes, in fact the structure constants do in general furnish a representation, the adjoint representation.

To understand this, we have to appeal to the Jacobi identity, which states that quite generally, for three matrices or operators A , B , and C ,

$$[[A, B], C] + [[B, C], A] + [[C, A], B] = 0 \quad (22)$$

You learned earlier that Lie algebras are defined by the commutation relations

$$[T^a, T^b] = i f^{abc} T^c \quad (23)$$

with the indices $a, b, c, \dots$ ranging over n values, where n denotes the number of generators of the algebra. Now plug this into Jacobi's identity: $[[T^a, T^b], T^c] + [[T^b, T^c], T^a] + [[T^c, T^a], T^b] = 0$. The first term is equal to $i f^{abd} [T^d, T^c] = (i f^{abd})(i f^{dcg}) T^g$. Similarly, we can write the other two terms as linear combinations of T^g s. Setting the coefficient of T^g to 0 then gives the remarkable identity involving the structure constants:

$$f^{abd} f^{dcg} + f^{bcd} f^{dag} + f^{cad} f^{dbg} = 0 \quad (24)$$

Now define the matrix $(T^b)^{cd}$ by specifying its entry in the c th row and d th column as follows:

$$(T^b)^{cd} = -i f^{bcd} \quad (25)$$

I'd like to alert the reader to a possible notational confusion.

In (23), T^b denotes a generator, an abstract mathematical entity, if you like, of the Lie algebra; (23) is a statement about the algebra, independent of any representation we might want to choose.

In contrast, (25) defines the specific matrix representing T^b in the adjoint representation.

The claim is that the matrices T^b defined in (25) represent the Lie algebra (23). To prove this, we use (25). Write the first term in (24) as $f^{abd} f^{d cg} = i f^{abd} (T^d)^{cg}$, and the second and third terms as $f^{bcd} f^{dag} + f^{cad} f^{dbg} = -(T^b)^{cd} (-T^a)^{dg} - (-T^a)^{cd} (-T^b)^{dg} = (T^b T^a)^{cg} - (T^a T^b)^{cg} = -([T^a, T^b])^{cg}$. (Note that we used the antisymmetry of the structure constant in the first two indices, for example, $f^{dag} = -f^{adg}$.) Thus, we obtain

$$([T^a, T^b])^{cg} = i f^{abd} (T^d)^{cg} \quad (26)$$

which is just (23) in the adjoint representation.

We have proved that the structure constants of a Lie algebra furnish a representation of the algebra, known as the adjoint, whose dimension is given by the number of generators.

The adjoint of SO(N)

We have just proved what we set out to prove for Lie algebras in general, so that's that.

No more need be said.

Still, it is instructive to see how this works out in the specific case of SO(N).

Let's start fresh and go about it using a route slightly different from that followed in the previous section.

Earlier in this part, we learned that the antisymmetric 2-indexed tensor T^{ij} in $SO(N)$ furnishes a $1/2 N(N-1)$ -dimensional irreducible representation.

Recall from earlier that the number of generators in $SO(N)$ is $1/2 N(N-1)$, with the generators represented in the defining N -dimensional representation by the anti-symmetric matrices

$$\mathcal{J}_{(mn)}^{ij} = (\delta^{mi} \delta^{nj} - \delta^{mj} \delta^{ni}) \quad (27)$$

Physicists also define the hermitean matrices $J_{(mn)} = -i \mathcal{J}_{(mn)}$.

As explained earlier, the matrix indices i and j are not to be confused with the indices m and n , which I put in parenthesis for pedagogical clarity.

The indices m and n indicate which of the $1/2 N(N-1)$ matrices we are talking about.

In particular, note that $\mathcal{J}_{(mn)}$ is an N -by- N matrix, and there are $1/2 N(N-1)$ of them.

The poor man might ask, "Is there a connection?" You bet!

To see this, first rewrite, for the sake of further clarity, the label (mn) , which takes on $1/2 N(N-1)$ values, as a , where the symbol a runs from 1 to $1/2 N(N-1)$.

Then we have $1/2 N(N-1)$ matrices $\mathcal{J}_a^{ij}$, each of which is a $1/2 N(N-1) \times 1/2 N(N-1)$ matrix.

Confusio says, “This is why my friends and I get thrown: there are so many $1/2 N(N-1)$ s all over the place.”

Indeed. Furthermore, for the most important case for physics, $N=3$, both $1/2 N(N-1)$ and N are equal to $3!$

Here comes the punchline: we can also regard the antisymmetric tensor T^{ij} as an $N \times N$ matrix, and hence write it as a linear combination of the $\mathcal{J}_a$ s, with coefficients denoted by A_a :

$$T^{ij} = \sum_{a=1}^{\frac{1}{2}N(N-1)} A_a \mathcal{J}_a^{ij} = A_a \mathcal{J}_a^{ij} \quad (28)$$

We invoked the repeated index summation convention in the second equality.

To forestall confusion, keep in mind the ranges of the indices in $A_a \mathcal{J}_a^{ij}$: $i = 1, \dots, N$ while $a = 1, \dots, 1/2 N(N-1)$.

In other words, we can trade T^{ij} for A_a and vice versa.

Check to make sure that we didn't lose anybody: indeed, there are $1/2 N(N-1)$ of each.

You may think of T^{ij} and A_a as the same thing expressed in two different bases: T^{ij} for A_a are linear combinations of each other.

How do the A_a s transform?

We go back to (4) and recall $T_{ij} \rightarrow T'_{ij} = R^{ik} R T^{kl} = R^{ik} T^{kl} (R^T)^{lj}$ and thus

$$T \rightarrow T' = R T R^T = R T R^{-1} \quad (29)$$

If you confuse the T for transpose with the T for tensor, go back to square one.

Regarded as an antisymmetric matrix, the antisymmetric tensor T transforms as if by a similarity transformation in linear algebra.

For an infinitesimal rotation, $R \simeq I + \theta_a J_a$, and so

$$T' \simeq (I + \theta_a \mathcal{J}_a) T (I + \theta_a \mathcal{J}_a) \simeq T + \theta_a [\mathcal{J}_a, T].$$

Thus, the variation of T under the rotation is given by

$$\delta T = \theta_a [\mathcal{J}_a, T] = \theta_a [\mathcal{J}_a, T_b \mathcal{J}_b] = \theta_a T_b [\mathcal{J}_a, \mathcal{J}_b].$$

In general, a Lie algebra is characterized by the commutation relations between its generators and structure constants.

In this context, $[J_a, J_b] = i f_{abc} J_c$ in physicists' notation, or after taking out some trivial factors of i,

$$[\mathcal{J}_a, \mathcal{J}_b] = f_{abc} \mathcal{J}_c.$$

Now we are almost done.

Writing

$$\delta T = \delta A_c \mathcal{J}_c = \theta_a A_b [\mathcal{J}_a, \mathcal{J}_b] = \theta_a A_b f_{abc} \mathcal{J}_c$$

and equating coefficients of $\mathcal{J}_c$, we obtain

$$\delta A_c = f_{abc} \theta_a A_b \tag{30}$$

The $1/2 N(N-1)$ objects A_a transform according to (30) and are said to furnish the adjoint representation. See (25).

This conclusion is actually totally familiar in the case of $SO(3)$, for which the structure constant f_{abc} is just the antisymmetric symbol ϵ_{abc} .

Using a vector notation (since for $SO(3)$ the indices a, b , and c take on three values, we recognize A_a as a vector), we can write $\delta \vec{A} = \vec{\theta} \otimes \vec{A}$.

With no loss of generality, let $\vec{\theta}$ point along the z -axis, and we obtain (as always, not hesitating to abuse notation if clarity is not compromised) $\delta A_x = -\theta A_y$, $\delta A_y = \theta A_x$.

We are back where we started earlier.

It is perhaps more illuminating if we revert to the notation J_x, J_y, J_z .

The statement here is simply that the three generators of $SO(3)$ transform like a vector: $\delta \vec{J} = \vec{\theta} \otimes \vec{J}$.

Entirely reasonable.

The adjoint representation is clearly of special significance, since it is how the generators transform, so to speak.

It is worth emphasizing again that in general the adjoint is not the same as the vector or fundamental representation (certainly $\frac{1}{2} N(N-1) \neq N$ for $N \neq 3$).

We will encounter the adjoint representation for other continuous groups later.

Lie Algebra of SO(3) and Ladder Operators: Creation and Annihilation

The Lie algebra of SO(3) is prototypical of all Lie algebras, and the method used in this part to deal with this algebra will give us a strong hint on how to deal with more involved algebras later, for example.

You will also learn how to multiply two irreducible representations of SO(3) together, the so-called Clebsch-Gordan decomposition.

Then we go on to discuss Casimir invariants, Legendre polynomials, and spherical harmonics.

Representing the Lie algebra of SO(3)

Earlier, we constructed the irreducible representations of the group SO(3) using the method of tensors.

In this part we construct the irreducible representations of the Lie algebra SO(3).

We found the algebra already earlier, namely,

$$[J_x, J_y] = i J_z, \quad [J_y, J_z] = i J_x, \quad [J_z, J_x] = i J_y \quad (1)$$

What does it mean to represent the algebra?

It means that we are to find three matrices J_x , J_y , and J_z such that the commutation relations in (1) are satisfied.

Since we know that rotations are given by exponentials of linear combinations of the J s, this would lead to matrices representing the rotation group.

We already know by construction the 3-dimensional defining representation of the algebra; the three 3×3 matrices J_x , J_y , and J_z were explicitly displayed earlier.

However, we know from earlier that there exists, for j a non-negative integer, a $(2j + 1)$ -dimensional representation of the rotation group.

Hence there must exist a corresponding $(2j + 1)$ -dimensional representation of the algebra: we simply consider infinitesimal rotations.

We will proceed pretending that we do not know this, so that we will end up confirming this result.

A word of clarification.

Strictly speaking, we should distinguish the matrices representing the abstract operators J_x , J_y , and J_z from the operators themselves.

But it would only clutter up things if we introduce more notation.

Instead, we follow the physicist's sloppy practice of using J_x , J_y , and J_z also to denote the matrices representing the abstract operators J_x , J_y , and J_z .

Climbing up and down on a ladder

Since the three generators J_x , J_y , and J_z do not commute, they cannot be simultaneously diagonalized (linear algebra).

But we can diagonalize one of them.

Choose J_z , and work in a basis in which J_z is diagonal.

The move that breaks the problem wide open may be familiar to some students of physics: it is akin to going from the 2-dimensional coordinates x , y to the complex variable $z = x + iy$, $z^* = x - iy$, and from a transversely polarized electromagnetic wave to a circularly polarized electromagnetic wave.

Define $J_{\pm} \equiv J_x \pm iJ_y$.

Then we can rewrite (1) as

$$[J_z, J_{\pm}] = \pm J_{\pm}, \quad [J_+, J_-] = 2J_z \quad (2)$$

Instead of working with matrices, we use Dirac's bra and ket notation.

Write the eigenvector of J_z with eigenvalue m as $|m\rangle$; in other words,

$$J_z |m\rangle = m |m\rangle \quad (3)$$

Since J_z is hermitean, m is a real number.

What we are doing is going to a basis in which J_z is diagonal; according to (2), $J_{\pm}$ cannot be diagonal in this basis.

Now consider the state $J_+|m\rangle$ and act on it with J_z :

$$\begin{aligned} J_z J_+ |m\rangle &= (J_+ J_z + [J_z, J_+]) |m\rangle = (J_+ J_z + J_+) |m\rangle = (m J_+ + J_+) |m\rangle \\ &= (m + 1) J_+ |m\rangle \end{aligned} \quad (4)$$

where the second equality follows from (2).

Henceforth, we will be using (2) repeatedly without bothering to refer to it.

Thus, $J_+|m\rangle$ is an eigenvector (or eigenstate; these terms are used interchangeably) of J_z with eigenvalue $m + 1$.

Hence, by the definition of $|m\rangle$, the state $J_+|m\rangle$ must be equal to the state $|m+1\rangle$ multiplied by some normalization constant; in other words, we have $J_+|m\rangle = c_{m+1} |m + 1\rangle$ with the complex number c_{m+1} to be determined.

Similarly, $J_z J_- |m\rangle = (J_- J_z + [J_z, J_-]) |m\rangle = (J_- J_z - J_-) |m\rangle = (m-1)J_- |m\rangle$, from which we conclude that $J_- |m\rangle = b_{m-1} |m-1\rangle$ with some other unknown normalization constant.

It is very helpful to think of the states $\dots, |m-1\rangle, |m\rangle, |m+1\rangle, \dots$ as corresponding to rungs on a ladder.

See figure 1.

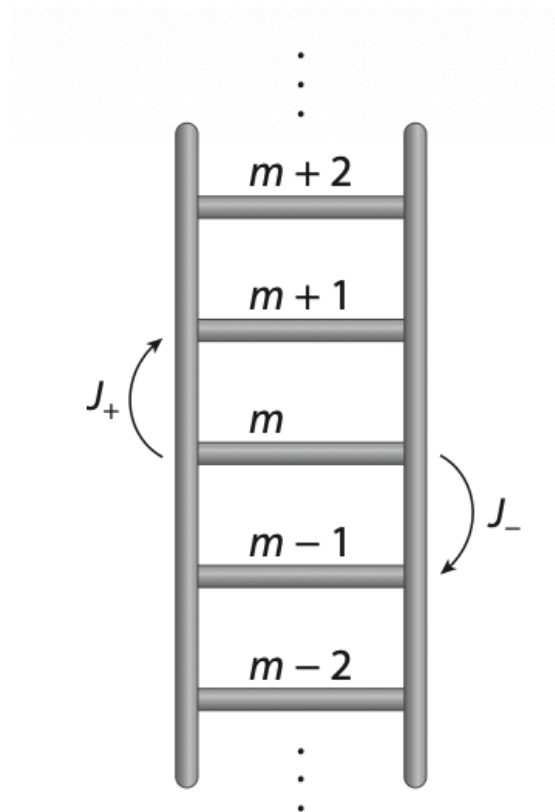


Figure 1

The result $J_+ |m\rangle = c_{m+1} |m+1\rangle$ tells us that we can think of J_+ as a “raising operator” that enables us to climb up one rung on the ladder, going from $|m\rangle$ to $|m+1\rangle$.

Similarly, the result $J_- |m\rangle = b_{m-1} |m-1\rangle$ tells us to think of J_- as a “lowering operator” that enables us to climb down one rung on the ladder.

Collectively, $J_{\pm}$ are referred to as ladder operators.

To relate b_m to c_m , we invoke the hermiticity of J_x , J_y , and J_z , which implies that

$$(J_+)^{\dagger} = (J_x + iJ_y)^{\dagger} = J_x - iJ_y = J_-.$$

Multiplying $J_+|m\rangle = c_{m+1}|m+1\rangle$ from the left by $\langle m+1|$ and normalizing the states by $\langle m|m\rangle = 1$, we obtain $\langle m+1|J_+|m\rangle = c_{m+1}$.

Complex conjugating this gives us $c_{m+1}^* = \langle m|(J_+)^{\dagger}|m+1\rangle = \langle m|J_-|m+1\rangle = \langle m|b_m|m\rangle = b_m$, that is, $b_{m-1} = c_m^*$, so that we can write $J_-|m\rangle = c_m^*|m-1\rangle$.

Acting on this with J_+ gives $J_+J_-|m\rangle = c_m^*J_+|m-1\rangle = |c_m|^2|m\rangle$.

Similarly, acting with J_- on $J_+|m\rangle = c_{m+1}|m+1\rangle$ gives $J_-J_+|m\rangle = c_{m+1}J_-|m+1\rangle = |c_{m+1}|^2|m\rangle$.

Requiring the ladder to terminate

Since we know that the representation is finite dimensional, the ladder must terminate, that is, there must be a top rung.

So, call the maximum value of m by j .

At this stage, all we know is that j is a real number. (Note that we have not assumed that m is an integer.)

Thus, there is a state $|j\rangle$ such that $J_+|j\rangle = 0$.

It corresponds to the top rung of the ladder.

At this point, we have only used the first part of (2).

Now we use the second half: $0 = \langle j | J_- J_+ | j \rangle = \langle j | J_+ J_- - 2J_z | j \rangle = |c_j|^2 - 2j$, thus determining $|c_j|^2 = 2j$.

Furthermore,

$\langle m | [J_+, J_-] | m \rangle = \langle m | (J_+ J_- - J_- J_+) | m \rangle = |c_m|^2 - |c_{m+1}|^2 = \langle m | 2J_z | m \rangle = 2m$. We obtain a recursion relation $|c_m|^2 = |c_{m+1}|^2 + 2m$, which, together with $|c_j|^2 = 2j$, allows us to determine the unknown $|c_m|$.

Here we go: $|c_{j-1}|^2 = |c_j|^2 + 2(j-1) = 2(2j-1)$, then $|c_{j-2}|^2 = |c_{j-1}|^2 + 2(j-2) = 2(3j-1-2)$, then eventually $|c_{j-s}|^2 = 2((s+1)j - \sum_{i=1}^s i)$. Recall the Gauss formula $\sum_{i=1}^s i = \frac{1}{2}s(s+1)$, and obtain $|c_{j-s}|^2 = 2((s+1)j - \frac{1}{2}s(s+1)) = (s+1)(2j-s)$.

We keep climbing down the ladder, increasing s by 1 at each step.

When $s = 2j$, we see that c_{-j} vanishes.

We have reached the bottom of the ladder.

More explicitly, we have $J_- | -j \rangle = c_{-j}^* | -j - 1 \rangle = 0$, according to what we just derived.

The minimum value of m is $-j$.

Since s counts the number of rungs climbed down, it is necessarily an integer, and thus the condition $s = 2j$ that the ladder terminates implies that j is either an integer or a half-integer, depending on whether s is even or odd.

If the ladder terminates, then we have the set of states $| -j \rangle, | -j + 1 \rangle, \dots, | j - 1 \rangle, | j \rangle$, which totals $(2j + 1)$ states.

For example, for $j = 2$, these states are $| -2 \rangle, | -1 \rangle, | 0 \rangle, | 1 \rangle$, and $| 2 \rangle$.

Starting from $| 2 \rangle$, we apply J_- four times to reach $| -2 \rangle$. (We will do this explicitly later.)

To emphasize the dependence on j , we sometimes write the kets $| m \rangle$ as $| j, m \rangle$.

Notice that the ladder is symmetric under $| m \rangle \rightarrow | -m \rangle$, a symmetry that can be traced to the invariance of the algebra in (1) under $J_x \rightarrow J_x, J_y \rightarrow -J_y$, and $J_z \rightarrow -J_z$ (namely, a rotation through π around the x-axis).

Earlier, we used the method of tensors to show that the representations of the group $SO(3)$ are $(2j + 1)$ -dimensional with j an integer.

So, for j an integer, the method of tensors and the method of Lie algebra agree. Yeah!

Mysterious appearance of the half integers

But what about the representations of the algebra corresponding to $j = \text{a half integer}$?

For example, for $j = 1/2$, we have a $2 \cdot 1/2 + 1 = 2$ -dimensional representation consisting of the states $| -1/2 \rangle$ and $| 1/2 \rangle$.

We climb down from $|1/2\rangle$ to $|-1/2\rangle$ in one step.

Certainly no sight of a 2-dimensional representation as earlier!

The mystery of the $j=1/2$ representation will be resolved later when we discuss $SU(2)$, but let's not be coy about it and keep the reader in suspense.

I trust that most readers have heard that it describes the electron spin.

We did not go looking for the peculiar number $1/2$; $1/2$ came looking for us.

Ladder operators

It should not escape your notice that as a by-product of requiring the ladder to terminate, we have also determined $|c_m|^2$.

Indeed, setting $s = j - m$, we had $|c_m|^2 = (j + m)(j - m)$.

Recalling the definition of c_m we obtain

$$J_+ |m\rangle = c_{m+1} |m+1\rangle = \sqrt{(j+1+m)(j-m)} |m+1\rangle \quad (5)$$

and

$$J_- |m\rangle = c_m^* |m-1\rangle = \sqrt{(j+1-m)(j+m)} |m-1\rangle \quad (6)$$

As a mild check on the arithmetic, indeed $J_+|j\rangle = 0$ and $J_-|-j\rangle = 0$.

You might also have noticed that, quite rightly, the phase of c_m is not determined, since it is completely up to us to choose the relative phase of the kets $|m\rangle$ and $|m-1\rangle$.

Beware that different authors choose differently.

I simply take c_m to be real and positive.

Tables of the c_m s for various j s are available, but it's easy enough to write them down when needed.

Note also that the square roots in (5) and (6) are related by $m \leftrightarrow -m$.

I list the two most common cases needed in physics:

$$j = \frac{1}{2}: \quad J_+ \left| -\frac{1}{2} \right\rangle = \left| \frac{1}{2} \right\rangle, \quad J_- \left| \frac{1}{2} \right\rangle = \left| -\frac{1}{2} \right\rangle \quad (7)$$

$$j = 1: \quad J_+ |-1\rangle = \sqrt{2} |0\rangle, \quad J_+ |0\rangle = \sqrt{2} |1\rangle \\ J_- |1\rangle = \sqrt{2} |0\rangle, \quad J_- |0\rangle = \sqrt{2} |-1\rangle \quad (8)$$

Note that the (nonzero) c_m for these two cases are particularly easy to remember (that is, if for some odd reason you want to): they are all 1 in one case, and $\sqrt{2}$ in the other.

Let us also write down the $j = 2$ case for later use:

$j = 2 :$

$$\begin{aligned} J_+ | -2 \rangle &= 2 | -1 \rangle, & J_+ | -1 \rangle &= \sqrt{6} | 0 \rangle, & J_+ | 0 \rangle &= \sqrt{6} | 1 \rangle, & J_+ | 1 \rangle &= 2 | 2 \rangle \\ J_- | 2 \rangle &= 2 | 1 \rangle, & J_- | 1 \rangle &= \sqrt{6} | 0 \rangle, & J_- | 0 \rangle &= \sqrt{6} | -1 \rangle, & J_- | -1 \rangle &= 2 | -2 \rangle \end{aligned} \quad (9)$$

Multiplying two SO(3) representations together

We will multiply SO(3) representations together using first the tensor approach from earlier and then the ladder approach discussed here.

Tensors are practically begging us to multiply them together.

Rather than treat the world's most general case (which will mostly amount to writing a lot ofs),

I think it pedagogically much clearer to treat a specific case.

Suppose we are given two SO(3) tensors: a symmetric traceless tensor S^{ij} and a vector T^k .

They furnish the 5-dimensional and the 3-dimensional irreducible representations, respectively.

The product $P^{ijk} = S^{ij} T^k$ is manifestly a 3-indexed tensor containing $5 \cdot 3 = 15$ components, but P^{ijk} is certainly not a totally symmetric traceless 3-indexed tensor.

It is symmetric and traceless in the first two indices, but has no particular symmetry on the interchange of j and k , for example.

Earlier, we learned that the irreducible representations of $SO(3)$ are furnished by symmetric traceless tensors.

So, the game is to beat P^{ijk} into a linear combination of symmetric traceless tensors.

First, construct the symmetric tensor $U^{ijk} = S^{ij} T^k + S^{jk} T^i + S^{ki} T^j$.

Its trace is given by the vector $U^k = \delta_{ij} U^{ijk} = 2S^i T^i$. (Note that it does not matter which pair of indices on U^{ijk} to contract when calculating the trace, since U^{ijk} is totally symmetric.)

Second, take out the trace by defining $\tilde{U}^{ijk} = U^{ijk} - 1/5 (\delta^{ij} U^k + \delta^{jk} U^i + \delta^{ki} U^j)$.

There are $(2 \cdot 3 + 1) = 7$ components in $\tilde{U}^{ijk}$: it furnishes a 7-dimensional irreducible representation.

Next we extract the antisymmetric part of the product $S^{ij} T^k$ by contracting it with the antisymmetric symbol: $V^{il} = S^{ij} T^k \varepsilon^{jkl}$. (There is of course no point in contracting the index pair ij with the ε symbol.)

The tensor V^{il} is traceless and so contains $3 \cdot 3 - 1 = 8$ objects.

Extract its symmetric and antisymmetric parts: $W^{il} = V^{il} + V^{li}$ and $X^{il} = V^{il} - V^{li}$.

The antisymmetric part can be written explicitly as a vector: $\frac{1}{2} X^{il} \varepsilon^{mil} = S^{ij} T^k \varepsilon^{jkl} \varepsilon^{mil} = S^{ij} T^k (\delta^{jm} \delta^{ki} - \delta^{ji} \delta^{km}) = S^{im} T^i$, which we recognize as the vector $\frac{1}{2} U^m$. (In the second equality we used a result from exercise IV.1.11.) This furnishes a 3-dimensional irreducible representation.

The symmetric part $W^{il} = S^{ij}T^k \epsilon^{jkl} + S^{lj}T^k \epsilon^{jki}$ is manifestly traceless (just set $i = l$ and sum).

Thus, it has $1/2 (3 \cdot 4) - 1 = 5$ components.

We have thus derived

$$5 \otimes 3 = 7 \oplus 5 \oplus 3 \tag{10}$$

We did not lose anybody: $5 \cdot 3 = 15 = 7 + 5 + 3$.

It should be clear how to treat the general case.

Given two totally symmetric traceless tensors, $S^{i_1 \cdots i_j}$ and $T^{k_1 \cdots k_{j'}}$, one with j indices, the other with j' indices, the product is then a tensor with $j + j'$ indices.

We first symmetrize this and take out its trace.

We get the irreducible representation labeled by $j + j'$.

Next, contract with ϵ^{ikl} , where i is an index on S and k an index on T .

We trade two indices, i and k , for one index l , and hence end up with a tensor with $j + j' - 1$ indices.

We get the irreducible representation labeled by $j + j' - 1$.

We repeat this process until there is nothing left to work with.

With no loss of generality, take $j \geq j'$. We have nothing left when all the indices on T are gone. We have thus shown that $j \otimes j'$ contains the irreducible representations $(j + j') \oplus (j + j' - 1) \oplus (j + j' - 2) \oplus \cdots \oplus (j - j' + 1) \oplus (j - j')$. Using the absolute value, we can drop the condition $j \geq j'$:

$$j \otimes j' = (j + j') \oplus (j + j' - 1) \oplus (j + j' - 2) \oplus \cdots \oplus (|j - j'| + 1) \oplus |j - j'| \quad (11)$$

For example, (10) can be written in the language of (11) as $2 \otimes 1 = 3 \oplus 2 \oplus 1$.

Later, we will derive this result in another way, but for now, let us count the number of components and check that we did not lose anybody.

We start with $(2j + 1)(2j' + 1)$ components.

To count the number of components contained in (11), we need $\sum_{n=1}^m 1 = (m - n + 1)$ and, using Gauss's trick, $\sum_{n=1}^m k = 1/2 (m - n + 1)(m + n)$, so that $\sum_{n=1}^m (2k + 1) = (m - n + 1)(m + n) + 1 = (m + 1)^2 - n^2$.

Then the number of components contained in (11) is given by.

$$\sum_{|j-j'|}^{j+j'} (2k + 1) = (j + j' + 1)^2 - (j - j')^2 = (2j + 1)(2j' + 1) \quad (12)$$

Precisely right.

I trust that you recognize that the very way in which we introduced the notion of a 2-indexed tensor in earlier represents a special case of (11).

We define the tensor as an object that transforms as if it were composed of the product of two vectors; thus $1 \otimes 1 = 0 \oplus 1 \oplus 2$.

Indeed, that was the content of the section “Reducible or irreducible?”: we showed that $3 \otimes 3 = 9 = 1 \oplus 3 \oplus 5$.

How do we name irreducible representations?

This is probably as good a place as any to comment on the conventions for naming irreducible representations.

Here you have already seen two for $SO(3)$: we can refer to the irreducible representation by the number of indices on the tensor furnishing the representation, as in j , or by its dimension, $2j + 1$.

For larger groups, such as $SO(N)$ for $N > 3$, as was explained earlier, a tensor can have rather involved symmetry properties under the interchange of its indices.

We need to specify its symmetry properties.

For instance, in $SO(N)$, the representation furnished by a totally antisymmetric tensor with n indices is often denoted by $[n]$ and the representation furnished by a totally symmetric traceless tensor with n indices by $\{n\}$. (Obviously, $[1] = \{1\}$. For $SO(3)$, we explained that $[2] = \{1\} = [1]$.)

Different conventions are used in different areas of physics and are favored by different authors.

I usually prefer the convention of identifying the irreducible representation by its dimension. (Usually, there is no risk of confusion, but ambiguities arise occasionally: distinct representations may happen to have the same dimension.)

With this convention, you can see instantly that nobody has gone missing when you multiply irreducible representations together, as in (10).

For the group $SO(3)$, with its long standing in physics and its connection with atomic spectroscopy, the confusion over names is almost worse than in reading Russian novels.

As you know, the trivial representation 0 or 1 is also called “s-wave”!

And the list goes on, 1 or 3 is called p, 2 or 5 is called d, 3 or 7 is called f, and so on.

One glaringly self-evident point is that the letters used in the preceding section, S, T , P , U , V , W , and X, are totally irrelevant; they are merely coatracks on which to hang the indices.

This is reminiscent of one motivation behind the Dirac bra and ket notation: no sense in writing the letter ψ all the time.

We saw earlier , that the generators of $SO(3)$, J_x , J_y , and J_z , transform like a vector under rotation: $\delta \vec{J} = \vec{\theta} \otimes \vec{J}$. This implies that rotations do not change $J^2 \equiv \vec{J}^2 = J_x^2 + J_y^2 + J_z^2$, known as the Casimir invariant for $SO(3)$. Indeed, we have* $[J_i, \vec{J}^2] = [J_i, J_j J_j] = i\varepsilon_{ijk}([J_i, J_j]J_j + J_j[J_i, J_j]) = i\varepsilon_{ijk}(J_k J_j + J_j J_k) = 0$. (In general, a Lie algebra may possess several Casimir invariants; one example will be shown in chapter VII.2.)

In particular, $[J_{\pm}, \vec{J}^2] = 0$. Thus, $\vec{J}^2$ evaluated on the states in an irreducible representation j should give the same value. We can readily verify this statement and determine this value. First, write

$$J^2 = \frac{1}{2}(J_+ J_- + J_- J_+) + J_z^2 \tag{13}$$

Then, $J^2 |j, m\rangle = \left(\frac{1}{2}(J_+J_- + J_-J_+) + J_z^2\right) |j, m\rangle = \left(\frac{1}{2}(|c_m|^2 + |c_{m+1}|^2) + m^2\right) |j, m\rangle = j(j+1) |j, m\rangle$. As expected, J^2 evaluated on $|j, m\rangle$ depends only on j , but not on m .

This result is sometimes written as

$$J^2 = j(j+1) \tag{14}$$

where it is understood that the equality holds in the irreducible representation j . (Some beginners are confused by this equality, partly because of the standard notation of dropping the arrow on $\vec{J}^2$. Thus, importantly, J^2 is not equal to j^2 , because $\vec{J}$ is not a set of numbers, but a set of three noncommuting operators; j is of course just an integer or a half integer.)

Legendre polynomials and spherical harmonics

I trust that most readers have encountered Legendre polynomials and spherical harmonics in their study of physics.

It should not come as a surprise that both are intimately connected to the rotation group $SO(3)$, since they pop up in problems with spherical symmetry.

To see the connection, let us go back to an earlier remark that the generators of $SO(3)$ can be written as differential operators rather than as matrices.

Here, to conform to age-old convention, we use L_i instead of J_i to denote the generators.

Earlier, I remarked that

$$L_x = i \left(z \frac{\partial}{\partial y} - y \frac{\partial}{\partial z} \right), \quad L_y = i \left(x \frac{\partial}{\partial z} - z \frac{\partial}{\partial x} \right), \quad L_z = i \left(y \frac{\partial}{\partial x} - x \frac{\partial}{\partial y} \right) \quad (15)$$

satisfy the commutation relations

$$[L_x, L_y] = iL_z, \quad [L_y, L_z] = iL_x, \quad [L_z, L_x] = iL_y \quad (16)$$

Behold the power of abstract mathematics: by earlier discussion in this part, we know immediately that there are functions of x, y, z , or equivalently r, θ, φ , that represent the Lie algebra (16).

These functions satisfy

$$\vec{L}^2 Y_l^m(\theta, \varphi) = l(l+1) Y_l^m(\theta, \varphi) \quad (17)$$

and

$$L_z Y_l^m(\theta, \varphi) = m Y_l^m(\theta, \varphi) \quad (18)$$

Without further ado, we know that $l = 0, 1, 2, \dots$ is a non-negative integer and that $m = -l, -l+1, \dots, l-1, l$. (The reason that we can take $Y_l^m(\theta, \varphi)$ to be independent of the radial coordinate r is because $\vec{L}$ acting on r gives nothing. To see this, note $L_z r^2 = i(y \frac{\partial}{\partial x} - x \frac{\partial}{\partial y})(x^2 + y^2 + z^2) = 2i(yx - xy) = 0$.)

The functions Y_l^m are known as spherical harmonics.

Note that, while the mathematics of the Lie algebra implies (17) and (18), if we actually want to know what Y_{m_1} is, we have to slog through what we might refer to as “rather tedious arithmetic, which we did in the Advanced Quantum lectures.

Tensors and Representations of the Special Unitary Groups $SU(N)$

Since this part is fairly involved, I first give you an overview or roadmap.

Now that we have worked out the special orthogonal groups $SO(N)$, we simply try to repeat everything for the special unitary groups $SU(N)$.

My pedagogical strategy is to work out the tensors for $SU(N)$ by following our noses; soon enough, we will stumble into the fact that the tensors of $SU(N)$, in contrast to the tensors of $SO(N)$, have to carry two kinds of indices, upper and lower.

The important message for the beginning reader is that this fact did not capriciously and mysteriously fall from the sky, but merely reflects the presence in $SU(N)$, but not in $SO(N)$, of the basic operation of complex conjugation.

From orthogonal to unitary groups

We worked out the tensors and representations of the special orthogonal groups $SO(N)$ earlier.

Here we extend that discussion to the tensors and representations of the special unitary groups $SU(N)$.

Roughly speaking, the extension corresponds to going from real to complex numbers.

We start with a quick review of the special orthogonal groups $SO(N)$.

We learned that $SO(N)$ consists of rotations in N -dimensional space, namely, all $N \times N$ real matrices O that are orthogonal

$$O^T O = 1 \tag{1}$$

and have unit determinant

$$\det O = 1 \tag{2}$$

The condition (1) ensures that the length squared $v^T v$ of the N -component vector v , which transforms according to $v \rightarrow v' = Ov$, is left unchanged or invariant.

As was remarked earlier, the condition that $v^T v$ is left invariant actually implies the apparently stronger condition that $u^T v$ is left invariant for u and v two arbitrary vectors.

Lie showed us how to solve (1) by writing rotations near the identity as $O \simeq I + A$.

To leading order in A , (1) becomes $A^T - A = 0$.

The most general A may then be written as a linear combination of $N(N-1)/2$ antisymmetric matrices.

Quantum physics and complex numbers

Classical physics is done with real numbers. With the advent of quantum mechanics, complex numbers came into physics in an essential way. Wave functions, for example, are in general complex.

We wish to generalize the discussion for $SO(N)$ by considering linear transformations on complex vectors ψ . While $v = \{v^j, j = 1, \dots, N\}$ consists of N real numbers, we now take $\psi = \{\psi^j, j = 1, \dots, N\}$ to consist of N complex numbers. We demand that the linear transformations $\psi \rightarrow \psi' = U\psi$ leave $\sum_{j=1}^N \psi^{j*}\psi^j = \psi^{*T}\psi \equiv \psi^\dagger\psi$ invariant. (The last step defines hermitean conjugation of a complex vector as complex conjugation followed by transposition; see the review of linear algebra.) Just as in the case of $SO(N)$, this implies[†] that for two arbitrary complex vectors ζ and ψ , under the transformations $\zeta \rightarrow U\zeta$ and $\psi \rightarrow U\psi$, the quadratic forms[‡] $\zeta^\dagger\psi$ and $\psi^\dagger\zeta$ are left invariant. Since ζ and ψ are arbitrary, this condition, that $\zeta^\dagger U^\dagger U \psi = \zeta^\dagger \psi$, leads to the requirement that $U^\dagger U = I$, in other words, that U is unitary.

The group $U(N)$

The group $U(N)$ is defined to consist of all $N \times N$ matrices U that are unitary

$$U^\dagger U = I \tag{3}$$

We can readily show that $U(N)$ is indeed a group.

To prove closure, we need to show that, if U_2 and U_1 are unitary matrices, then the matrix $U_2 U_1$ is also unitary.

Simply compute:

$$(U_2 U_1)^\dagger U_2 U_1 = U_1^\dagger U_2^\dagger U_2 U_1 = U_1^\dagger U_1 = I.$$

It is crucial that hermitean conjugation reverses the order in a product, that is, $(U_2 U_1)^\dagger = U_1^\dagger U_2^\dagger$, but this follows from the fact that transposition of a product of matrices reverses the order, that is, $(U_2 U_1)^T = U_1^T U_2^T$, a fact which also plays a crucial role in proving closure for the orthogonal groups.

The existence of the identity I and of the inverse $U^\dagger$ for each U holds almost by definition.

SU(N) as a subgroup of U(N)

For orthogonal matrices, on taking the determinant of the condition (1), $O^T O = I$, we obtain $(\det O)^2 = 1$, which implies the binary possibilities: $\det O = \pm 1$.

Choosing to impose the condition (2), $\det O = 1$, then allows us to eliminate all those transformations that involve reflections.

In contrast, for unitary matrices, on taking the determinant of the condition (3), $U^\dagger U = I$, we obtain $\det(U^\dagger U) = (\det U^\dagger)(\det U) = (\det U)^*(\det U) = |\det U|^2 = 1$, which implies a continuum of possibilities: $\det U = e^{i\alpha}$, for $0 \leq \alpha < 2\pi$.

The condition

$$\det U = 1 \tag{4}$$

eliminates this dangling phase factor $e^{i\alpha}$.

The two conditions (3) and (4), which are analogous to the two conditions (1) and (2), define the group $SU(N)$, a natural subgroup of $U(N)$.

An alternative way of saying this is as follows.

Among the groups $U(N)$, we have, in particular, the group $U(1)$ consisting of all 1×1 unitary matrices, that is, all complex numbers with absolute value equal to 1.

In other words, $U(1)$ consists of phase factors $e^{i\varphi}$, $0 \leq \varphi < 2\pi$.

The group $U(N)$ contains two sets of group elements.

The first set consists of unitary matrices of the form $e^{i\varphi}I$, where I denotes the $N \times N$ identity matrix.

These matrices clearly form a subgroup of $U(N)$; they satisfy all the group axioms.

Indeed, the matrix part (namely, I) is trivial: the different elements are distinguished by the phase factor $e^{i\varphi}$.

This is a wordy way of saying that this subgroup is in a one-to-one correspondence with the group $U(1)$ and might as well be called $U(1)$.

The second set of group elements of $U(N)$ consists of $N \times N$ unitary matrices with determinant equal to 1, namely, the special unitary matrices forming the subgroup $SU(N)$.

The matrices in this subgroup trivially commute with matrices of the form $e^{i\varphi}I$.

At this point, we might as well focus on the group $SU(N)$, defined by (3) and (4). (We will return to $U(N)$ later)

The story of $SU(N)$ follows more or less the same plot as the story of $SO(N)$, with the crucial difference that we will be dealing with complex rather than real numbers.

In discussing the orthogonal groups, I started with $SO(2)$ and $SO(3)$, and then proceeded to $SO(N)$. This order of presentation certainly makes sense, given that we live in 3-dimensional Euclidean space.

But in discussing the unitary groups, I reverse my pedagogical strategy and start with $SU(N)$, armed as it were with our experience with $SO(N)$.

From a general discussion of $SU(N)$, to be given here, we will then, in subsequent parts, specialize to $SU(2)$ and $SU(3)$, pointing out the special features that they do not share with $SU(N)$.

After our discussion of the orthogonal groups, that the “lower” unitary groups enjoy special features not granted to the “higher” unitary groups is hardly surprising.

As in the earlier discussion, this fact has to do with the number of indices the relevant antisymmetric symbols carry.

the

the

the

the

the

the

The Ω - and counting with our fingers

Let us now merrily repeat what we did back in chapter IV.1 to construct the tensor representations of $SO(N)$. (You might want to do a quick review.) Starting with the fundamental representation* $\psi^i \rightarrow \psi'^i = U^{ij} \psi^j$, we consider tensor $\varphi^{i_1 i_2 \cdots i_m}$ carrying m indices and transforming according to $\varphi^{i_1 i_2 \cdots i_m} \rightarrow \varphi'^{i_1 i_2 \cdots i_m} = U^{i_1 j_1} U^{i_2 j_2} \cdots U^{i_m j_m} \varphi^{j_1 j_2 \cdots j_m}$. The whole point of tensors, as you will recall, is that their indices transform independently,[†] so that any symmetry of $\varphi^{i_1 i_2 \cdots i_m}$ under permutations of its indices is carried over to $\varphi'^{i_1 i_2 \cdots i_m}$. (For example, if $\varphi^{i_1 i_2 \cdots i_5}$ is totally symmetric in the first three indices and antisymmetric in the last two indices, so too will $\varphi'^{i_1 i_2 \cdots i_5}$ be totally symmetric in the first three indices and antisymmetric in the last two indices.)

Ta dah!

We have thus constructed lots of representations of $SU(N)$.

Once again, we have to undergo the ordeal of mastering Young tableaux to study the most general symmetry patterns that a tensor of $SU(N)$ carrying m indices can exhibit.

And once again, fortunately, for many, perhaps even most, applications (say, in particle physics), we only have to deal with the simplest possible symmetry patterns, such as totally symmetric or totally antisymmetric, and with m small.

As an example, let us count the dimension of the representation furnished by the 3-indexed totally symmetric tensor φ^{ijk} of $SU(3)$.

Indeed, this counting is part of a Nobel prize winning piece of work!

Instead of watching the rich man going through his fancy footwork, let's watch the poor man simply enumerate:[‡] All indices equal to 3: $\{\varphi^{333}\}$; two indices equal to 3: $\{\varphi^{332}, \varphi^{331}\}$; one index equal to 3: $\{\varphi^{322}, \varphi^{321}, \varphi^{311}\}$; and no index equal to 3: $\{\varphi^{222}, \varphi^{221}, \varphi^{211}, \varphi^{111}\}$. We have altogether $1 + 2 + 3 + 4 = 10$ objects. Thus, $SU(3)$ has a 10-dimensional representation.

In the early 1960s, nine short-lived baryonic particles with similar properties and masses close to one another were known experimentally.

This suggests, as per the argument given earlier, that these nine particles belong to an irreducible representation of some symmetry group.

By “pure thought” (as shown here) and on guessing that the symmetric group of the strong interaction is $SU(3)$, Gell-Mann insisted that there must be ten, not nine, baryonic particles.

The dramatic discovery of the tenth baryonic particle, dubbed the Ω^- , confirming Gell-Mann's group theoretic prediction, was not only one of the most stirring episodes in the annals of particle physics, but also ushered in the widespread study of group theory among particle physicists.

Actually, we are only pretending to be poor; you and I can readily work out by induction the general formula for the dimension of $\{m\}$, the irreducible representation furnished by the totally symmetric tensor carrying m indices in $SU(N)$.

the

the

the

the

the

the

the

The trace for SU(N)

You ask, what about taking out the trace?

Good question.

Earlier, we had to subtract out the trace $\delta^{ij}\varphi^{ijk}$, so that the irreducible representation of SO(3) is $7 = (10 - 3)$ -dimensional, not 10-dimensional.

OK, let's do the same here, but first, let's check how the trace transforms: $\delta^{ij}\varphi^{ijk} \rightarrow \delta^{ij}(U^{if}U^{jg}U^{kh}\varphi^{fgh}) = (\delta^{ij}U^{if}U^{jg})(U^{kh}\varphi^{fgh})$.

But oops! What in the spinning world is $\delta^{ij}U^{if}U^{jg} = (U^T)^{fi}\delta^{ij}U^{jg} = (U^T U)^{fg}$?

It certainly is not δ^{fg} , which is what we need if $\delta^{ij}\varphi^{ijk}$ is to transform the way we want, namely, as a vector carrying the index k.

We are talking about unitary matrices here, and the combination $U^T U$ has no special meaning.

So, this particular “trace” does not make any mathematical sense.

We don't have to take out the trace, contrary to the discussion in the orthogonal case (there, $O^T O$ does have a special meaning, of course—namely, that it is equal to the identity matrix).

The representation $\{3\}$ of SU(3) is indeed irreducible, with dimension given by 10.

Gell-Mann got it right.

Fine, but this raises a more pressing question: how do we define a trace for $SU(N)$?

The clue—no surprise for those readers with mathematical sense!—is that it is the combination $U^\dagger U$, not $U^T U$, that is equal to the identity matrix.

We have to complex conjugate as well as transpose. (Physically, a quark is to be combined with an antiquark, not with another quark!)

Moving downstairs!

Now we come to something that has traditionally confused quite a few beginning students: the sudden appearance of two types of indices.

We could go through some fancy schmancy talk about covariant and contravariant, vector spaces and dual vector spaces, et cetera— what Einstein called “more-or-less dispensable erudition”—but we won’t; instead, we follow Einstein and adopt a low-brow practical approach, sticking to the nitty-gritty.

In linear algebra, the commonly used notation is complex vectors carry an upper index, while their complex conjugates carry a lower index.

But in view of the confusion alluded to above, it is worthwhile to go back to basics.

As noted at the beginning of this chapter, the quadratic invariant is $\zeta^\dagger \psi = \zeta^{*T} \psi = \sum_{j=1}^N \zeta^{j*} \psi^j$, not $\zeta^T \psi = \sum_{j=1}^N \zeta^j \psi^j$. Given that $\psi^i \rightarrow U^{ij} \psi^j$, we complex conjugate to obtain $\psi^{i*} \rightarrow (U^{ij})^* \psi^{j*}$. Next notice that this expression can be written as $\psi^{i*} \rightarrow (U^{T*})^{ji} \psi^{j*} = (U^\dagger)^{ji} \psi^{j*} = \psi^{j*} (U^\dagger)^{ji}$; nothing prevents us, in the last step, from interchanging the two complex numbers $(U^\dagger)^{ji}$ and ψ^{j*} . Indeed, we can now verify explicitly

that $\psi^{i*}\zeta^i \rightarrow \psi^{j*}(U^\dagger)^{ji}U^{ik}\zeta^k = \psi^{j*}\delta^{jk}\zeta^k = \psi^{j*}\zeta^j$ is invariant; after all, that was how we got the condition $U^\dagger U = I$ in the first place.

At this point, people invented a very clever notation: write

$$\psi_i \equiv \psi^{i*} \tag{5}$$

A new object ψ_i appears, but it is literally just ψ^{i*} , nothing mysterious about it at all.

At the least, we no longer have to write “*”.

It may not seem to you like all that much of a notational saving: not having to write “*”, we now have to write lower indices.

But the nice feature is that the invariant $\zeta^\dagger\psi = \sum_{j=1}^N \zeta^{j*}\psi^j$ can be written as $\zeta^\dagger\psi = \sum_{j=1}^N \zeta_j \psi^j = \zeta_j \psi^j$ (invoking the repeated index summation convention at the last step).

Thus, as this example shows, when we contract and sum over indices, with this new notation, we should impose the rule that we are allowed to contract and sum over an upper index with a lower index, but never an upper index with an upper index, nor a lower index with a lower index.

You will see shortly how this works in general.

the

To summarize, the defining or fundamental representation of $SU(N)$ consists of N objects $\psi^j, j = 1, \dots, N$, which transform under the action of the group element U according to

$$\psi^i \rightarrow \psi'^i = U^i_j \psi^j \quad (6)$$

Note that the column index on the matrix U has been typeset to the right of the row index, and lowered to satisfy our convention that the upper index on ψ^j is to be summed with a lower index.

Now we define an object written as ψ_i , which transforms in the same way as ψ^{*i} . The transformation $\psi^{i*} \rightarrow \psi^{j*}(U^\dagger)^{ji}$ we had earlier is then translated to $\psi_i \rightarrow \psi_j (U^\dagger)^j_i$. Thus, the counterpart of (6) is

$$\psi_i \rightarrow \psi'_i = \psi_j (U^\dagger)^j_i \quad (7)$$

As a check, $\zeta_i \psi^i \rightarrow \zeta'_i \psi'^i = \zeta_j (U^\dagger)^j_i U^i_k \psi^k = \zeta_j (U^\dagger U)^j_k \psi^k = \zeta_j \delta^j_k \psi^k = \zeta_j \psi^j$. Things work out nicely. Note that the Kronecker delta carries one upper and one lower index. This reinforces the fact (as we shall see shortly in an example) that we are not allowed to contract an upper index with an upper index, nor a lower index with a lower index.

In summary, upper indices transform with U , lower indices with $U^\dagger$.

Tensors with upper and lower indices

We now realize that the discussion earlier about tensors in $SU(N)$ was incomplete: in this notation, we must allow the tensors to carry lower indices as well as upper indices; we only had upper indices.

In general, we have $\varphi^{i_1 i_2 \dots i_m}_{j_1 j_2 \dots j_n}$ carrying m upper indices and n lower indices.

Instead of obscuring the fairly obvious with a general discussion, let us show how tensors work with a specific example.

The tensor φ_{ijk} transforms as

$$\varphi_k^{ij} \rightarrow \varphi_k'^{ij} = U_l^i U_m^j (U^\dagger)_k^n \varphi_n^{lm} = U_l^i U_m^j \varphi_n^{lm} (U^\dagger)_k^n \quad (8)$$

As at the end of the last section, upper indices transform with U , lower indices with $U^\dagger$.

In other words, it transforms as if it were equal to the product $\psi^i \psi^j \psi_k$; note that I did not say that φ_{ijk} is equal to $\psi^i \psi^j \psi_k$, merely that they transform in the same way.

We now know how to take a trace: set an upper index equal to a lower index and sum over them.

In our example, consider $\delta^k_j \varphi^{ij}_k \equiv \varphi^{ij}_j$. (Note that, as alluded to earlier, the fact that the Kronecker delta carries one upper and one lower index reminds us to contract and sum over an upper index with a lower index.)

It transforms as

$$\varphi_j^{ij} \rightarrow U_l^i U_m^j (U^\dagger)_j^n \varphi_n^{lm} = U_l^i \delta_m^n \varphi_n^{lm} = U_l^i \varphi_m^{lm} \quad (9)$$

where we have used (3).

In other words, the φ^{ij}_j (the trace of φ^{ij}_k), carrying the index $i=1, 2, \dots, N$, are N objects that transform into linear combinations of one another in the same way as a vector.

Thus, given a tensor, we can always subtract out its trace.

Now we see what we did wrong earlier.

We erroneously defined a trace by summing over two upper indices, which is not allowed in this formalism.

To summarize, traceless tensors with upper and lower indices furnish representations of SU(N).

The discussion proceeds in a way that should by now be familiar.

The symmetry properties of a tensor under permutation of its indices are not changed by the group transformation.

In other words, given a tensor $\varphi_{j_1 j_2 \dots j_n}^{i_1 i_2 \dots i_m}$ we can always require it to have definite symmetry properties under permutation of its upper indices (the i s) and under permutation of its lower indices (the j s).

Note that the upper indices and lower indices are to be permuted separately.

It does not make sense to interchange an upper index and a lower index; they are different beasts, so to speak.

In our specific example, we are allowed to take φ^{ij}_k to be either symmetric or anti-symmetric under the exchange of i and j and to be traceless. (Here, with one single lower index, we don't have to worry about the symmetry pattern of the lower indices.)

Thus,

the symmetric traceless tensor $S_k^{ij} = +S_k^{ji}$ furnishes a representation with dimension $\frac{1}{2}N^2(N+1) - N = \frac{1}{2}N(N-1)(N+2)$ and the antisymmetric traceless tensor $A_k^{ij} = -A_k^{ji}$ a representation with dimension $\frac{1}{2}N^2(N-1) - N = \frac{1}{2}N(N-2)(N+1)$.

Thus, in summary, the irreducible representations of $SU(N)$ are realized by traceless tensors with definite symmetry properties under permutation of indices. Convince yourself that for $SU(N)$, the dimensions of the representations defined by the following tensors φ^i , φ^{ij} (antisymmetric), φ^{ij} (symmetric), φ_j^i , and φ_k^{ij} (antisymmetric in the upper indices) are, respectively, N , $N(N-1)/2$, $N(N+1)/2$, N^2-1 , and $\frac{1}{2}N(N-2)(N+1)$.

You are now almost ready for the grand unified theories of the strong, weak, and electromagnetic interactions!

These are the irreducible representations commonly used in the popular Georgi-Glashow $SU(5)$ theory, with dimensions 5, 10, 15, 24, and 45, respectively.

As in the $SO(N)$ story, representations of $SU(N)$ have many names.

For example, we can refer to the representation furnished by a tensor with m upper and n lower indices as (m, n) .

Alternatively, we can refer to them by their dimensions, with a $*$ to distinguish representations with mostly lower indices from those with mostly upper indices.

For example, an alias for $(1, 0)$ is N , and for $(0, 1)$ is N^* .

Of course, two entirely different representations could happen to have the same dimension, but this ambiguity does not usually occur for representations with relatively low dimensions.

For example, in the popular SU(5) grand unified theory, the known quarks and leptons transform like 5 and 10^* .

A square bracket is often used to indicate that the indices are antisymmetric, and a curly bracket that the indices are symmetric.

Thus, the 10 of SU(5) is also known as $[2, 0] = [2]$, where as indicated, the 0 (no lower index) is suppressed.

Similarly, 10^* is also known as $[0, 2] = [2]^*$

Moving indices up and down stairs

The astute reader may have noticed that we have not yet used the condition (4), namely, the S in SU(N).

It can be written as either

$$\varepsilon_{i_1 i_2 \dots i_N} U_{i_1}^{i_1} U_{i_2}^{i_2} \dots U_{i_N}^{i_N} = 1 \quad (10)$$

or

$$\varepsilon^{i_1 i_2 \dots i_N} U_{i_1}^1 U_{i_2}^2 \dots U_{i_N}^N = 1 \quad (11)$$

Recall that this parallels what we did with $SO(N)$ earlier, except that here we have upper and lower indices.

Again, we can immediately generalize (10) to

$$\varepsilon_{i_1 i_2 \dots i_N} U_{j_1}^{i_1} U_{j_2}^{i_2} \dots U_{j_N}^{i_N} = \varepsilon_{j_1 j_2 \dots j_N} \quad (12)$$

Multiplying this identity by $(U^\dagger)_{p_N}^{j_N}$ and summing over j_N , we obtain

$$\varepsilon_{i_1 i_2 \dots i_{N-1} p_N} U_{j_1}^{i_1} U_{j_2}^{i_2} \dots U_{j_{N-1}}^{i_{N-1}} = \varepsilon_{j_1 j_2 \dots j_N} (U^\dagger)_{p_N}^{j_N} \quad (13)$$

Note that the index p_N is “exposed” and not summed over.

Clearly, by repeating this process, we can peel off as many U s on the left hand side as we like, and put them back as $U^\dagger$ s on the right hand side.

We can play a similar game with (11).

In contrast to $SO(N)$, we now have, not one, but two antisymmetric symbols $\varepsilon_{i_1 i_2 \dots i_N}$ and $\varepsilon^{i_1 i_2 \dots i_N}$, which we can use to raise and lower indices as our hearts desire.

To avoid drowning in a sea of indices, let me show how we can raise and lower indices in a specific example rather than in general.

Consider the tensor φ^{ij}_k in $SU(4)$.

We expect that the tensor $\varphi_{kpq} \equiv \varepsilon_{ijpq} \varphi_k^{ij}$ will transform as a tensor with three lower indices. We could trade two upper indices for two lower indices. Let's check that this indeed works: $\varphi_{kpq} \equiv \varepsilon_{ijpq} \varphi_k^{ij} \rightarrow \varepsilon_{ijpq} U_l^i U_m^j (U^\dagger)_k^n \varphi_n^{lm} = (U^\dagger)_p^s (U^\dagger)_q^t ((U^\dagger)_k^n \varepsilon_{lmst} \varphi_n^{lm}) = (U^\dagger)_p^s (U^\dagger)_q^t (U^\dagger)_k^n \varphi_{nst}$, as expected. Here we used a version of (13) to trade two U s for two $U^\dagger$ s. Similarly, you could verify that $\varphi^{ijklmn} \equiv \varepsilon^{klmn} \varphi_k^{ij}$ transforms like a tensor with five upper indices. Try a few more examples until you catch on.

To summarize, using the two (not one!) antisymmetric symbols, we can move indices on $SU(N)$ tensors up and down stairs.

From group to algebra

As before, Lie instructs us to solve (3) by writing $U \simeq I + iH$, with H some arbitrary “small” complex matrix. (Taking out a factor of i is merely a convenience, as we will see presently.)

To leading order in the parameter specifying the smallness of H , $U^\dagger U \simeq (I - iH^\dagger)(I + iH) \simeq I - i(H^\dagger - H) = I$.

Thus, (3) leads to $H^\dagger = H$, in other words, to the requirement that H is hermitean.

As in the case of $SO(N)$, by multiplying an infinitesimal transformation repeatedly by itself, we can write in general

$$U = e^{iH} \tag{14}$$

The statement is that U is unitary if H is hermitean.

It is worth checking this statement explicitly, although it is already implied by the preceding discussion. As explained in the review of linear algebra, the exponential of a matrix may be defined by a power series: $U = e^{iH} = \sum_{k=0}^{\infty} (iH)^k / k!$. Hermitean conjugating term by term, we have $U^\dagger = \sum_{k=0}^{\infty} (-iH^\dagger)^k / k! = \sum_{k=0}^{\infty} (-iH)^k / k! = e^{-iH}$, and thus $U^\dagger U = e^{-iH} e^{iH} = I$.

Next, to evaluate $\det U$ in order to impose the condition (4), we exploit the fact that H can always be diagonalized and write $H = W^\dagger \Lambda W$, with Λ the diagonal matrix $\text{diag}(\lambda_1, \lambda_2, \dots, \lambda_N)$ and W a unitary matrix. Then*

$$\begin{aligned} \det U &= \det e^{iH} = \det e^{iW^\dagger \Lambda W} = \det(W^\dagger e^{i\Lambda} W) = \det(W W^\dagger) \det e^{i\Lambda} \\ &= \det e^{i\Lambda} = \prod_{j=1}^N e^{i\lambda_j} = e^{i \sum_{j=1}^N \lambda_j} = e^{i \text{tr } \Lambda} = e^{i \text{tr } W^\dagger \Lambda W} \\ &= e^{i \text{tr } H} \end{aligned} \tag{15}$$

To obtain the third equality, expand the exponential in a power series and use the unitarity of W .

Thus, the condition $\det U = 1$ implies

$$\text{tr } H = 0 \tag{16}$$

This crucial result breaks the problem apart.

It says that all we have to do is to write down the most general $N \times N$ traceless hermitean matrix H .

To see how this works, it is easiest to proceed by examples.

For $N = 2$, the hermiticity condition $\begin{pmatrix} u & w \\ z & v \end{pmatrix}^\dagger = \begin{pmatrix} u^* & z^* \\ w^* & v^* \end{pmatrix} = \begin{pmatrix} u & w \\ z & v \end{pmatrix}$ implies that u and v are real and $w = z^*$, while the traceless condition gives $v = -u$. Thus, in general,

$$H = \begin{pmatrix} u & z^* \\ z & -u \end{pmatrix} = \frac{1}{2} \begin{pmatrix} \theta_3 & \theta_1 - i\theta_2 \\ \theta_1 + i\theta_2 & -\theta_3 \end{pmatrix} \quad (17)$$

where θ_1 , θ_2 , and θ_3 denote three arbitrary real numbers. (The factor of $\frac{1}{2}$ is conventional, due partly to historical reasons. I explain the reason for its inclusion in chapter IV.5.)

It is standard to define the three traceless hermitean matrices, known as Pauli matrices, as

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (18)$$

The most general 2-by-2 traceless hermitean matrix H can then be written as a linear combination of the three Pauli matrices: $H = \frac{1}{2}(\theta_1\sigma_1 + \theta_2\sigma_2 + \theta_3\sigma_3) = \sum_{a=1}^3 \frac{1}{2}\theta_a\sigma_a$. An element of $SU(2)$ can then be written as $U = e^{i\theta_a\sigma_a/2}$ (with the repeated index summation convention).

For $SU(3)$, we can go through analogous steps and define the eight 3 x 3 traceless hermitean matrices, known as Gell-Mann matrices,

$$\begin{aligned}
\lambda_1 &= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, & \lambda_2 &= \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, & \lambda_3 &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \\
\lambda_4 &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, & \lambda_5 &= \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, & & \\
\lambda_6 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, & \lambda_7 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, & \lambda_8 &= \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}
\end{aligned} \tag{19}$$

A general 3-by-3 traceless hermitean matrix and an element of $SU(3)$ can then be written, respectively, as $H = \theta_a \frac{\lambda_a}{2}$ and $U = e^{i\theta_a \frac{\lambda_a}{2}}$, where the index a now runs from 1 through 8. Note that the Pauli and Gell-Mann matrices are normalized by $\text{tr } \sigma_a \sigma_b = 2\delta_{ab}$ and $\text{tr } \lambda_a \lambda_b = 2\delta_{ab}$, respectively.

To make sure that you understand what is going on, you should write down a linearly independent set of 4-by-4 traceless hermitean matrices. (I regret to inform you, however, that physicists no longer have the habit of naming traceless hermitean matrices after anybody, including you.)

How many are there?

Indeed, how many are there for a general N ?

We simply go through mentally the steps we went through for $SU(2)$.

The N diagonal elements (the analogs of u and v) are real, and with the traceless condition, there are $N - 1$ of these.

The $1/2 N (N - 1)$ entries (each of which is specified by two real numbers) below the diagonal (namely, the analogs of z) are equal to the complex conjugate of the entries above the diagonal (the analogs of w).

Thus, the number of real numbers (namely, the analogs of θ_a) required to specify a general $N \times N$ traceless hermitean matrix is given by $(N - 1) + 2 \cdot 1/2 N (N - 1) = N^2 - 1$, which is of course also the number of linearly independent $N \times N$ traceless hermitean matrices.

The group $SU(N)$ is thus characterized by $N^2 - 1$ real parameters (the analog of the θ_a s), namely, 3, 8, 15, 24, 35, . . . for $N = 2, 3, 4, 5, 6, \dots$, respectively.

Compare this with the group $SO(N)$ characterized by $1/2 N (N - 1)$ real parameters, namely, 1, 3, 6, 10, 15, . . . for $N = 2, 3, 4, 5, 6, \dots$, respectively.

Notice that $SU(2)$ and $SO(3)$ are characterized by the same number of parameters.

File this fact away in your mind.

We will come back to this important observation later.

As I've said, my pedagogical plan is to continue discussing $SU(N)$ here, describing those aspects that hold for general N .

Then we will discuss, in the next two parts, $SU(2)$ and $SU(3)$ in turn, focusing on features specific to these two groups.

A trivial but potentially confusing point for some students.

We just went through a long song and dance about the importance of distinguishing upper and lower indices on tensors.

An upper index and a lower index transform quite differently (one might say oppositely) under $SU(N)$ transformations.

In contrast, the index a on the Pauli matrices is just a label, telling us which one of the three Pauli matrices we are talking about.

It is completely arbitrary whether we write σ_a or σ^a .

Similarly, the index a on the Gell-Mann matrices λ_a can be written as a subscript or superscript as we please.

Correspondingly, the “angles” θ_a , $a = 1, \dots, N^2 - 1$, can also be written with a superscript or a subscript.

I will intentionally not stick to a consistent convention to emphasize this point.

The structure constants of the Lie algebra

As just explained, any element of $SU(N)$ can be written as $U = e^{iH}$, with H a linear combination of $(N^2 - 1)$ linearly independent N -by- N hermitean traceless matrices T^a ($a = 1, 2, \dots, N^2 - 1$), so that

$$U = e^{i\theta^a T^a} \quad (20)$$

where θ^a are real numbers, and the index a is summed over. The T^a s are known as the generators of $SU(N)$. The index a , which ranges over $(N^2 - 1)$ values, is clearly not to be confused with the index i , which ranges over N values.

The discussion now parallels that in chapter I.3 for the orthogonal groups. As mentioned in chapter I.1, group elements, in general, do not commute. Lie proposed to capture this essence of group multiplication by focusing on infinitesimal elements. We are thus led to consider two simple unitary matrices, $U_1 \simeq I + A$ and $U_2 \simeq I + B$, near the identity. Just as in chapter I.3, the lack of commutativity is measured by the deviation of the quantity

$$\begin{aligned} U_2^{-1} U_1 U_2 &\simeq (I - B)(I + A)(I + B) \simeq I + A + AB - BA \\ &= I + A + [A, B] \end{aligned} \quad (21)$$

from $U_1 \simeq I + A$, namely, the commutator $[A, B]$. You should notice that this is formally identical to the manipulations in chapter I.3. Indeed, the only difference is that the generators of $SO(N)$ were called J , while the generators of $SU(N)$ are called T . So write $A = i \sum_a \theta^a T^a$ and similarly, $B = i \sum_b \theta'^b T^b$ with θ and θ' small. Hence $[A, B] = i^2 \sum_{ab} \theta^a \theta'^b [T^a, T^b]$, and it suffices to calculate the commutators $[T^a, T^b]$.

Recall from the review of linear algebra that the commutator $[H, K]$ of two hermitean matrices H and K is antihermitean and traceless (since $([H, K])^\dagger = [K^\dagger, H^\dagger] = [K, H] = -[H, K]$, and $\text{tr}[H, K] = \text{tr } HK - \text{tr } KH = 0$).

Since the commutator $[T^a, T^b]$ is antihermitean and traceless, it can also be written as a linear combination of the T^c s multiplied by i :

$$[T^a, T^b] = if^{abc}T^c \quad (22)$$

(with the index c summed over).

For an alternative derivation of (22), argue that since $(U_2U_1)^{-1}U_1U_2$ is an element of $SU(N)$, it must be equal to $I + i\varphi^cT^c + \dots$ for some φ^c .

The commutation relations (22) define the Lie algebra of $SU(N)$ and, as was mentioned earlier, the real numbers f^{abc} are known as the structure constants of the algebra.

Note that by construction, $f^{abc} = -f^{bac}$.

We will show generally later that the structure constant is in fact totally antisymmetric; it changes sign under the interchange of any pair of its three indices.

Given the Pauli matrices (18) and the Gell-Mann matrices (19), you can readily compute the structure constants f^{abc} for $SU(2)$ and $SU(3)$.

While physicists fill their brains with specific matrices, mathematicians use their minds to imagine abstract entities T^a satisfying (22).

We started this part defining $SU(N)$ as the group formed by unitary matrices with unit determinant.

In other words, we thought of the group elements as represented in the fundamental representation by $N \times N$ matrices.

Using tensors, we constructed higher-dimensional representations, in which the group elements U are represented by higher-dimensional matrices.

For example, for $SU(5)$, U can be represented by 10×10 , or 15×15 , or 24×24 matrices, and so on, as well as by 5×5 matrices.

For each of these representations, by considering elements near the identity, $U \simeq I + i\theta^a T^a$, we have a corresponding representation of T^a , for $a = 1, \dots, N^2 - 1$.

More pedantically and explicitly, consider a d -dimensional representation. The group element $U \simeq I + i\theta^a T^a$ is represented by the d -by- d matrix $D(I + i\theta^a T^a)$. Setting θ to 0, we find that $D(I) = I_d$, the d -by- d identity matrix. Hardly a surprise! Subtracting, or equivalently, differentiating with respect to θ^a , we obtain[†] $D(T^a)$, the d -by- d matrix representing T^a . For example, in the fundamental representation of $SU(2)$, the $D(T^a)$ are just the 2-by-2 Pauli matrices (with the “funny factor” of one half) $\frac{1}{2}\sigma_a$.

Thus, for $SU(5)$, we have a set of 24 10×10 matrices T^a constructed to satisfy (22). (Or 15×15 matrices, or 24×24 matrices, and so on.)

In particular, in the fundamental representation, we denote the entry in the i th row and j th column of the matrix T^a by $(T^a)_{ij}$.

At the risk of beating a point to death, note that a and i are entirely different beasts.

The label a , which ranges over $a = 1, \dots, N^2 - 1$, tells us which of the $N^2 - 1$ matrices we are talking about, while the indices i and j , which range from 1 to N , are matrix indices.

In light of these remarks, it is occasionally convenient, when working with a specific representation, to introduce yet another set of indices $p, q, r, \dots$ running from 1 to $d =$ the dimension of the representation being discussed.

For example, we might write for the representation 15 of $SU(5)$, 24 15×15 matrices $(T^a)_{pq}$ (thus, $a = 1, \dots, 24$, $p, q = 1, \dots, 15$).

Note that it is important to specify the group (of course) as well as the representation.

For example, the fundamental representation 15 of $SU(15)$ would consist of 224 1×15 matrices $(T^a)_{ij}$ (thus, $a = 1, \dots, 224$, $i, j = 1, \dots, 15$).

These two sets of 15-by-15 matrices have nothing to do with each other.

In a way, this is neither new nor amazing.

For $SO(3)$, for example, we have already learned that the three generators J_x , J_y , and J_z can be represented by three $(2j + 1) \times (2j + 1)$ matrices, for $j = 0, 1, 2, \dots$

Henceforth, we will follow the standard usage in the physics literature and will not distinguish between the generators T^a and the matrices $D(T^a)$ representing the generators.

Consider a tensor φ in an arbitrary d -dimensional irreducible representation.

Under a group transformation infinitesimally close to the identity, it changes by $\varphi \rightarrow (I + i\theta^a T^a)\varphi$, and thus its variation is given by $i\theta^a T^a \varphi$, with T^a as given in that irreducible representation, that is, by $d \times d$ matrices $(T^a)_{pq}$, with $p, q = 1, \dots, d$.

Again, we are at risk of belaboring a point, but some students are invariably confused here.

The tensor φ , written out explicitly as $\varphi_{j_1 j_2 \dots j_n}^{i_1 i_2 \dots i_m}$, carries m upper indices and n lower indices, each of which runs from 1 through N .

But because of various symmetry restrictions and of the traceless condition, these components of the tensor φ are not independent, and thus we often find it convenient to label the components of φ as φ^p , with p running from 1 through d , the dimension of the irreducible representation furnished by φ .

Then the change under the transformation may be written as

$$\delta\varphi^p = i\theta^a (T^a)^p_q \varphi^q \quad (23)$$

Note that for N , m , and n equal to some largish numbers, the dimension d can be rather large.

The adjoint representation of $SU(N)$

We've already talked about the adjoint representation in chapters I.3 and IV.1, but let us start our discussion of the adjoint representation of $SU(N)$ without reference to our previous knowledge and take a slightly different tack. We will see, in the end, that the discussions here and in chapter IV.1 are conceptually entirely the same.

For $SU(N)$, the fundamental (or defining) irreducible representation N and its conjugate N^* (defined by φ^i and φ_i , respectively) are of course of, well, fundamental importance. There is another irreducible representation of great importance, namely, the representation defined by the traceless tensor φ^i_j , known as the adjoint representation, or adjoint for short. As explained earlier, it has dimension $N^2 - 1$.

The poor man would make a flying guess here. In light of the earlier discussion, we are to find $(N^2 - 1)$ $(N^2 - 1)$ -by- $(N^2 - 1)$ matrices $(T^a)^p_q$ with $p, q = 1, \dots, N^2 - 1$. The poor man notices that the indices p, q have the same range as the index a . Guessing that they are all the same beasts, the poor man looks around for an object carrying three indices a, b , and c that could be conscripted to serve as $(T^a)^b_c$.

Good guess! The structure constants f^{abc} are practically staring you in the face.

The rich man proceeds more systematically. After all, that was just a wild guess, inspired though it may be. The rich man notes that, by definition, the adjoint transforms according

to $\varphi_j^i \rightarrow \varphi_j'^i = U_l^i (U^\dagger)^n_j \varphi_n^l = U_l^i \varphi_n^l (U^\dagger)^n_j$. We are thus invited to regard φ_j^i as a matrix transforming according to

$$\varphi \rightarrow \varphi' = U \varphi U^\dagger \quad (24)$$

Note that if φ is hermitean and traceless, it stays hermitean and traceless (since $\varphi'^\dagger = (U \varphi U^\dagger)^\dagger = U \varphi^\dagger U^\dagger = U \varphi U^\dagger = \varphi'$, and $\text{tr } \varphi' = \text{tr } U \varphi U^\dagger = \text{tr } \varphi U^\dagger U = \text{tr } \varphi$). Thus, with no loss of generality, we can take φ to be a hermitean traceless matrix. (If φ is antihermitean, we can always multiply it by i .) Another way of saying this is that given a hermitean traceless matrix X , then $U X U^\dagger$ is also hermitean and traceless if U is an element of $SU(N)$.

But if φ_j^i is regarded as a hermitean traceless matrix, then we can write it as a linear combination of the complete set of matrices T^a , $a = 1, 2, \dots, N^2 - 1$:

$$\varphi_j^i = \sum_{a=1}^{N^2-1} A^a (T^a)^i_j = A^a (T^a)^i_j \quad (25)$$

Here the A^a denote $(N^2 - 1)$ real coefficients. (The last step is simply a reminder that we use the repeated summation convention unless stated otherwise.) For example, for $SU(2)$, this just expresses the 2-by-2 hermitean traceless matrix φ_j^i as a linear combination of the three Pauli matrices.

We are free to think of the $(N^2 - 1)$ objects* A^a as furnishing the $(N^2 - 1)$ -dimensional adjoint representation; after all, the A^a are just linear combinations of the $(N^2 - 1)$ objects φ_j^i .

It is illuminating to work out how A^a transforms. Plugging the infinitesimal transformation $U \simeq 1 + i\theta^a T^a$ into the transformation rule (24), we have

$$\varphi \rightarrow \varphi' \simeq (1 + i\theta^a T^a)\varphi(1 + i\theta^a T^a)^\dagger \simeq \varphi + i\theta^a T^a \varphi - \varphi i\theta^a T^a = \varphi + i\theta^a [T^a, \varphi] \quad (26)$$

In other words, the a th generator acting on the adjoint representation gives $[T^a, \varphi]$. It is convenient to express (26) as giving the infinitesimal change of φ due to an infinitesimal transformation:

$$\delta\varphi = i\theta^a [T^a, \varphi] \quad (27)$$

To find out how A^a transforms, simply plug $\varphi = A^a T^a$ (namely, (25)) into (27) and use (22):

$$(\delta A^b)T^b = \delta(A^b T^b) = i\theta^a [T^a, A^c T^c] = i\theta^a A^c [T^a, T^c] = i\theta^a A^c i f^{acb} T^b \quad (28)$$

Equating the coefficient of T^b , we obtain $\delta A^b = -\theta^a f^{acb} A^c$. Comparing with the definition $\delta A^b = i\theta^a (T^a)^b_c A^c$, we obtain

$$(T^a)^{bc} = -i f^{abc} \quad (29)$$

Multiplying representations together

Given two tensors φ and η of $SU(N)$, with m upper and n lower indices and with m' upper and n' lower indices, respectively, we can consider a tensor T with $(m + m')$ upper and $(n + n')$ lower indices that transforms in the same way as the product $\varphi\eta$.

We can then reduce T by the various operations described earlier, namely, tracing and contracting with the antisymmetric symbol.

The multiplication of two representations is of course of fundamental importance in physics.

We will do it in detail for $SU(3)$ later.

U(1) in theoretical physics

I close by mentioning that while the group $U(1)$ is almost trivial, it is of prime importance in theoretical physics.

The group consists of elements $e^{i\theta}$ (with $2\pi > \theta \geq 0$) obeying the multiplication rule, $e^{i\theta} e^{i\theta'} = e^{i(\theta + \theta')}$, and hence commuting with one another.

The group theory is about as simple as it can get.

But electric charge conservation, one of the fundamental facts about the universe, corresponds to the invariance of physical laws under the multiplication of charged fields by some appropriate powers of $e^{i\theta(x)}$, with the angle θ a function of the spacetime coordinates x .

Earlier, I mentioned the far-reaching insight of Emmy Noether that conservation laws are associated with symmetries; charge conservation is one of the most celebrated examples.

SU(2): Double Covering and the Spinor

From electron spin to internal symmetries

The group $SU(2)$ came into physics with the dramatic discovery of electron spin.

Heisenberg, reasoning by analogy, then leaped from the notion of electron spin to the notion of isospin.

This epochal insight opens up a vast vista of internal symmetries, without which none of our modern physical theories could even be conceived.

In particular, we would not have been able to write down the nonabelian gauge theories that contemporary physics is entirely based on at the fundamental level.

Just as $SU(2)$ has played an honored role in quantum physics, it is also of great importance in mathematics, due to its many remarkable properties, such as its rather intricate relationship with $SO(3)$.

Historically, it is only after physicists had thoroughly mastered the mathematics of $SU(2)$ that they were able to move onto $SU(3)$ and beyond.

Every part of these notes is important, needless to say, but this part is particularly important.

I preview some of the highlights to be covered:

1. $SU(2)$ locally isomorphic to $SO(3)$
2. $SU(2)$ covers $SO(3)$ twice
3. irreducible representations of $SU(2)$
4. half angles versus full angles
5. the group elements of $SU(2)$
6. quantum mechanics and double-valued representations
7. $SU(2)$ does not have a downstairs, and its upstairs is all symmetric
8. pseudoreality
9. the precise relation between $U(2)$ and $SU(2)$

I dare to say that more than a vanishing fraction of theoretical physicists have not totally mastered some of these points, for example, 9.

$SU(2)$ is locally isomorphic to $SO(3)$

We start with one striking, I would even say astonishing, property of $SU(2)$: it is locally isomorphic to $SO(3)$.

The discussion to follow will make clear what these two words mean.

Any 2×2 hermitean traceless matrix X can be written as a linear combination of the three Pauli matrices, as explained earlier.

For convenience, I list the Pauli matrices again:

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

The statement is that we can write $X = \vec{x} \cdot \vec{\sigma}$ with three real coefficients, which we can assemble into an array $\vec{x} = (x^1, x^2, x^3)$ and provisionally call a vector. Or, call them the Cartesian coordinates (x, y, z) , if you prefer. (I do.) Explicitly,

$$X = x\sigma_1 + y\sigma_2 + z\sigma_3 = \begin{pmatrix} z & x - iy \\ x + iy & -z \end{pmatrix} \quad (2)$$

It is manifestly hermitean and traceless. Conversely, given any hermitean and traceless 2-by-2 matrix, we can write it in the form of (2) with the vector $\vec{x} = (x, y, z)$ uniquely determined.

Calculate the determinant of X : $\det X = -(x^2 + y^2 + z^2) = -\vec{x}^2$. Perhaps not surprisingly, the determinant of X , being an invariant, is given by the length squared¹ of $\vec{x}$. We now make this remark explicit.

Pick an arbitrary element U of $SU(2)$. In other words, U is unitary with determinant equal to 1. Consider $X' \equiv U^\dagger X U$.

I claim that X' is also hermitean and traceless. The proof is by explicit computation: $(X')^\dagger = (U^\dagger X U)^\dagger = U^\dagger X^\dagger U = U^\dagger X U = X'$, and $\text{tr } X' = \text{tr } U^\dagger X U = \text{tr } X U U^\dagger = \text{tr } X = 0$. Since X' is hermitean and traceless, we can write it as $X' = \vec{x}' \cdot \vec{\sigma}$, with three real coefficients (x', y', z') , which we also assemble into a vector $\vec{x}' = (x', y', z')$. The vector $\vec{x}'$ is linearly related to the vector $\vec{x}$: if we scale $\vec{x} \rightarrow \lambda \vec{x}$ by multiplying it by a real number λ , then $\vec{x}' \rightarrow \lambda \vec{x}'$.

Thus far, we have used the unitarity of U , but not the fact that it has unit determinant. In other words, we have used the U , but not the S , of $SU(2)$. You have done enough math exercises (or read enough detective mysteries) to know that S has to come in somehow somewhere sometime. Here and now in fact! Compute $\det X' = -(\vec{x}')^2 = \det U^\dagger X U = (\det U^\dagger)(\det X)(\det U) = \det X = -\vec{x}^2$. Thus, $\vec{x}'$ and $\vec{x}$ are not only linearly related to each other, but they also have the same length.

By definition, this means that the 3-vector $\vec{x}$ is rotated into the 3-vector $\vec{x}'$. Since X (and hence $\vec{x}$) is arbitrary, this defines a rotation R . In other words, we can associate an element R of $SO(3)$ with any element U of $SU(2)$.

The map $U \rightarrow R$ clearly preserves group multiplication: if two elements U_1 and U_2 of $SU(2)$ are mapped to the rotations R_1 and R_2 , respectively, then the element $U_1 U_2$ is mapped to the rotation $R_1 R_2$. We can readily verify this: $(U_1 U_2)^\dagger X (U_1 U_2) = U_2^\dagger (U_1^\dagger X U_1) U_2 = U_2^\dagger X' U_2 = X''$, so that the 3-vectors associated with X , X' , and X'' get rotated in stages, $\vec{x} \rightarrow \vec{x}' \rightarrow \vec{x}''$, first by R_1 and then by R_2 .

$SU(2)$ covers $SO(3)$ twice

Interestingly, this map $f: U \rightarrow R$ of $SU(2)$ into $SO(3)$ is actually 2-to-1, since U and $-U$ are mapped into the same R ; $f(U) = f(-U)$.

This is not obvious to you? Well, $(-U)^\dagger X (-U) = U^\dagger X U$.

In $SU(2)$, U and $-U$ are manifestly not the same: indeed, $U = -U$ would imply, on multiplication by $U^\dagger$, $I = -I$, which is clearly absurd.

The unitary group $SU(2)$ is said to double cover the orthogonal group $SO(3)$.

Since the map $U \rightarrow R$ is 2-to-1, the groups $SU(2)$ and $SO(3)$ are not isomorphic, but only locally isomorphic.

In particular, for U near the identity, $-U$ is far from the identity, and so in a neighborhood around the identity, the two groups are isomorphic.

I alluded to this local isomorphism between $SU(2)$ and $SO(3)$ as astonishing, but in fact, we have already encountered several hints of this.

For example, earlier, I asked you to file away in your mind the apparently coincidental fact that both groups have a 3-dimensional representation.

Now we see that this is not a coincidence: identifying the 3 of $SU(2)$ with the vector of $SO(3)$ in fact defines the local isomorphism.

Properties of the Pauli matrices

To proceed further, it is convenient to pause and list some “arithmetical” properties of the three Pauli matrices.

You should verify the following statements as you read along.

By inspection, we see that each of the Pauli matrices squares to the identity matrix: $(\sigma_a)^2 = I$, for $a = 1, 2, 3$.

Distinct Pauli matrices anticommute with each other: $\sigma_a \sigma_b = -\sigma_b \sigma_a$ for $a \neq b$.

The product of any two distinct Pauli matrices gives the third with a coefficient given by $\pm i$; thus, $\sigma_1 \sigma_2 = i\sigma_3$, with all other cases given by cyclic permutation and the fact that Pauli matrices anticommute.

These facts are summarized by (with the repeated index c summed)

$$\sigma_a \sigma_b = \delta_{ab} I + i \varepsilon_{abc} \sigma_c \quad (3)$$

Here I denotes the 2×2 identity matrix; it will often be omitted if there is no risk of confusion.

Interchanging a and b , and then adding and subtracting the resulting equation to and from (3) gives

$$\{\sigma_a, \sigma_b\} = 2\delta_{ab} \quad (4)$$

and

$$[\sigma_a, \sigma_b] = 2i \varepsilon_{abc} \sigma_c \quad (5)$$

Note the factors of 2.

Earlier, we showed that the Lie algebra for $SU(N)$ is given by $[T^a, T^b] = if^{abc} T^c$.

Dividing (5) by 2, we obtain

$$\left[\frac{\sigma_a}{2}, \frac{\sigma_b}{2} \right] = i \varepsilon_{abc} \frac{\sigma_c}{2} \quad (6)$$

This shows that for the fundamental 2-dimensional representation, the generators T^a are represented by $\sigma_a/2$, a fact often written somewhat sloppily as $T^a = \sigma_a/2$, and that the structure constant of $SU(2)$ is just the antisymmetric symbol, that is, $f^{abc} = \varepsilon^{abc}$.

As remarked earlier, whether the indices $a, b, \dots$ are written as subscripts or superscripts is of no import.

Again, I have capriciously moved the indices $a, b, \dots$ up- and downstairs to stress this point.

Also, I had emphasized that the T^a may be regarded as abstract mathematical entities.

In a d -dimensional representation, they are realized as $d \times d$ matrices.

In particular, in the fundamental representation, they are represented by $\sigma_a/2$.

Confusio: “So, T^3 is sometimes a 4×4 matrix and sometimes a 17×17 matrix?”

Exactly! It depends on the phase of the moon.

The Lie algebra of $SU(2)$ is thus given by $[T^a, T^b] = i\varepsilon^{abc} T^c$; in other words,

$$[T^1, T^2] = iT^3, \quad [T^2, T^3] = iT^1, \quad [T^3, T^1] = iT^2 \quad (7)$$

For comparison, the Lie algebra of $SO(3)$ was given earlier as

$$[J_x, J_y] = iJ_z, \quad [J_y, J_z] = iJ_x, \quad [J_z, J_x] = iJ_y \quad (8)$$

We see immediately that, with the identification $T^1 \leftrightarrow J_x$, $T^2 \leftrightarrow J_y$, $T^3 \leftrightarrow J_z$, the two Lie algebras are manifestly identical, that is, isomorphic.

In fact, who cares which letters we use to denote the generators of SU(2) and SO(3)?

It is the same algebra, period.

Representing the Lie algebra of SU(2)

Given the preceding remark, we can now find the representations of the Lie algebra of SU(2) without doing a stitch of work.

Simply take over what we did earlier for the Lie algebra of SO(3), replacing the letter J by T if you insist. Define $T^\pm \equiv T^1 \pm iT^2$ so that

$$[T^3, T^\pm] = \pm T^\pm, \quad [T^+, T^-] = 2T^3 \quad (9)$$

and proceed.

We discover soon enough that the representations of SU(2) are $(2j + 1)$ -dimensional, with $j = 0, 1/2, 1, 3/2, 2, 5/2, \dots$

Earlier, I introduced to you the representations of the algebra of SO(3) corresponding to $j = \text{half integer}$ as a disturbing puzzle.

But now we see why they pop up.

They are representations of $SU(2)$!

In particular, the 2-dimensional representation with $j = 1/2$ consists of the states $| -1/2 \rangle$ and $| 1/2 \rangle$.

It is the fundamental or defining representation of the group $SU(2)$.

More on this shortly.

The group elements of $SU(2)$

Before writing down the group elements of $SU(2)$, we need a useful formula. Given two 3-vectors $\vec{u}$ and $\vec{v}$, contract (3) with u^a and v^b to obtain

$$(\vec{u} \cdot \vec{\sigma})(\vec{v} \cdot \vec{\sigma}) = u^a v^b \sigma_a \sigma_b = u^a v^b (\delta_{ab} I + i \varepsilon_{abc} \sigma_c) = (\vec{u} \cdot \vec{v}) I + i (\vec{u} \otimes \vec{v}) \cdot \vec{\sigma} \quad (10)$$

The dot and cross products (or the scalar and vector products) familiar from elementary vector analysis both appear on the right hand side. This result shows the convenience of thinking of the three Pauli matrices as a vector $\vec{\sigma}$. Also, it follows that $X^2 = x^a x^b \sigma_a \sigma_b = \vec{x}^2$.

A crucial step in proving, in the preceding section, that the 3-vector $\vec{x}$ is rotated into the 3-vector $\vec{x}'$, involves showing that they have the same length. We did this by calculating the determinant of X and the determinant of $X' \equiv U^\dagger X U$. We now see that an alternative method is to note that $\text{tr } X^2 = \text{tr } X'^2$. (In fact, the trace is not even necessary, since X^2 is proportional to the identity matrix.)

We already noted earlier that any element of $SU(2)$ can be written as $U =$

$e^{i\varphi_a\sigma_a/2} = e^{i\vec{\varphi}\cdot\vec{\sigma}/2}$ (with the repeated index summation convention). Denote the magnitude and direction of $\vec{\varphi}$ by φ and the unit vector $\hat{\varphi}$, respectively (so that $\vec{\varphi} = \varphi\hat{\varphi}$). Use $(\vec{\varphi} \cdot \vec{\sigma})^2 = \vec{\varphi}^2 = \varphi^2$. We now expand the exponential series into even and odd powers (compare with (I.3.12)) and sum to obtain

$$\begin{aligned} U &= e^{i\vec{\varphi}\cdot\vec{\sigma}/2} = \sum_{n=0}^{\infty} \frac{i^n}{n!} \left(\frac{\vec{\varphi} \cdot \vec{\sigma}}{2} \right)^n \\ &= \left\{ \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} \left(\frac{\varphi}{2} \right)^{2k} \right\} I + i \left\{ \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} \left(\frac{\varphi}{2} \right)^{2k+1} \right\} \hat{\varphi} \cdot \vec{\sigma} \\ &= \cos \frac{\varphi}{2} I + i \hat{\varphi} \cdot \vec{\sigma} \sin \frac{\varphi}{2} \end{aligned} \quad (11)$$

We see that this is structurally the same as Euler's identity $e^{i\varphi/2} = \cos \frac{\varphi}{2} + i \sin \frac{\varphi}{2}$, which can also be derived by splitting the exponential series into even and odd terms.

An easier way to derive (11) is to note that we are free to call the axis of rotation $\hat{\varphi}$ the 3rd or z-axis. Then $\hat{\varphi} \cdot \vec{\sigma} = \sigma_3$, and since σ_3 is diagonal, the exponential can be immediately evaluated by Euler's identity:

$$\begin{aligned} U &= e^{i\varphi\sigma_3/2} = \begin{pmatrix} e^{i\frac{\varphi}{2}} & 0 \\ 0 & e^{-i\frac{\varphi}{2}} \end{pmatrix} = \begin{pmatrix} \cos \frac{\varphi}{2} + i \sin \frac{\varphi}{2} & 0 \\ 0 & \cos \frac{\varphi}{2} - i \sin \frac{\varphi}{2} \end{pmatrix} \\ &= \cos \frac{\varphi}{2} I + i\sigma_3 \sin \frac{\varphi}{2} \end{aligned} \quad (12)$$

For the general case, we simply replace σ_3 by $\hat{\varphi} \cdot \vec{\sigma}$ and obtain (11).

Confusio: “For years, I confused $U = \cos \frac{\varphi}{2} I + i \hat{\varphi} \cdot \vec{\sigma} \sin \frac{\varphi}{2}$ and $X = \vec{x} \cdot \vec{\sigma}$ (as defined in (2)) merely because they looked similar.”

Indeed, they are quite different: U is unitary and hence an element of $SU(2)$, but it is in general not traceless, while, in contrast, X is hermitean and traceless.

Incidentally, back in the review of linear algebra, I said that the polar decomposition was not particularly attractive. For example, we can write any 2-by-2 complex matrix as $M = (t + \vec{x} \cdot \vec{\sigma}) e^{i\theta} e^{i \frac{\vec{\varphi}}{2} \cdot \vec{\sigma}} = (t + \vec{x} \cdot \vec{\sigma}) (\cos \frac{\varphi}{2} + i \sin \frac{\varphi}{2} \hat{\varphi} \cdot \vec{\sigma}) e^{i\theta}$. A complex matrix is characterized by $2 \cdot 2^2 = 8$ real parameters, namely, t , $\vec{x}$, $\vec{\varphi}$, and θ .

How the half angles become full angles

We had proved earlier that if $X' = \vec{x}' \cdot \vec{\sigma}$ and $X = \vec{x} \cdot \vec{\sigma}$ are related by $X' = U^\dagger X U$, then the two vectors $\vec{x}'$ and $\vec{x}$ are related by a rotation. By construction, it must work. But still, now that we know U explicitly, it is instructive to see how it actually works—how all the wheels and gears fit together, so to speak. In particular, even though we now see how the half angles in (11) come in (to render the Lie algebras of $SU(2)$ and $SO(3)$ the same), it does seem strange at first sight that (11) can lead to the usual expression for rotations

So, let us evaluate $U^\dagger X U$ by brute force. As just noted, with no loss of generality, we can take $\hat{\varphi}$ to point along the third axis. So, use (12). For $a = 3$, we obtain trivially $U^\dagger \sigma_3 U = \sigma_3$. For $a = 1, 2$, $U^\dagger \sigma_a U = (\cos \frac{\varphi}{2} I - i \sigma_3 \sin \frac{\varphi}{2}) \sigma_a (\cos \frac{\varphi}{2} I + i \sigma_3 \sin \frac{\varphi}{2})$. To obtain with σ_1 and σ_2 and squares to the identity). Another line of arithmetic (or algebra, whatever you want to call it) gives, for $a = 1, 2$, $U^\dagger \sigma_a U = (\cos^2 \frac{\varphi}{2} - \sin^2 \frac{\varphi}{2}) \sigma_a - i \sin \frac{\varphi}{2} \cos \frac{\varphi}{2} [\sigma_3, \sigma_a]$.

From (5), we have $[\sigma_3, \sigma_1] = 2i\sigma_2$ and $[\sigma_3, \sigma_2] = -2i\sigma_1$. If you remember your trigonometric identities (the “double angle formulas”) $\cos \varphi = \cos^2 \frac{\varphi}{2} - \sin^2 \frac{\varphi}{2}$ and $\sin \varphi = 2 \sin \frac{\varphi}{2} \cos \frac{\varphi}{2}$, you can practically see the answer emerging:

$$U^\dagger \sigma_1 U = \cos \varphi \sigma_1 + \sin \varphi \sigma_2, \quad U^\dagger \sigma_2 U = -\sin \varphi \sigma_1 + \cos \varphi \sigma_2 \quad (13)$$

We are not quite there yet; we still have to plug in $X' = U^\dagger(x\sigma_1 + y\sigma_2 + z\sigma_3)U = (\cos \varphi x - \sin \varphi y)\sigma_1 + (\sin \varphi x + \cos \varphi y)\sigma_2 + z\sigma_3$. In other words, $x' = \cos \varphi x - \sin \varphi y$, $y' = \sin \varphi x + \cos \varphi y$, $z' = z$, precisely what we should get* for a rotation around the z -axis through angle φ .

Hey, math works. The half angle $\varphi/2$ becomes the full angle φ .

Confusio: “I have often read the sloppy and potentially confusing statement that the Pauli matrices transform like a vector.”

Indeed, the statement is sloppy, but sometimes useful as a shorthand for (13). Physically, we can talk about two different observers recording the same physical quantity (such as momentum) as $\vec{p}$ and as $\vec{p}'$, but there is no such thing as σ'_a ; we simply have a bunch of 2-by-2 matrices with entries like 0, ± 1 , and $\pm i$. There exists only one Pauli, so to speak.

Quantum mechanics and the double covering

Now comes a striking fact.

From (12) we have $U(\varphi) = e^{i\varphi\sigma_3/2}$, which, as we just checked, leads to a rotation through an angle φ around the z -axis.

But note that

$$U(2\pi) = e^{i(2\pi)\sigma_3/2} = e^{i\pi\sigma_3} = \begin{pmatrix} e^{i\pi} & 0 \\ 0 & e^{-i\pi} \end{pmatrix} = -I \quad (14)$$

By the time the rotational angle φ has gone from 0 to 2π , so that the corresponding rotation has gone back to the identity, the SU(2) element U has reached only -I .

To reach I , the angle φ has to go all the way to 4π :

$$U(4\pi) = I \quad (15)$$

This striking fact, as expressed by (14) and (15), of course just restates what we had learned earlier, that SU(2) double covers SO(3).

By the time we get around SU(2) once, the corresponding rotation has gone around SO(3) twice.

Another way of saying this is as follows. The 2-dimensional fundamental representation $\psi^i \rightarrow U_{ij} \psi^j$ of SU(2) is strictly speaking not a representation of SO(3). (Some authors call it a double-valued representation of SO(3).)

This discussion once again underlines the fact that the groups SU(2) and SO(3) are not globally isomorphic, only locally isomorphic; in other words, only their algebras are isomorphic.

SU(2) does not have a downstairs, and its upstairs is all symmetric

In our discussion of the orthogonal groups, I mentioned that SO(3) has special properties that are not shared by SO(N), a fact that sometimes confuses beginning students.

Similarly, the two special unitary groups that most students learn first, namely, $SU(2)$ and $SU(3)$, have special properties that are not shared by $SU(N)$, essentially for the same reasons as in the orthogonal case.

Earlier, we learned that traceless tensors $T^{i_1 i_2 \cdots i_m}_{j_1 j_2 \cdots j_n}$ with upper and lower indices furnish representations of $SU(N)$.

Furthermore, these tensors are to have definite symmetry properties under permutation of their upper indices (the i 's) and under permutation of their lower indices (the j 's).

For $SU(2)$, because the antisymmetric symbols ϵ^{ij} and ϵ_{ij} carry two indices, we can in fact remove all lower indices.

To see how this works, it is clearer not to treat the general case, but to pick a specific example, say, the tensor T^{ijk}_{mn} . (Each of the indices can take on two values, 1 or 2, of course.)

Then, following the discussion from earlier, we can construct $T^{pqijk} = \epsilon^{pm} \epsilon^{qn} T^{ijk}_{mn}$, which transforms like a tensor with five upper indices and no lower index.

Thus, it suffices to consider only tensors without lower indices.

We have used ϵ^{ij} but we still have ϵ_{ij} up our sleeves.

We now claim further that it suffices to consider only tensors with upper indices all symmetrized.

The argument is inductive (just as in our discussion of the orthogonal groups).

Suppose this holds for tensors with less than four upper indices.

Now we look at tensors with four indices.

Suppose T^{ijkl} has no particular symmetry on the interchange of, say, i and k .

We simply write $S^{ijkl} = T^{ijkl} + T^{kjil}$ and $A^{ijkl} = T^{ijkl} - T^{kjil}$.

Then $\varepsilon_{ik}A^{ijkl}$ is a tensor with two upper indices, which, by induction, is not something we need to worry about.

By construction, S^{ijkl} is of course symmetric on the interchange of i and k .

Repeating this argument leads to the bottom line that we need consider only tensors with upper indices and that do not change sign on the interchange of any pair of indices.

Confusio: “Why doesn’t this argument work for $SU(3)$?” I will let the reader answer Confusio’s question.

Ironically, it is harder, in some sense, to understand $SU(2)$ than $SU(5)$, say, precisely because $SU(2)$ is so “simple” and thus has idiosyncratic features.

Dimension of $SU(2)$ irreducible representations

What is the dimension of the representation furnished by $T^{i_1 i_2 \cdots i_m}$? Just count: $T^{11 \cdots 1}$, $T^{11 \cdots 12}$, $T^{11 \cdots 122}$, $\dots$, $T^{22 \cdots 2}$. The number of 2s ranges from 0 to m , and hence this representation has dimension $m + 1$. To make contact with the earlier discussion, simply write m as $2j$. Since m ranges over the non-negative integers, we have $j = 0, \frac{1}{2}, 1, \frac{3}{2}, 2, \dots$.

To summarize, an irreducible representation of $SU(2)$ is characterized by j , which can take on integral or half-integral values and has dimension $2j + 1$. Note that $j = \frac{1}{2}$ is the fundamental representation, while $j = 1$ is the vector representation. The existence of the vector representation offers another way of seeing the local isomorphism of $SU(2)$ and $SO(3)$.

Not complex, only pseudoreal

Earlier, I went over carefully the reason for having both lower and upper indices.

You learned that we had to introduce ψ_i as a stand-in, so to speak, for ψ^{i*} .

In general, for $SU(N)$, the representations furnished by ψ_i and ψ^i are quite different.

But then we just learned that for $SU(2)$, we don't need lower indices, and in particular, ψ_i is not needed.

Indeed, we just learned $\varepsilon^{ij}\psi_j$ transforms in exactly the same way as ψ^i .

What is going on?

Does this mean that for SU(2), the representations furnished by ψ^{i*} and ψ^i are the same, which would seem to suggest that complex conjugation “doesn’t do anything” in some sense?

However, we see from (11) that the fundamental representation is explicitly not real.

A clue to what is going on may be found in earlier results, where we learned about real, pseudoreal, and complex representations.

Recall that if a representation $D(g)$ does not look real, that is, even if $D(g)^* \neq D(g)$, the representation and its complex conjugate could be “secretly” related by a similarity transformation: that is, there may exist an S such that

$$D(g)^* = S D(g) S^{-1} \tag{16}$$

for all $g \in G$.

The representation is then said to be pseudoreal.

The fundamental representation of SU(2) is pseudoreal.

We now show this by explicitly finding the similarity transformation S . The fundamental or defining representation $D(g)$ is given by $e^{i\vec{\varphi}\vec{\sigma}/2}$, while its conjugate representation $D(g)^*$ is given by $(e^{i\vec{\varphi}\vec{\sigma}})^* = e^{-i\vec{\varphi}\vec{\sigma}^*}$.

But of the three Pauli matrices (1), σ_1 and σ_3 are manifestly real, while σ_2 is manifestly imaginary.

We could say that σ_2 is the “odd man out.”

Using the facts that different Pauli matrices anticommute and that the Pauli matrices square to the identity, we have $\sigma_2\sigma_1^*\sigma_2 = \sigma_2\sigma_1\sigma_2 = -\sigma_1\sigma_2\sigma_2 = -\sigma_1$.

Similarly, $\sigma_2\sigma_3^*\sigma_2 = -\sigma_3$, and $\sigma_2\sigma_2^*\sigma_2 = -\sigma_2\sigma_2\sigma_2 = -\sigma_2$.

Thus, we obtain

$$\sigma_2\sigma_a^*\sigma_2 = -\sigma_a \tag{17}$$

and so

$$\sigma_2(e^{i\vec{\varphi}\vec{\sigma}})^*\sigma_2 = e^{i\vec{\varphi}\vec{\sigma}} \tag{18}$$

Comparing this with (16), we see that we can choose $S = S^{-1} = \sigma_2$, which is anti-symmetric, in accordance with the proof in earlier that S has to be either symmetric or antisymmetric.

Indeed, the fundamental representation of $SU(2)$ is pseudoreal.

We note that (16), together with the condition $S^\dagger S = I$ (also proved earlier) determines $S = e^{i\alpha}\sigma_2$ only up to a phase factor.

It is then convenient to stick in an i and write $S = i\sigma_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$.

We suddenly see the connection of this discussion with the tensor approach: the S written here is precisely the ε^{ij} that connects ψ_i to ψ^i .

The groups $U(N)$ and $SU(N)$

In the preceding part, we mentioned that the group $U(N)$ contains two sets of group elements.

The first set consists of unitary matrices of the form $e^{i\varphi}I$, composing the group $U(1)$.

The second set consists of $N \times N$ unitary matrices with determinant equal to 1, composing the group $SU(N)$.

At this point, almost all physicists would say that $U(N)$ is equal to the direct product $SU(N) \otimes U(1)$.

But this is incorrect.

Nobody said that the two sets just mentioned have a nontrivial intersection.

The elements of $U(1)$ of the form $e^{i2\pi k/N} I$ with $k = 1, \dots, N-1$ have determinant equal to 1 and hence also belong to the group $SU(N)$.

These elements form the finite group Z_N .

Recall the definition, given earlier, of the direct product group $H \equiv F \otimes G$, consisting of the elements (f, g) , with f and g elements of the groups F and G , respectively.

The group F consists of the elements of H of the form (f, I_G) , and G of the elements of H of the form (I_F, g) .

The only element belonging to both F and G is the identity element of H , namely, $I_H = (I_F, I_G)$. (Here we are using the insufferably pedantic language in which we distinguish between the identity elements I_H , I_F , and I_G of the three groups H , F , and G , respectively, and which was alluded earlier.)

Since we already included the elements of Z_N in $U(1)$, to avoid overcounting, we have to exclude them from $SU(N)$.

Thus, we should actually write

$$U(N) = \left(SU(N)/Z_N \right) \otimes U(1) \quad (19)$$

In the quotient group $SU(N)/Z_N$, the N elements $e^{i2\pi k/N}U$, with U an arbitrary unitary matrix with unit determinant and with $k=0, 1, \dots, N-1$, are to be identified as one and the same element.

For example, for $N=2$, the two matrices U and $-U$ are actually the same element, but this is precisely the group $SU(2)/Z_2 = SO(3)$, as discussed earlier in this part.

The center of a group is defined as the set of all elements that commute with all the other elements of the group.

Thus, Z_N is actually the center of $SU(N)$.

We will discuss the physical meaning of the center of $SU(3)$ later.

At the level of Lie algebra, there is in fact no distinction between $SU(N)$ and $SU(N)/Z_N$: the algebra explores only the neighborhood of the identity of the group.

Put another way, to determine the algebra of $U(N)$, write an element as $U = e^{iH} \simeq I + iH$.

Since U is not required to have unit determinant, H is hermitean but is not required to be traceless.

Thus, the generators of the algebra are the N^2-1 traceless hermitean matrices plus the identity matrix I .

As mentioned earlier, the theory of the strong, weak, and electromagnetic interactions is based on $SU(3) \otimes SU(2) \otimes U(1)$.

But in typical applications, such as those involving the interaction between gauge bosons, the global properties of the group do not come in at all; only the Lie algebra enters into the field theory Lagrangian.

Henceforth, we will follow the sloppy physicists and not distinguish between $U(N) = SU(N)/Z_N \otimes U(1)$ and $SU(N) \otimes U(1)$.

The Electron Spin and Kramer's Degeneracy

The mystique of spin 1/2

Niels Bohr solved the hydrogen atom in 1913 by boldly postulating that angular momentum is quantized in integral units of $\hbar$.

Physicists were shocked, shocked, shocked.

But then their minds were really blown in 1925, when George Uhlenbeck and Sam Goudsmit proposed that the electron spins with angular momentum 1/2.

A priori, it would seem that any representation in which a rotation through 2π is represented by $-I$ would be worthless for physics, but the incredible mystery of quantum mechanics saved the angular momentum 1/2 representation from oblivion.

The wave function of a particle can perfectly well change sign upon a 2π rotation, since physical observables are postulated to be bilinears constructed out of $\Psi^\dagger$ and Ψ , such as the spin vector $\Psi^\dagger \frac{\vec{\sigma}}{2} \Psi$ (see later), which manifestly does not change sign.

Now we know that not only the electron, but also quarks and leptons—the fundamental constituents of matter—all carry spin 1/2 .

To many theoretical physicists, spin 1/2 continues to carry a certain mystique.

As will be explained later, spin $1/2$ particles are described by the strikingly beautiful Dirac equation.

Several decades later, when Gell-Mann first introduced quarks, many in the theoretical physics community were skeptical, and some were even hostile.

For one thing, the fractional electric charges carried by quarks went against preconceived notions.

Dirac was among the few eminences who liked quarks.

When Gell-Mann asked him why, Dirac replied, “They carry spin $1/2$, don’t they?”

The electron spin

The discovery that the electron has spin $1/2$ is surely one of the most stirring episodes in physics.

Consider how confused people were in those early days of quantum mechanics, with the mess of quantum numbers.

Then they realized that to explain the spectroscopic data, a mysterious additional quantum number had to be introduced.

Even after electron spin was proposed, belief would have to hinge on one’s philosophical stance.

An empirical positivist would demand to know how a spinning particle could carry spin $1/2$ and what that could mean.

The mathematical pragmatist would say that a particle whose quantum state transforms like the 2-dimensional representation of the covering group of rotations carries spin $1/2$, period.

In a galaxy far far away, quite possibly physicists knew all about the double covering of rotations and were able to predict the possible occurrence of spin $1/2$.

In fact, in our own civilization, there were people who knew, but they were called mathematicians.

Possibly, a few physicists, such as Hamilton, might also have known.

Communication between mathematicians and physicists, then as now, was blocked by language barriers and any number of other obstacles.

The electron wave function

The proposal that the electron has spin $1/2$ means that the electron wave function ψ has two components and that it transforms like $\psi \rightarrow e^{i\vec{\varphi}\cdot\vec{\sigma}/2}\psi$.

But, as was already noted earlier, for a rotation through 2π around the z-axis,

$$\psi \rightarrow e^{i(2\pi)\sigma_3/2}\psi = \begin{pmatrix} e^{i\pi} & 0 \\ 0 & e^{-i\pi} \end{pmatrix} \psi = -\psi \quad (1)$$

But a rotation through 2π is naively no rotation at all, so how can something flip sign?

We have to appeal to the magic of quantum mechanics.

The measurable quantities in quantum mechanics are all bilinear in ψ and its complex conjugate $\psi^\dagger$, such as the probability density $\psi^\dagger\psi$.

Thus, it is physically acceptable for ψ to flip sign under a rotation through 2π .

We now show explicitly that the electron spin does what we expect it to do.

Spin precession

An electron sitting at rest satisfies the Schrödinger equation (set $\hbar = 1$ to lessen clutter)

$$i \frac{\partial}{\partial t} \Psi(t) = H \Psi(t) \quad (2)$$

Since the electron is not moving, $\Psi_\alpha(t)$ does not depend on its position $\vec{x}$ but does depend on an index α indicating that it transforms like the defining representation of $SU(2)$. The spin angular momentum of the electron is then given by $\vec{S} = \Psi^\dagger \frac{\vec{\sigma}}{2} \Psi$. Impose an external magnetic field $\vec{B}$, so that the Hamiltonian is given by

$$H = \mu \vec{B} \cdot \frac{\vec{\sigma}}{2} = \frac{1}{2} \mu (B_1 \sigma_1 + B_2 \sigma_2 + B_3 \sigma_3) \quad (3)$$

with μ a measure of the magnetic moment. The Hamiltonian, effectively a 2-by-2 matrix, breaks rotation invariance explicitly, since $\vec{B}$ picks out a direction. Let $\vec{B}$ be time independent for simplicity; then the Schrödinger equation is immediately solved by

$$\Psi(t) = e^{-i\mu \vec{B} \cdot \frac{\vec{\sigma}}{2} t} \Psi(0) = U(t) \Psi(0) \quad (4)$$

where the unitary time evolution matrix $U(t)$ is precisely the U studied in (IV.5.11) with $\vec{\varphi} = -\mu \vec{B}t$ now linear in time.

Indeed, calling the direction of the magnetic field the third axis, we simply lift (IV.5.12) here, and obtain

$$\Psi(t) = \begin{pmatrix} e^{-i\frac{\mu Bt}{2}} & 0 \\ 0 & e^{i\frac{\mu Bt}{2}} \end{pmatrix} \Psi(0) = \left(\cos \frac{\mu Bt}{2} I - i \sin \frac{\mu Bt}{2} \sigma_3 \right) \Psi(0) \quad (5)$$

The magnetic field simply multiplies the upper and lower components by a relative phase.

Note once again that, when a spin $\frac{1}{2}$ particle is rotated through 2π , after a time $t = 2\pi(\mu B)^{-1}$, its wave function changes sign.

From (IV.5.13), we see that the spin $\vec{S}$ precesses, as would be expected for a classical magnetic moment:

$$\begin{aligned} \vec{S}_1(t) &= \cos \mu Bt \vec{S}_1(0) - \sin \mu Bt \vec{S}_2(0) \\ \vec{S}_2(t) &= \sin \mu Bt \vec{S}_1(0) + \cos \mu Bt \vec{S}_2(0) \\ \vec{S}_3(t) &= \vec{S}_3(0) \end{aligned} \quad (6)$$

After a time $t = 2\pi(\mu B)^{-1}$, the spin $\vec{S}$ indeed returns to its original direction.

Since this is a class on group theory rather than quantum mechanics, I will restrain myself from going into the many interesting phenomena involving the electron spin since we already did that.

Time reversal and antiunitary operator

In a famous paper in 1932, Wigner showed that time reversal is represented by an anti-unitary operator.

Since this subject is potentially confusing, I will go slow and let the physics, as embodied in the equations, lead us.

Take the Schrödinger equation $i \frac{\partial}{\partial t} \Psi(t) = H \Psi(t)$. For definiteness, think of $H = -\frac{1}{2m} \nabla^2 + V(\vec{x})$, just simple one-particle nonrelativistic quantum mechanics. We suppress the dependence of Ψ on $\vec{x}$.

Consider the transformation $t \rightarrow t' = -t$. We want to find a $\Psi'(t')$ such that

$$i \frac{\partial}{\partial t'} \Psi'(t') = H \Psi'(t') \quad (7)$$

Write $\Psi'(t') = T \Psi(t)$, where T is some operator to be determined (up to some arbitrary phase factor η). We require Ψ' to satisfy (7) if Ψ satisfies (2).

Plugging in, we have $i \frac{\partial}{\partial(-t)} T \Psi(t) = H T \Psi(t)$. Multiplying by T^{-1} , we obtain $T^{-1}(-i) T \frac{\partial}{\partial t} \Psi(t) = T^{-1} H T \Psi(t)$. Since H does not involve time, we want $HT = TH$, that is, $T^{-1}HT = H$. Then $T^{-1}(-i) T \frac{\partial}{\partial t} \Psi(t) = H \Psi(t)$. If T does what it is supposed to do, then we are forced to conclude, as Wigner was, that

$$T^{-1}(-i)T = i \quad (8)$$

Speaking colloquially, we can say that in quantum physics, time goes with an i , and so flipping time means you have to flip i also.

Let $T = UK$, where K complex conjugates everything to its right. Using $K^2 = I$, we have $T^{-1} = KU^{-1}$, and so (8) holds if $U^{-1}iU = i$, that is, if U^{-1} is just an ordinary (unitary) matrix that does nothing to i . We will determine U as we go along. The presence of K makes T antiunitary.

We check that this works for a spinless particle in a plane wave state $\Psi(t) = e^{i(\vec{p} \cdot \vec{x} - Et)}$. The formalism gives us $\Psi'(t') = T \Psi(t) = UK \Psi(t) = U \Psi^*(t) = U e^{-i(\vec{p} \cdot \vec{x} - Et)} = \eta e^{-i(\vec{p} \cdot \vec{x} - Et)}$; since Ψ has only one component, U is 1-by-1 and hence is just a phase factor* η , which we can choose to be 1. Rewriting, we have (note the argument of Ψ' !)

$\Psi'(t) = e^{-i(\vec{p}\cdot\vec{x}+Et)} = e^{i(-\vec{p}\cdot\vec{x}-Et)}$. Indeed, Ψ' describes a plane wave with wave vector $(-\vec{p})$ that moves in the opposite direction, as we would expect from time reversal. Note crucially that $\Psi'(t) \propto e^{-iEt}$ and thus has positive energy, as it should.⁶ Note also that acting on a spinless particle, $T^2 = UKUK = UU^*K^2 = +1$, as we would expect.

Kramer's degeneracy

Next consider a spin $\frac{1}{2}$ nonrelativistic electron, which we can take to be at rest for simplicity. The spin observable is given by $\vec{S} = \psi^\dagger(\frac{1}{2}\vec{\sigma})\psi$, with ψ the 2-component spinor of the electron. Under the time reversal operator $T = UK$, $\vec{S} \rightarrow \frac{1}{2}\psi^\dagger(UK)^\dagger\vec{\sigma}UK\psi = \frac{1}{2}\psi^\dagger KU^\dagger\vec{\sigma}UK\psi$. We wish this to be $-\vec{S}$, a wish that would come true if $U = \eta\sigma_2$ with some phase factor η , since

$$KU^\dagger\vec{\sigma}UK = \eta^*\eta K\sigma_2\vec{\sigma}\sigma_2K = K \begin{pmatrix} -\sigma_1 \\ +\sigma_2 \\ -\sigma_3 \end{pmatrix} K = -\vec{\sigma} \quad (9)$$

Everything fits!

The presence of K is crucial.

Note that the pseudoreality of the 2-dimensional spinor (and defining) representation of SU(2) is crucial here.

In contrast to the spin 0 case, in the spin 1/2 case, we need a 2 x 2 matrix $U = \eta\sigma_2$ to flip the spin.

The key observation is that, acting on a spin 1/2 particle,

$$T^2 = \eta \sigma_2 K \eta \sigma_2 K = \eta \sigma_2 \eta^* \sigma_2^* K K = -1 \quad (10)$$

Two successive time reversals do not give the identity!

Reversing the flow of time twice does not get a spin 1/2 particle back to itself.

This is the origin of Kramer's degeneracy, which states the following.

For an electron moving in an electric field, no matter how complicated the field, each of the energy levels must be two-fold degenerate.

The proof is very simple.

Since electric fields do not change under time reversal (think of the electric field as that produced between two charged plates), the Hamiltonian is time reversal invariant ($H T = T H$), and so Ψ and $T\Psi$ have the same energy.

Next, ask whether Ψ and $T\Psi$ represent two distinct states or just one single state.

Suppose they actually represent the same state; this means that $T\Psi$ is proportional to Ψ .

Thus, $T T\Psi = e^{i\alpha} \Psi$, with $e^{i\alpha}$ some phase factor.

But then

The time of mysteries and eeriness

It is difficult to convey to contemporary readers how confusing the half-integral electron spin was at the time; the only way would be to quote the distinguished physicists of the time.

For example, C. G. Darwin, who published an early paper on electron spin with Pauli, wrote,

“It is rather disconcerting to find that apparently something has slipped through the net, so that physical quantities exist which would be . . . artificial and inconvenient to express as tensors.”

Paul Ehrenfest in 1932 wrote of

“this eerie report that a mysterious tribe by the name of spinor family inhabit isotropic [three-dimensional] space and the Einstein-Minkowski [four-dimensional] world.”

From the Lagrangian to Quantum Field Theory: It Is but a Skip and a Hop

Field theory: The bare bones minimum

Here I attempt the nearly impossible, to convey the essence of quantum field theory in a few pages. All we need here, and for later use, is merely the rudiments of field theory.

I will try to convey the bare bones minimum in the simplest possible terms.

Just read this impressionistically.

It is of course understood that the discussion here is only a caricature of a serious discussion.

It suffices to have a vague impression of the material presented here.

For the convenience of some readers, I provide an executive summary at the end of this chapter.

One step at a time.

As described earlier, Heisenberg promoted q and p to operators satisfying

$$[q, p] = i \tag{1}$$

Dirac then introduced creation and annihilation operators, respectively, $a^\dagger = \frac{1}{\sqrt{2}}(q - ip)$ and $a = \frac{1}{\sqrt{2}}(q + ip)$, which according to (1) satisfy $[a, a^\dagger] = \frac{1}{2}[q + ip, q - ip] = -i[q, p] = 1$. For the harmonic oscillator, the Hamiltonian $H = \frac{1}{2}\omega a^\dagger a$ is proportional to the number operator. In Heisenberg's formulation of quantum mechanics, an operator O evolves in time according to $\frac{dO}{dt} = i[H, O]$. In particular, the position operator has the time dependence

$$q(t) = \frac{1}{\sqrt{2}}(ae^{-i\omega t} + a^\dagger e^{i\omega t}) \tag{2}$$

Note that $q(0) = q$.

At this breakneck pace, we are already almost at the door of quantum field theory. I assume that the reader is familiar with the concept of a field, the electric field $\vec{E}(t, \vec{x})$, for example. For our purposes, we can define a field as a dynamical variable that also depends on the spatial coordinates $\vec{x}$.

From discrete to continuous labels

The progression from particle to field¹ may be thought of as a two-step process:

$$q(t) \rightarrow q_a(t) \rightarrow \varphi(t, \vec{x}) \quad (3)$$

First, we replace the single particle by N particles labeled by $a = 1, 2, \dots, N$. Second, we promote the discrete label a to the continuous label $\vec{x}$ (and move it to live inside the parenthesis). By the way, the switch from q to φ for the dynamical variable is merely a bow to tradition.

The important point is to understand that the spatial coordinate $\vec{x}$ is a mere label, not a dynamical variable. Thus, the notation $\vec{E}(t, \vec{x})$ tells us that it is the time dependence of the electric field at $\vec{x}$ we are focusing on: for each $\vec{x}$, $\vec{E}(t, \vec{x})$ is the dynamical variable, not² $\vec{x}$.

In the many-particle case, the Lagrangian would contain terms linking the different particles, such as $\sim -\frac{1}{2}k(q_a - q_b)^2$, due to springs connecting them, for example. After the discrete variable a got promoted to the continuous variable $\vec{x}$, these terms take on the form³ $\sim (\varphi(\vec{x} + \vec{\varepsilon}) - \varphi(\vec{x}))^2 \sim \varepsilon^i (\frac{\partial \varphi}{\partial x^i})^2$; at an intermediate step, we might think of space as a lattice with a microscopic lattice spacing ε , which in the end we will take to zero, rendering space continuous.

In the intermediate stage, the Lagrangian is given as a sum $L = \sum_a(\cdots)$. In the continuum limit, the sum over a becomes an integral over $\vec{x}$. The Lagrangian ends up being given by an integral over space $L = \int d^3x \mathcal{L}$, with $\mathcal{L}$ known as the Lagrangian density. If you have followed the “poetry” in the preceding paragraphs, you could perhaps see that the Lagrangian density assumes the form

$$\mathcal{L} = \frac{1}{2} \left(\frac{\partial \varphi}{\partial t} \right)^2 - \frac{c^2}{2} \sum_{i=1}^3 \left(\frac{\partial \varphi}{\partial x^i} \right)^2 - V(\varphi) \quad (4)$$

The first term is just the usual kinetic energy. A certain amount of cleaning up has already been performed. For example, a constant multiplying $\frac{1}{2}(\frac{\partial \varphi}{\partial t})^2$ can be absorbed into the definition of φ . The lattice spacing ε has been similarly absorbed. Here c is needed for the dimension to come out right; hence it has the dimension of length over time, that is, that of a speed. In a relativistic theory, c is the speed of light. Units are usually chosen (with length measured in lightseconds rather than in terms of some English king’s foot, for example) so that $c = 1$. In other words, we can absorb c by scaling $\vec{x}$.

It is common practice to abuse language somewhat and to refer to $\mathcal{L}$ as the Lagrangian, dropping the word density. Note that the action $S = \int dt L = \int dt d^3x \mathcal{L} = \int d^4x \mathcal{L}$ may be written as an integral of the Lagrangian (density) $\mathcal{L}$ over 4-dimensional spacetime.

In a word, field theory is what you get if you’ve got an oscillator at each point $\vec{x}$ in space.⁴ The field $\varphi(t, \vec{x})$ describes the amplitude of the oscillation at time t of the oscillator located at $\vec{x}$. The term $(\frac{\partial \varphi}{\partial x^i})^2$ links the oscillators at neighboring points in space.

Packaging the time and space coordinates into spacetime coordinates $x^\mu = (t, x^i)$ with $x^0 = t$ and introducing the compact notation $\partial_\mu \equiv \frac{\partial}{\partial x^\mu} = (\frac{\partial}{\partial t}, \frac{\partial}{\partial x^i})$, we can write the Lagrangian density in (4) as

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \varphi)^2 - V(\varphi) \quad (5)$$

Here we are anticipating the discussion of relativistic physics in chapter VII.2, devoted to the Lorentz group. (Not to strain the reader even further, I would say that the simplest tactic here would be to regard⁵ $(\partial_\mu \varphi)^2$ as shorthand for the combination $(\frac{\partial \varphi}{\partial t})^2 - \sum_{i=1}^3 (\frac{\partial \varphi}{\partial x^i})^2$ in (4).)

Recall the Euler-Lagrange equation

$$\frac{d}{dt} \left(\frac{\delta L}{\delta \frac{dq}{dt}} \right) = \frac{\delta L}{\delta q} \quad (6)$$

from chapter III.3. When we jump from mechanics to field theory, $q(t)$ is replaced by $\varphi(t, \vec{x})$. Thus, I presume that you would not be surprised to learn that the Euler-Lagrange equation is naturally extended to include space derivatives as well as a time derivative, thus replacing the left hand side of (6) by $\partial_\mu \left(\frac{\delta \mathcal{L}}{\delta \partial_\mu \varphi} \right) \equiv \frac{\partial}{\partial t} \left(\frac{\delta \mathcal{L}}{\delta \frac{\partial \varphi}{\partial t}} \right) + \frac{\partial}{\partial x^i} \left(\frac{\delta \mathcal{L}}{\delta \frac{\partial \varphi}{\partial x^i}} \right)$. For a field theory, we have the Euler-Lagrange equation

$$\partial_\mu \left(\frac{\delta \mathcal{L}}{\delta \partial_\mu \varphi} \right) = \frac{\delta \mathcal{L}}{\delta \varphi} \quad (7)$$

with the corresponding equation of motion

$$\left(\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial \vec{x}^2} \right) \varphi + V'(\varphi) = 0 \quad (8)$$

In the simplest case $V(\varphi) = \frac{1}{2}m^2\varphi^2$ (which we might think of as the harmonic approximation), then the equation of motion becomes $(\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial \vec{x}^2} + m^2)\varphi = 0$, which is solved by $\varphi(t, \vec{x}) \propto e^{-i(\omega_k t - \vec{k} \cdot \vec{x})}$, with $\omega_k^2 = \vec{k}^2 + m^2$.

To quantize this field theory, we “merely” have to follow Heisenberg’s prescription. The momentum conjugate to $\varphi(t, \vec{x})$ is $\pi(t, \vec{x}) \equiv \frac{\delta \mathcal{L}}{\delta \frac{\partial \varphi}{\partial t}} = \frac{\partial \varphi}{\partial t}$. (This is almost an immediate generalization of $p(t) = m \frac{dq}{dt}$.) The end result, not surprisingly, is that the field can again be expanded, generalizing (2), in terms of creation and annihilation operators:

$$\varphi(t, \vec{x}) = \int \frac{d^3k}{\sqrt{(2\pi)^3 2\omega_k}} \left(a(\vec{k}) e^{-i(\omega_k t - \vec{k} \cdot \vec{x})} + a(\vec{k})^\dagger e^{i(\omega_k t - \vec{k} \cdot \vec{x})} \right) \quad (9)$$

The assumption that φ is real, or more precisely, hermitean ($\varphi^\dagger = \varphi$), fixes the second term in (9) in terms of the first.

A big difference is that the creation and annihilation operators, respectively, $a(\vec{k})^\dagger$ and $a(\vec{k})$, are now functions of $\vec{k}$. Evidently, the dependence of φ on $\vec{x}$ gets translated into a dependence of $a^\dagger$ and a on $\vec{k}$. Physically, $a(\vec{k})^\dagger$ acting on the vacuum state $|0\rangle$ creates a

state containing a particle of momentum $\vec{k}$ and energy $\omega_k = +\sqrt{\vec{k}^2 + m^2}$. (Note that we have set $\hbar = 1$ throughout; the momentum and energy are actually $\hbar\vec{k}$ and $\hbar\omega_k$, respectively.) Incidentally, the normalization factor $\sqrt{(2\pi)^3 2\omega_k}$ in (9) is chosen so that this state is normalized conventionally.⁶

Fields and symmetry groups

Since this part of the class is on group theory rather than quantum field theory, we are less interested in the dynamics of fields, about which volumes could be, and have been, written, but more in how symmetries are incorporated into field theory.

We have already talked about rotational invariance and relativistic invariance.

We mentioned that Heisenberg, by introducing isospin, opened up a vast vista of internal symmetries, so that physicists have many more symmetries to play with beyond these spacetime symmetries.

The exploration of these internal symmetries turned out to be a central theme in subatomic physics.

Here we discuss the group theoretic framework in general and come to specific examples later.

A symmetry group G is postulated, deduced from experimental observation or theoretical considerations, with various fields furnishing various representations $\mathcal{R}_1, \mathcal{R}_2, \dots$ of G (and of the Lorentz group of special relativity).

The task at hand is then to construct the Lagrangian (density).

By definition, that G is a symmetry means that the Lagrangian is invariant under the transformations of G .

Thus, the mathematical problem is to simply construct, out of the $\mathcal{R}$ s, an object $\mathcal{L}$ that does not change under G .

Let us hasten from this rather airy yak yak yak to a specific example.

Suppose that $G = \text{SU}(N)$ (as discussed earlier) and that our theory contains two fields, φ_{ij} as a symmetric tensor with two lower indices) and η^k (belonging to the defining representation).

Then $\varphi_{ij}\eta^i\eta^j$ is an invariant term we can include in L , describing the interaction of a particle of the φ type with two particles of the η type.

Physically, we can see that the Lagrangian, after we expand the fields ϕ and η into creation and annihilation operators, contains a term that would annihilate a ϕ particle and create two η particles.

This describes the decay of a φ particle into two η particles.

We now elaborate on these remarks.

Lagrangians with internal symmetries

The simplest example is a $U(1)$ invariant theory containing a field φ belonging to the defining representation: under the group element $e^{i\theta}$, $\varphi \rightarrow e^{i\theta}\varphi$. Since $\varphi^\dagger \rightarrow e^{-i\theta}\varphi^\dagger$, the hermiticity condition $\varphi = \varphi^\dagger$ cannot be maintained. Necessarily, $\varphi \neq \varphi^\dagger$, and φ is a nonhermitean field, commonly referred to as a complex field.

An immediate consequence is that we have to write

$$\varphi(t, \vec{x}) = \int \frac{d^3k}{\sqrt{(2\pi)^3 2\omega_k}} \left(a(\vec{k}) e^{-i(\omega_k t - \vec{k} \cdot \vec{x})} + b(\vec{k})^\dagger e^{i(\omega_k t - \vec{k} \cdot \vec{x})} \right) \quad (10)$$

That $\varphi^\dagger \neq \varphi$ compels us to introduce two sets of creation and annihilation operators, $(a, a^\dagger)$ and $(b, b^\dagger)$. In contrast to (9), the second term here is not related to the first term.

You should verify that the two creation operators, $a^\dagger$ and $b^\dagger$, transform oppositely under the $U(1)$. If we say that the state $a^\dagger(\vec{k}) |0\rangle$ describes a particle of momentum $\vec{k}$, then the state $b^\dagger(\vec{k}) |0\rangle$ describes an antiparticle⁸ of momentum $\vec{k}$.

Electromagnetism in fact exhibits a $U(1)$ symmetry, as explained in any quantum field theory textbook. The two states just described carry equal but opposite electric charges.

Since $U(1)$ is isomorphic to $SO(2)$, a theory of a complex field can also be written in terms of two real fields, just like a complex number can be written in terms of two real numbers $z = x + iy$. Simply set

$$\varphi = \frac{1}{\sqrt{2}}(\varphi_1 + i\varphi_2) \quad (11)$$

Each of the two real fields, φ_1 and φ_2 , can be expanded as in (9), thus confirming the need for two sets of creation and annihilation operators, $(a_1, a_1^\dagger)$ and $(a_2, a_2^\dagger)$.

The important factor of $\frac{1}{\sqrt{2}}$ in (11) can be understood as follows. Call the states $a_1^\dagger |0\rangle$ and $a_2^\dagger |0\rangle$ created by φ_1 and φ_2 , respectively, $|1\rangle$ and $|2\rangle$. Let these states be normalized⁹ such that $\langle 1|1\rangle = 1$ and $\langle 2|2\rangle = 1$. If φ as defined in (11) is expanded as in (10), then the creation operator $a^\dagger$ in φ creates the state $|s\rangle = \frac{1}{\sqrt{2}}(|1\rangle + |2\rangle)$. The factor $\frac{1}{\sqrt{2}}$ here comes from the $\frac{1}{\sqrt{2}}$ in (11). The state $|s\rangle$ is then normalized correctly: $\langle s|s\rangle = \frac{1}{2}(\langle 1|1\rangle + \langle 2|2\rangle) = \frac{1}{2}(1 + 1) = 1$.

From this discussion it is but a short leap to field theories symmetric¹⁰ under some group G . The various fields appearing in the Lagrangian $\mathcal{L}$ transform under various irreducible representations of G , and they are to be combined in such a way that $\mathcal{L}$ is invariant. In other words, $\mathcal{L}$ transforms as a singlet under G .

All the known internal symmetry groups in physics are compact. Thus, the irreducible representations that the known fields belong to are unitary. In other words, under the group element g , the field φ transforms like $\varphi \rightarrow D(g)\varphi$, with $D(g)^\dagger = D(g)^{-1}$. This means that the mass term $m^2\varphi^\dagger\varphi$, and the “kinetic” terms $\frac{\partial\varphi^\dagger}{\partial t}\frac{\partial\varphi}{\partial t}$ and $\frac{\partial\varphi^\dagger}{\partial x^i}\frac{\partial\varphi}{\partial x^i}$, are always allowed.¹¹

Given the symmetry group G and the collection of irreducible representations realized as fields in the theory, the construction of the invariant Lagrangian is thus a group theoretic problem. We multiply the various irreducible representations together to form a singlet.

Concrete examples will soon be given.

An executive summary

Given the lightning speed with which I introduced quantum field theory, it would be good to provide an executive summary here:

1. The coordinates $q(t)$ in mechanics are promoted to a collection of fields $\varphi(t, \vec{x})$.
2. Under an internal symmetry group G , the fields $\varphi(t, \vec{x})$ belong to various irreducible representations and transform accordingly.
3. The Lagrangian density $\mathcal{L}$, constructed out of the fields $\varphi(t, \vec{x})$, must be invariant under G .
This requirement poses a group theoretic problem.

Multiplying Irreducible Representations of Finite Groups: Return to the Tetrahedral Group

Now that we know how to multiply irreducible representations together and to Clebsch-Gordan decompose, we can go back and do these things for finite groups.

It is particularly easy if the finite group is naturally a subgroup of a continuous group, about which we know a lot. A good example is given by the tetrahedral group $T = A_4$, discussed in chapter II.2. We learned there that T , a subgroup of $SO(3)$, has four irreducible representations, 1 , $1'$, $1''$, and 3 . Let us now work out $3 \otimes 3$ in T .

In $SO(3)$, $3 \otimes 3 = 1 \oplus 3 \oplus 5$. But 5 does not exist in T . Restricting $SO(3)$ to T , we expect the 5 to decompose either as $5 \rightarrow 3 \oplus 1 \oplus 1$ or as $5 \rightarrow 3 \oplus 1' \oplus 1''$. Note that since 5 is real, if its decomposition contains a complex representation, such as $1'$, it must also contain the conjugate of that representation.

It is instructive to work out the decomposition explicitly. Given two vectors $\vec{u}$ and $\vec{v}$ of $SO(3)$, the 3 is of course given by the vector cross product $\vec{u} \times \vec{v}$, while the 5 , the symmetric traceless tensor, consists of the three symmetric combinations $u_2v_3 + u_3v_2$, $u_3v_1 + u_1v_3$, and $u_1v_2 + u_2v_1$, together with the two diagonal traceless combinations $2u_1v_1 - u_2v_2 - u_3v_3$ and $u_2v_2 - u_3v_3$.

The last two guys are clearly the “odd men out.” Referring back to the discussion of $A_4 = T$ in chapter II.3, recall that the Z_3 subgroup of A_4 is generated by cyclic permutation on the irreducible representation 3 : for example, $u_1 \rightarrow u_2 \rightarrow u_3 \rightarrow u_1$ (or in the reverse order). Under these cyclic permutations, $u_2v_3 + u_3v_2$ transforms into his two friends, thus generating the 3 in the product $3 \otimes 3$, while the two odd guys transform into linear combinations of each other, for example, $u_2v_2 - u_3v_3 \rightarrow u_3v_3 - u_1v_1$. This also means that we cannot have $5 \rightarrow 3 \oplus 1 \oplus 1$. Thus, we have established that $5 \rightarrow 3 \oplus 1' \oplus 1''$.

Therefore, in $T = A_4$,

$$3 \otimes 3 = 1 \oplus 3 \oplus 3 \oplus 1' \oplus 1'' = 1 \oplus 1' \oplus 1'' \oplus 3 \oplus 3$$

The second equality merely reflects our desire to write things more “symmetrically.”

To see how the two odd guys correspond to $1' \oplus 1''$, recall that in our construction of the character table for A_4 in (II.3.15) the cube root of unity $\omega \equiv e^{i2\pi/3} = -\frac{1}{2} + \frac{\sqrt{3}}{2}i$, the number naturally associated with Z_3 , plays a prominent role. Some useful identities are

$$1 + \omega + \omega^2 = 0, \quad \omega^2 = \omega^* \quad (2)$$

so that

$$\begin{pmatrix} 1 \\ \omega \\ \omega^2 \end{pmatrix} + \begin{pmatrix} 1 \\ \omega^2 \\ \omega \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix} \quad (3)$$

and

$$\begin{pmatrix} 1 \\ \omega \\ \omega^2 \end{pmatrix} - \begin{pmatrix} 1 \\ \omega^2 \\ \omega \end{pmatrix} = \sqrt{3}i \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \quad (4)$$

Thus, the $1'$ and $1''$ can be taken to be

$$1' \sim q' = u_1 v_1 + \omega u_2 v_2 + \omega^2 u_3 v_3 \quad (5)$$

and

$$1'' \sim q'' = u_1 v_1 + \omega^2 u_2 v_2 + \omega u_3 v_3 \quad (6)$$

The combinations $2u_1v_1 - u_2v_2 - u_3v_3$ and $u_2v_2 - u_3v_3$ contained in the 5 of $SO(3)$ correspond to $q' + q''$ and $q' - q''$.

It is perhaps worth emphasizing that while $1'$ and $1''$ furnish 1-dimensional representations of A_4 , they are not invariant under A_4 . For example, under the cyclic permutation c , $q' \rightarrow \omega q'$ and $q'' \rightarrow \omega^2 q''$. Evidently, $1' \otimes 1'' = 1$, $1' \otimes 1' = 1''$, and $1'' \otimes 1'' = 1'$, and also $(1')^* = 1''$.

Note that the two 3s on the right hand side of $3 \otimes 3 = 1 \oplus 1' \oplus 1'' \oplus 3 \oplus 3$ may be taken to be (u_2v_3, u_3v_1, u_1v_2) and (u_3v_2, u_1v_3, u_2v_1) . Recall from exercise I.3.1 that neither of these transforms as a vector under $SO(3)$. Under the rotations restricted to belong to T , however, they do transformation properly, like a 3.

Decomposition of the product using characters

We can also derive the decomposition of $3 \otimes 3$ by using the character table of A_4 given in (II.3.15). As explained in chapter II.1, the characters of the product representation $3 \otimes 3$ are given by the square of the characters of the irreducible representation 3, namely, $\begin{pmatrix} 3^2 \\ (-1)^2 \\ 0^2 \\ 0^2 \end{pmatrix} = \begin{pmatrix} 9 \\ 1 \\ 0 \\ 0 \end{pmatrix}$. We now use two results, derived in (II.2.8) and (II.2.9), which I reproduce here for your convenience, telling us about the number of times n_r the irreducible representation r appears in a given representation:

$$\sum_c n_c \chi^*(c) \chi(c) = N(G) \sum_r n_r^2 \quad (7)$$

and

$$\sum_c n_c \chi^{*(r)}(c) \chi(c) = N(G) n_r \quad (8)$$

Applied here, (7) gives $1 \cdot 9^2 + 3 \cdot 1^2 + 0 + 0 = 81 + 3 = 84 = 12 \cdot 7$ and thus $\sum_r n_r^2 = 7$. The two solutions are $1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 = 7$ and $1^2 + 1^2 + 1^2 + 2^2 = 7$. The former seems highly unlikely, and so we will root for the latter.

To determine n_r , we simply plug into (8). Orthogonality of the product representation with 1 gives $1 \cdot 1 \cdot 9 + 3 \cdot 1 \cdot 1 + 0 + 0 = 12 \cdot 1$, and thus $3 \otimes 3$ contains 1 once. Similarly, $1'$ and $1''$ each occur once. Finally, orthogonality of the product representation with 3 gives $1 \cdot 3 \cdot 9 + 3 \cdot (-1) \cdot 1 + 0 + 0 = 27 - 3 = 24 = 12 \cdot 2$. Thus, $3 \otimes 3$ contains 3 twice.

We obtain $3 \otimes 3 \rightarrow 1 \oplus 1' \oplus 1'' \oplus 3 \oplus 3$, in agreement with our earlier discussion. Indeed,

$$\begin{pmatrix} 9 \\ 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ \omega \\ \omega^* \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ \omega^* \\ \omega \end{pmatrix} + \begin{pmatrix} 3 \\ -1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 3 \\ -1 \\ 0 \\ 0 \end{pmatrix} \quad (9)$$

Note that this method is of more direct applicability. In most cases, we may not know a continuous group that contains the given finite group, or, if we know the continuous group, we may not be familiar with its Clebsch-Gordan decomposition rules.

5. Isospin and the Discovery of a Vast Internal Space

A small number in the subnuclear world

In 1932, James Chadwick discovered one of the most important small numbers in the history of physics, the mass difference between the proton and the neutron in units of the neutron mass:

$$(M_n - M_p)/M_n \simeq (939.6 - 938.3)/939.6 \simeq 0.00138 \quad (1)$$

Almost immediately,¹ in the same year, Werner Heisenberg (and others²) proposed that the strong interaction is invariant under a symmetry group, which transforms the proton and the neutron into linear combinations of each other. Since the proton is charged while the neutron is not, the electromagnetic interaction was thought to be responsible for the small mass difference. It was postulated that in a world with electromagnetism turned off, the proton p and neutron n would have equal mass,³ and they would furnish the spinor representation $N \equiv \begin{pmatrix} p \\ n \end{pmatrix}$ of a symmetry group $SU(2)$ that leaves the strong interaction invariant. Since the electromagnetic interaction is much weaker than the strong interaction, we expect that neglecting electromagnetic effects would give a reasonably approximate description of the real world.

I have often been struck by Nature's kindness toward physicists; it is almost as if we were offered a step-by-step instruction manual. Theoretical physicists did not even have to learn more group theory: they already knew about $SU(2)$ and half-integral spin. Hence the name isospin* for this symmetry. The invariance under the strong interaction is the first example of an approximate symmetry⁴ in physics, in contrast to rotation invariance, which, as far as we know, is an exact symmetry.

Prior to isospin, the symmetries of physics (translation invariance, rotation invariance, Lorentz invariance, and so on) were confined to the spacetime we live and love in. Heisenberg's enormous insight* led to the discovery of a vast internal space, the ongoing exploration of which has been a central theme of fundamental physics for close to a hundred years now.

The pions and the electric charge

The charged pions, π^+ and π^- , were discovered in 1947 after much experimental confusion.⁵ Yukawa, with remarkable insight, had postulated in 1935 that the exchange of a then-unknown meson[†] could generate the strong interaction, much like the exchange of the photon generates the electromagnetic interaction. Another example of Nature's kindness to theoretical physicists! Or, you might say that Nature knows only so many readily solvable problems, sort of like a professor making up an exam.

The power of the symmetry argument is that, since the π^+ and π^- interact with the nucleons, they are immediately required to also transform under isospin. Much as charged particles could emit and absorb photons, nucleons could emit and absorb pions via the following fundamental processes: $p \rightarrow n + \pi^+$, $n \rightarrow p + \pi^-$, and so on.

The initial states in these two processes shown have isospin $I = \frac{1}{2}$. Denote the unknown isospin of the pions by I_π . Then, according to the group theoretic result in chapter IV.3, the final states have isospin $\frac{1}{2} \otimes I_\pi = (I_\pi + \frac{1}{2}) \oplus |I_\pi - \frac{1}{2}|$. For isospin to be a symmetry of the strong interaction, this must contain the $I = \frac{1}{2}$ representation of the initial state; hence I_π can only be either 0 or 1. But $I_\pi = 0$ cannot accommodate both π^+ and π^- , and so we conclude that $I_\pi = 1$, a 3-dimensional irreducible representation.

Isospin thus predicts that π^+ and π^- must have an electrically neutral partner, the π^0 . The predicted neutral pion was soon discovered⁶ in 1950. Another triumph (small by today's standards perhaps, but nevertheless a triumph) for group theory.

How does the electric charge operator Q fit into the Lie algebra of $SU(2)$?

Isospin is supposed to be exact in an ideal world without electromagnetism. It is important to realize that, even in that world, it is meaningful and legitimate to ask how the electric-charge operator Q transforms under isospin. The following simple observation provides the answer.

Observe that the difference in charge of the proton and of the neutron is given by $\Delta Q = Q_p - Q_n = 1 - 0 = \frac{1}{2} - (-\frac{1}{2}) = I_{3,p} - I_{3,n} = \Delta I_3$, from which we deduce[‡]

$$Q = I_3 + \frac{1}{2}Y \quad (2)$$

where Y , known as hypercharge, is an operator lying outside $SU(2)$. Evidently, $Y = 1$ for nucleons and $Y = 0$ for pions.⁷ Note that the formula $Q = I_3$ for the pions indicates that the missing $I_3 = 0$ partner of the $\pi^\pm$ indeed has $Q = 0$.

Scattering cross sections and isospin

The predictions of isospin have been verified by an almost countless number of measurements. A particularly elegant, and simple, prediction⁸ concerns deuteron production in nucleon-nucleon collision, involving the two processes

$$(A) \quad p + p \rightarrow d + \pi^+$$

and

$$(B) \quad p + n \rightarrow d + \pi^0$$

The deuteron, as you may know, is a bound state of the proton and the neutron.

We can now put our vast knowledge of group theory to good use. The first step is to establish the isospin of the deuteron. Since $\frac{1}{2} \otimes \frac{1}{2} = 1 \oplus 0$, the deuteron, as a p - n bound state, has isospin equal to either 1 or 0. But if the deuteron's isospin is 1, then it would be a member of a triplet (since $2 \cdot 1 + 1 = 3$) with nearly equal mass. Applying the isospin raising and lowering operators, we see that the other two members are a p - p bound state and an n - n bound state. Since neither of these were seen,* we deduce that the deuteron has isospin 0.

Thus, the final state in both process (A) and process (B), containing a deuteron and a pion, has isospin $0 \otimes 1 = 1$.

On the other hand, as far as isospin is concerned, the initial state is, in process (A), given by $\left| I = \frac{1}{2}, I_3 = \frac{1}{2} \right\rangle \otimes \left| I = \frac{1}{2}, I_3 = \frac{1}{2} \right\rangle = \left| \frac{1}{2}, \frac{1}{2}; \frac{1}{2}, \frac{1}{2} \right\rangle = \left| \frac{1}{2}, \frac{1}{2} \right\rangle$ (to use the notation of chapter IV.3), and in process (B), it is given by $\left| \frac{1}{2}, -\frac{1}{2} \right\rangle$.

But we know about Clebsch-Gordan decomposition! We had obtained in chapter IV.3 that

$$\begin{aligned} |1, 1\rangle &= \left| \frac{1}{2}, \frac{1}{2} \right\rangle \\ |1, 0\rangle &= \frac{1}{\sqrt{2}} \left(\left| -\frac{1}{2}, \frac{1}{2} \right\rangle + \left| \frac{1}{2}, -\frac{1}{2} \right\rangle \right) \\ |0, 0\rangle &= \frac{1}{\sqrt{2}} \left(\left| -\frac{1}{2}, \frac{1}{2} \right\rangle - \left| \frac{1}{2}, -\frac{1}{2} \right\rangle \right) \end{aligned} \tag{3}$$

The initial proton-proton state in process (A), $|p, p\rangle = \left| \frac{1}{2}, \frac{1}{2} \right\rangle = |1, 1\rangle$, is an $I = 1$ state. In contrast, to obtain the initial proton-neutron state in process (B), $|p, n\rangle = \left| \frac{1}{2}, -\frac{1}{2} \right\rangle$, we have to invert (3) using high school algebra. We obtain $|p, n\rangle = \left| \frac{1}{2}, -\frac{1}{2} \right\rangle = \frac{1}{\sqrt{2}} (|1, 0\rangle - |0, 0\rangle)$, a mixture of an $I = 1$ state and an $I = 0$ state. But since the final state is $I = 1$, the $I = 0$ component of the initial state cannot contribute because of isospin conservation. Thus, compared to the amplitude for $p + p \rightarrow d + \pi^+$, the amplitude for $p + n \rightarrow d + \pi^0$ is down by a factor $\frac{1}{\sqrt{2}}$.

More formally, we can introduce a transition operator $\mathcal{T}$, which takes the initial state to the final state. Then $\langle d\pi^+ | \mathcal{T} | pp \rangle = \langle 1, 1 | \mathcal{T} | 1, 1 \rangle$, while

$$\langle d\pi^0 | \mathcal{T} | pn \rangle = \langle 1, 0 | \mathcal{T} \left(\frac{1}{\sqrt{2}} (|1, 0\rangle - |0, 0\rangle) \right) = \frac{1}{\sqrt{2}} \langle 1, 0 | \mathcal{T} | 1, 0 \rangle \quad (4)$$

Since in quantum mechanics, the cross section is determined by the absolute square of the amplitude, we obtain

$$\frac{\sigma(p + p \rightarrow d + \pi^+)}{\sigma(p + n \rightarrow d + \pi^0)} = 2 \quad (5)$$

The point is that although almost nothing was known about the strong interaction when the prediction was made, isospin allows us to fix the ratio of two cross sections, regardless of how complicated the actual details of the dynamics might be. Knowing that isospin is an approximate symmetry, we might expect this to hold to an accuracy of a few percent. When the experiments were done (in 1951 and 1953), sure enough, the ratio of the two measured cross sections comes out to be about 2 within the error bars.

I have marveled elsewhere⁹ about the almost magical power of theoretical physics. Picture the heroic efforts involved in building the necessary accelerator, all the wires, the vacuum pumps, and what not; and designing and setting up the detectors. In the end, the hardworking experimentalists measured the two cross sections, and bam! They confirmed that Nature knew about group theory.

It is worth emphasizing that had Heisenberg merely postulated a Z_2 symmetry under which $p \leftrightarrow n$, we would not get the prediction above. That the strong interaction is invariant under an $SU(2)$ is by now an established fact, and all known strongly interacting particles, known as hadrons, have been classified into $SU(2)$ irreducible representations. To give just one example, the $\pi^+ + p \rightarrow \pi^+ + p$ scattering cross section shows an enormous bump¹⁰ when the center of mass energy reaches about 1238 MeV. This resonance phenomenon is interpreted as the formation of $\pi^+ p$ into a short-lived particle known as N^{*++} with mass $\simeq 1238$ MeV, which then quickly decays back into $\pi^+ + p$.

Since the initial and final states contain a pion and nucleon, and since $1 \otimes \frac{1}{2} = \frac{3}{2} \oplus \frac{1}{2}$, the total isospin is either $I = \frac{3}{2}$ or $I = \frac{1}{2}$. But in the initial and final states, we have $I_3 = 1 + \frac{1}{2} = \frac{3}{2}$, and hence I can only be $I = \frac{3}{2}$, not $I = \frac{1}{2}$. We conclude that the particle N^{*++} has isospin $I = \frac{3}{2}$, and is thus a member of a quadruplet (since $2 \cdot \frac{3}{2} + 1 = 4$). Sure enough, three other particles, N^{*+} , N^{*0} , and N^{*-} (with the indicated charges and with masses approximately equal to that of N^{*++}), were eventually found.

A particularly clean prediction is that at resonance, the ratio of the total cross section for $\pi^+ + p$ and the total cross section for $\pi^- + p$ should be equal to 3. You should be able to work this out.

Nucleon nucleon scattering and Feynman diagrams

Historically, measuring and understanding the coupling strength of the three pions to the proton and to the neutron was of great importance in elucidating the strong interaction. Group theoretically, this simply corresponds to picking out the $\frac{1}{2}$ on the right hand side of the multiplication $1 \otimes \frac{1}{2} = \frac{1}{2} \oplus \frac{3}{2}$: we want “pion times nucleon = nucleon,” so to speak. Thus, the coupling strengths are proportional to a bunch of Clebsch-Gordan coefficients.

In fact, you already worked out the relevant coefficients in exercise IV.3.5. Group theory does not care whether you are combining the spin angular momentum $S = \frac{1}{2}$ of an electron with orbital angular momentum $L = 1$ or coupling the isospin $I = \frac{1}{2}$ of the nucleon with the isospin $I = 1$ of the pion. Simply plug in the result of the exercise.

$$\left| \frac{1}{2}, m \right\rangle = -\sqrt{\frac{\frac{3}{2} - m}{3}} \left| m - \frac{1}{2}, \frac{1}{2} \right\rangle + \sqrt{\frac{\frac{3}{2} + m}{3}} \left| m + \frac{1}{2}, -\frac{1}{2} \right\rangle \quad (6)$$

to obtain

$$\begin{aligned} \left| \frac{1}{2}, \frac{1}{2} \right\rangle &= -\sqrt{\frac{1}{3}} \left| 0, \frac{1}{2} \right\rangle + \sqrt{\frac{2}{3}} \left| 1, -\frac{1}{2} \right\rangle \\ \left| \frac{1}{2}, -\frac{1}{2} \right\rangle &= -\sqrt{\frac{2}{3}} \left| -1, \frac{1}{2} \right\rangle + \sqrt{\frac{1}{3}} \left| 0, -\frac{1}{2} \right\rangle \end{aligned} \quad (7)$$

Now change the names to protect the innocent and write (7) as

$$\begin{aligned} |p\rangle &\sim -\sqrt{\frac{1}{3}} |\pi^0, p\rangle + \sqrt{\frac{2}{3}} |\pi^+, n\rangle \\ |n\rangle &\sim -\sqrt{\frac{2}{3}} |\pi^-, p\rangle + \sqrt{\frac{1}{3}} |\pi^0, n\rangle \end{aligned} \quad (8)$$

Of course, this does not mean that the proton and neutron are simply equal to whatever appear on the right hand sides of (8); even Confusio would not think that. The strong interaction is far too strong and complicated for that. Rather, under isospin, the state $|p\rangle$ transforms like a linear combination of $|\pi^0, p\rangle$ and $|\pi^+, n\rangle$. Group theory fixes the coefficients, and tells us, for example, that the amplitude of $|p\rangle$ to be in the quantum state $|\pi^+, n\rangle$ is larger than the amplitude of $|p\rangle$ to be in the quantum state $|\pi^0, p\rangle$ by a factor $(-\sqrt{2})$.

This implies that the relative strength of the various couplings of a pion to a nucleon can be read off from (8):

$$g_{p,\pi^0 p} = g, \quad g_{p,\pi^+ n} = -\sqrt{2}g, \quad g_{n,\pi^- p} = \sqrt{2}g, \quad g_{n,\pi^0 n} = -g \quad (9)$$

Here g is a coupling strength characteristic of the strong interaction, which we are not able to calculate from first principles to this very day. We have chosen the sign of g for later convenience. The notation is such that $g_{p,\pi^+ n}$, for example, is proportional to the probability amplitude of the process $\pi^+ + n \rightarrow p$, in which a neutron n absorbs a π^+ to become a proton p . By a basic tenet of quantum field theory, this is the same as the amplitude for $p \rightarrow \pi^+ + n$, in which p becomes n by emitting a π^+ . (Furthermore, this

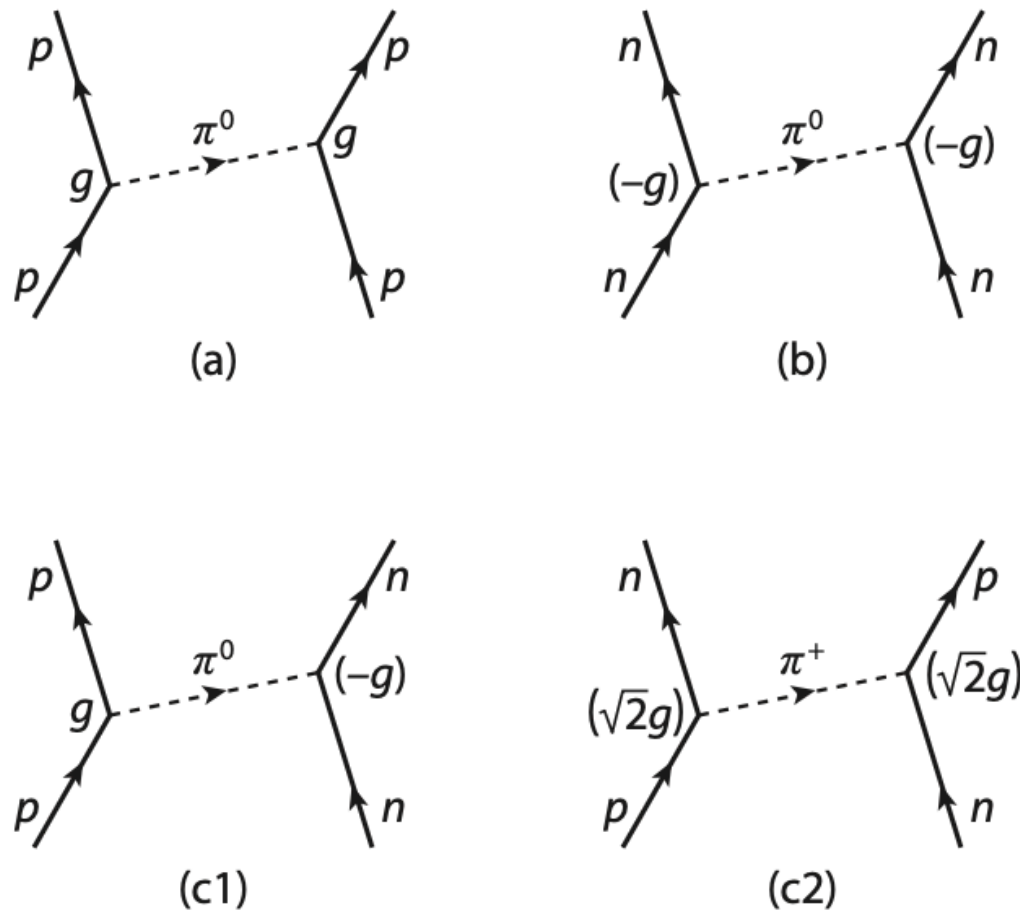


Figure 1

is the same as the amplitude for $n \rightarrow \pi^- + p$ and for $\pi^- + p \rightarrow n$. In other words, we could remove a particle from one side of a process and put it back on the other side as an antiparticle, for example, turning $p \rightarrow \pi^+ + n$ into $\pi^- + p \rightarrow n$. See any textbook on quantum field theory.)

The relative signs and factors of $\sqrt{2}$ sure look strange in (9), but in fact are entirely physical. There are three nucleon-nucleon scattering processes that experimentalists can measure, at least in principle, namely, $p + p \rightarrow p + p$, $n + n \rightarrow n + n$, and $p + n \rightarrow p + n$. According to Yukawa, the scattering is caused by the exchange of a pion between the two nucleons, as shown in the little diagrams (surely you've heard of Feynman diagrams!) in Figure 1. For our purposes here, you could think of the diagrams as showing what's happening in spacetime, with the time axis running upward.

For example, in the diagram in (c2), the proton becomes a neutron by emitting a π^+ , which is then absorbed by the neutron it is scattering with, turning it into a proton. In other words, we have $p + n \rightarrow n + \pi^+ + n \rightarrow n + p$. According to the discussion above, the emission and the absorption are both characterized by the coupling strength $g_{p,\pi^+n} = -\sqrt{2}g$. The various relevant coupling strengths are shown in these Feynman diagrams.

Now comes the fun. Let us write the relative strength of the quantum amplitude for each of these scattering processes. For $p + p \rightarrow p + p$, we have $(-g)^2 = g^2$. For $n + n \rightarrow n + n$, we have g^2 . So far so good: proton-proton scattering and neutron-neutron scattering are the same, as we would expect in an isospin-invariant world. Next, for $p + n \rightarrow p + n$, we have from the π^0 exchange in (c1) $g(-g) = -g^2$. Oops, this is not equal to g^2 . But let's not forget the contribution due to exchange of the charged pion shown in (c2): this gives $(\sqrt{2}g)(\sqrt{2}g) = 2g^2$. And so we have $-g^2 + 2g^2 = g^2$. Indeed, the three scattering processes come out to be the same. Once again, math works.

Pion nucleon coupling and the tensor approach

It is instructive also to use tensors and spinors to work out the relative strengths of the pion nucleon couplings.

Heisenberg started this story by proposing that the nucleons $N = \begin{pmatrix} p \\ n \end{pmatrix}$ transform like the 2-dimensional $I = \frac{1}{2}$ defining representation N^i of $SU(2)$, written here as a column spinor. It transforms as $N^i \rightarrow U^i_j N^j$. Recall that the hermitean conjugate of N^i is denoted by N_i^* , which can be written as a row spinor $\bar{N} = (\bar{p} \quad \bar{n})$. It transforms as $\bar{N}_i \rightarrow \bar{N}_j (U^\dagger)^j_i$, so that $N_i N^i$ is invariant. As explained in detail in chapter IV.4, upper indices transform with U , lower indices with $U^\dagger$.

We also learned in chapter IV.4 that the $I = 1$ vector representation of $SU(2)$, which happens to be also the adjoint representation, can be written as a traceless tensor ϕ^i_j , which can also be viewed as a 2-by-2 hermitean matrix. (This is an example of what I alluded to before; the simpler groups are often more confusing to beginners!) In other words, we can package the three pions π as a traceless 2-by-2 hermitean matrix:

$$\phi = \vec{\pi} \cdot \vec{\tau} = \pi_1 \tau_1 + \pi_2 \tau_2 + \pi_3 \tau_3 = \begin{pmatrix} \pi_3 & \pi_1 - i\pi_2 \\ \pi_1 + i\pi_2 & -\pi_3 \end{pmatrix} = \begin{pmatrix} \pi^0 & \sqrt{2}\pi^+ \\ \sqrt{2}\pi^- & -\pi^0 \end{pmatrix} \quad (10)$$

Note that we may pass freely between the Cartesian basis (π_1, π_2, π_3) and the circular basis (π^+, π^0, π^-) . The appearance of $\sqrt{2}$ here was explained for (IV.9.11). The pion fields transform according to

$$\phi^i_j \rightarrow \phi'^i_j = U^i_l \phi^l_n (U^\dagger)^n_j \quad (11)$$

As far as group theory is concerned, the representations that the pions and the nucleons belong to under isospin determine their interaction with one another; the interaction term in the Lagrangian can only be, precisely as we learned in chapter IV.4,

$$\mathcal{L} = \cdots + f N_i \phi_j^i N^j \quad (12)$$

with f some unknown coupling constant.[‡] The important point is that we do not have to know the details of the strong interaction: group theory rules!

Let us then write out the isospin invariant $N_i \phi_j^i N^j$, using (10), as

$$N_i \phi_j^i N^j = \begin{pmatrix} \bar{p} & \bar{n} \end{pmatrix} \vec{\pi} \cdot \vec{\tau} \begin{pmatrix} p \\ n \end{pmatrix} = (\bar{p} \pi^0 p - \bar{n} \pi^0 n) + \sqrt{2}(\bar{p} \pi^+ n + \bar{n} \pi^- p) \quad (13)$$

It is gratifying to see that the relative strength of $\sqrt{2}$ of the coupling of the charged pions to the neutral pion emerges here also, in agreement* with (9).

What we have done here is known as the effective field theory approach.¹¹ You, the astute reader, will realize that the essential point is that group theory does not care whether the objects being transformed are states or fields or whatever.

The Eightfold Way of SU(3)

While most of our colleagues were put off by the unfamiliar math, [Sidney Coleman and I] became traveling disciples of the Eightfold Way.

—S. L. Glashow

SU (3) in particle physics

Laugh out loud; the math that put off the leading particle theorists in the early 1960s was just the group SU(3), something that students these days are expected to breeze through lickety split.

Indeed, you the reader have already breezed through SU(N).

While the fame and glory of the group SU(2) permeate quantum physics (and parts of classical physics as well), the group SU(3) essentially appears only in particle physics.

But what fame and glory it enjoys there!

It appears twice, originally as the internal symmetry group that led to the discovery of quarks, and then, later, as the gauge group of quantum chromodynamics.

We will work out the representations of SU(3) by following two lines of attack: first, in this part, by using tensors, and second, in the next part, by generalizing the ladder approach used for SU(2).

We now need two floors

Moving from the house of SU(2) to the house of SU(3), we find ourselves now living on two floors.

For SU (2) tensors, recall that we can remove all lower indices using ϵ^{ij} .

In particular, ψ_i is equivalent to ψ^i .

This no longer holds for SU(3).

The reason is that, for SU(3), in contrast to SU(2), the antisymmetric symbols ϵ^{ijk} and ϵ_{ijk} carry three indices.

Thus, the irreducible representations furnished by ψ^i and ψ_i are not equivalent to each other: if we try the same trick that worked for us for SU(2) and contract ψ_i with ϵ^{ijk} , the resulting tensor $\psi^{jk} \equiv \epsilon^{ijk} \psi_i$ transforms like a tensor with two upper indices, not just one.

As explained earlier, the contracted index i has disappeared.

The irreducible representations furnished by ψ^i and ψ_i are commonly denoted by 3 and 3*.

They are not equivalent. (Physically, as we will see, the quark corresponds to 3, while the antiquark corresponds to 3*.)

Thus, for SU(3), in contrast to SU(2), we have to include tensors with both upper and lower indices.

However, in contrast to SU(N) for $N > 3$, SU(3) does have an important simplifying feature: it suffices to consider only traceless tensors $\varphi_{j_1 j_2 \dots j_n}^{i_1 i_2 \dots i_m}$ with all upper indices symmetrized and all lower indices symmetrized.

We prove this claim presently.

Thus, the representations of SU(3) are uniquely labeled by two integers (m, n), where m and n denote the number of upper and lower indices, respectively.

Recall that for SO(N), it suffices to consider only tensors with all upper indices symmetrized.

To show this, we used an iterative approach.

Essentially, the same approach works here.

The only catch is that with two integers (m, n) instead of one, we have to move on a 2-dimensional lattice, so to speak.

We iterate in $p \equiv m + n$; thus, starting at the lattice point $(m, n) = (0, 0)$, we move on to $(1, 0)$ and $(0, 1)$; then on to $(2, 0)$, $(1, 1)$, and $(0, 2)$; and so on and so forth, increasing p by 1 at each step.

To save verbiage, I assume from the outset that the reader, when presented with a tensor symmetric in its upper indices and symmetric in its lower indices, knows how to render it traceless by subtracting out its trace.

For example, given T^{ij}_{kl} , define

$$\tilde{T}^{ij}_{kl} \equiv T^{ij}_{kl} - \frac{1}{5}(\delta^i_k T^j_l + \delta^j_k T^i_l + \delta^i_l T^j_k + \delta^j_l T^i_k) + \frac{1}{20}(\delta^i_k T^j_l + \delta^j_k \delta^i_l + \delta^i_l \delta^j_k) T$$

where $T_{j_l} \equiv T^{ij}_{il}$ and $T \equiv T_{j_j}$.

We now prove our claim that, for SU(3), it suffices to consider only traceless tensors with all upper indices symmetrized and all lower indices symmetrized.

The claim holds trivially for the tensors characterized by $(0, 0)$, $(1, 0)$, and $(0, 1)$.

Now onward to $p = 2$.

First up is the tensor φ^{ij} with two upper indices.

As earlier, we break it up into the symmetric combination $\varphi^{\{ij\}} \equiv (\phi^{ij} + \phi^{ji})$ and the antisymmetric combination $\varphi^{[ij]} \equiv (\phi^{ij} - \phi^{ji})$ (using the notation $\{ij\}$ and $[ij]$ introduced earlier).

We can get rid of the antisymmetric combination by writing $\varphi_k \equiv \epsilon_{ijk} \varphi^{[ij]}$; in other words, the antisymmetric tensor with two upper indices transforms just like a tensor with one lower index, namely, $(0, 1)$, which we have dealt with already.

The remaining symmetric tensor $\varphi^{\{ij\}}$, now called $(2, 0)$, satisfies our claim.

Similarly, we can take care of $(0, 2)$ by using ϵ^{ijk} .

Note that, while we have to deal with both upper and lower indices, we also have two ϵ symbols, which we can use to trade two upper antisymmetric indices for one lower index, and two lower antisymmetric indices for one upper index, respectively.

Finally, we come to the tensor with one upper index and one lower index.

If it is not traceless, we simply subtract out its trace and obtain $(1, 1)$.

We are done with $p = 2$.

Moving from $p = 2$ to $p = 3$, we first encounter the tensor φ^{ijk} .

The analysis is the same as that we gave for $SO(3)$ earlier.

We break it up into the symmetric combination $\varphi^{\{ij\}k} \equiv (\varphi^{ijk} + \varphi^{jik})$ and the antisymmetric combination $\varphi^{[ij]k} \equiv (\varphi^{ijk} - \varphi^{jik})$.

Once again, we don't care about the antisymmetric combination $\varphi^{[ij]k}$, because it transforms just like a tensor $\varphi^k_l \equiv \epsilon_{ijl} \varphi^{[ij]k}$ with one upper index and one lower index, namely, $(1, 1)$, which we have already disposed of.

As for the symmetric combination $\varphi^{\{ij\}k}$, add and subtract its cyclic permutations and write, as earlier, $3\varphi^{\{ij\}k} = (\varphi^{\{ij\}k} + \varphi^{\{j k\}i} + \varphi^{\{ki\}j}) + (\varphi^{\{ij\}k} - \varphi^{\{j k\}i}) + (\varphi^{\{ij\}k} - \varphi^{\{ki\}j})$.

Note that the expression in the first set of parentheses is completely symmetric in all three indices, while the expressions in the other two sets of parentheses are antisymmetric in ki and kj , respectively.

We can thus multiply the latter two expressions by ϵ_{kil} and ϵ_{kjl} , respectively, and turn them into 2-indexed tensors, which we have already taken care of.

Speaking colloquially, 123 is 123, and we don't much care if it is the 123 of $SO(3)$ or of $SU(3)$.

You can carry the ball from here on.

For instance, you will next encounter ϕ^{ij}_k .

Its two upper indices can be symmetrized and antisymmetrized.

The antisymmetric piece contains (0, 2) and (1, 0).

So, the crucial observation, as noted above, is that we can always trade two antisymmetric indices for one index on a different floor, thus reducing $p = m + n$ by 1.

The dimensions of SU(3) tensors

Let us now compute the dimension of the SU(3) tensor $\varphi^{i_1 i_2 \cdots i_m}_{j_1 j_2 \cdots j_n}$ labeled by (m, n).

Once again, you should quickly review what we did for SO(3).

At the risk of repetition, note that our fingers don't care whether we are talking about SO(3) or SU(3): we have “merely” a counting problem here.

The number of distinct configurations of m indices symmetrized, according to a result earlier, is $\frac{1}{2}(m+1)(m+2)$.

So at this stage, our tensor has $\frac{1}{4}(m+1)(m+2)(n+1)(n+2)$ components.

You remember, of course, that we have yet to impose the traceless condition $\delta_{j_i} \varphi_{j_1 j_2 \dots j_n}^{i_1 i_2 \dots i_m} = 0$, and here comes the crucial difference between SO(3) and SU(3): an upper index and a lower index are contracted, rather than two upper indices.

The left hand side of this condition transforms like a tensor labeled by $(m-1, n-1)$; in other words, $1/4 m(m+1)n(n+1)$ linear combinations of the tensor $\varphi_{j_1 j_2 \dots j_n}^{i_1 i_2 \dots i_m}$ are to be set to 0.

We conclude that the dimension of the irreducible representation (m, n) of SU(3) is equal to

$$\begin{aligned} \mathcal{D}(m, n) &= \frac{1}{4}(m+1)(m+2)(n+1)(n+2) - \frac{1}{4}m(m+1)n(n+1) \\ &= \frac{1}{2}(m+1)(n+1)(m+n+2) \end{aligned} \tag{1}$$

Keep in mind that the tensor furnishing (m, n) is symmetric in its m upper indices and in its n lower indices.

Let us list the dimension of a few low-lying irreducible representations:

(1, 0)	$\frac{1}{2} \cdot 2 \cdot 1 \cdot 3 = 3$
(1, 1)	$\frac{1}{2} \cdot 2 \cdot 3 \cdot 4 = 8$
(2, 0)	$\frac{1}{2} \cdot 3 \cdot 1 \cdot 4 = 6$
(3, 0)	$\frac{1}{2} \cdot 4 \cdot 1 \cdot 5 = 10$
(2, 1)	$\frac{1}{2} \cdot 3 \cdot 2 \cdot 5 = 15$
(2, 2)	$\frac{1}{2} \cdot 3 \cdot 3 \cdot 6 = 27$

The numbers 3, 8, and 10 all have celebrity status in particle physics, and 27 is involved in an interesting Feynman story.

We have already mentioned 10 in connection with GellMann's stunning prediction of the hitherto unknown earlier.

Multiplication of SU(3) irreducible representations

You're almost an old hand by now at multiplying irreducible representations together using the tensor approach.

In fact, the most important cases for physics are the simplest.

Consider, for example, $3 \otimes 3^*$.

The product of ψ^i and χ_j , namely $\psi^i \chi_j$, transforms like the tensor T^i_j .

We render it traceless by subtracting out its trace, which transforms like the trivial representation.

We write this important result using two different notations:

$$(1, 0) \otimes (0, 1) = (1, 1) \oplus (0, 0) \tag{2}$$

and

$$3 \otimes 3^* = 8 \oplus 1 \tag{3}$$

To write the second form, we used (1).

We need hardly mention that nobody is missing ($3 \cdot 3 = 9 = 8 + 1$).

The next easiest example is

$$(1, 0) \otimes (1, 0) = (2, 0) \oplus (0, 1) \quad (4)$$

which, according to (1), corresponds to

$$3 \otimes 3 = 6 \oplus 3^* \quad (5)$$

(Now $3 \cdot 3 = 9 = 6 + 3$.)

How about $3 \otimes 6$, that is, $(1, 0) \otimes (2, 0)$?

Try to figure it out before reading on.

Write out the tensors ψ^i and φ^{jk} (with $\varphi^{jk} = \varphi^{kj}$) explicitly. The product $T^{ijk} = \psi^i \varphi^{jk}$ is a tensor with three indices upstairs, symmetric in jk but not totally symmetric. So symmetrize in ij ; extract the antisymmetric part by contracting with the antisymmetric symbol: $\varepsilon_{mij} \psi^i \varphi^{jk} \equiv \zeta_m^k$. Show that ζ_m^k is traceless, that is, $\zeta_k^k = 0$, and thus it corresponds to

$(1, 1) = 8$. We remove these eight components from T^{ijk} . What is left is a totally symmetric tensor S^{ijk} with three indices, namely, $(3, 0) = 10$.

We thus obtain

$$(1, 0) \otimes (2, 0) = (3, 0) \oplus (1, 1) \quad (6)$$

and

$$3 \otimes 6 = 10 \oplus 8 \quad (7)$$

Putting (5), (3), and (7) together, we have

$$3 \otimes 3 \otimes 3 = (6 \oplus 3^*) \otimes 3 = (6 \otimes 3) \oplus (3^* \otimes 3) = 10 \oplus 8 \oplus 8 \oplus 1 \quad (8)$$

$$(3^3 = 27 = 10 + 8 + 8 + 1.)$$

As we shall see, this is a highly celebrated result in particle physics.

Our next example, which played an important role historically, is more involved, but it will give us a hint on how to do the general case: $8 \otimes 8 = ?$

Let us go slow and be as explicit as possible. Multiply ψ_j^i and χ_l^k together to form $\tilde{T}_{jl}^{ik} = \psi_j^i \chi_l^k$. Take out the two traces $P_j^k = \psi_j^i \chi_i^k$ and $Q_l^i = \psi_j^i \chi_l^j$. Render these two tensors traceless by taking out the trace $P_j^j = Q_i^i$. Thus far, we have gotten $8 \oplus 8 \oplus 1$ on the right hand side of $8 \otimes 8 = ?$. (The reader should note that, as remarked in an earlier chapter, the letters $\tilde{T}$, ψ , χ , P , and Q are just coatracks on which to hang indices; we are “wasting” letters, so to speak, in the service of pedagogy, even though they come for free.) Next, subtract out all these traces from $\tilde{T}_{jl}^{ik}$ and call the resulting traceless tensor T_{jl}^{ik} .

Antisymmetrize in the upper indices i and k : $A_{jl}^{ik} = T_{jl}^{[ik]}$. You might think that we would have to symmetrize and antisymmetrize in the lower indices j and l next. But thanks to an identity involving the ε symbols, it turns out that, without further massaging, A_{jl}^{ik} is already symmetric in its two lower indices j and l .

To prove this, we first define $B_{mjl} \equiv \varepsilon_{ikm} A_{jl}^{ik}$. The claim is that B_{mjl} is symmetric in j and l . We verify the claim by showing that $B_{mjl} \varepsilon^{jln} = \varepsilon_{ikm} \varepsilon^{jln} A_{jl}^{ik} = 0$, since by construction A_{jl}^{ik} is traceless. (You will show this in exercise 2.) Furthermore, $B_{mjl} \varepsilon^{mjp} = 0$ for a similar reason. Thus, B_{mjl} is totally symmetric and furnishes the irreducible representation $(0, 3) = 10^*$.

Going through similar steps with $T_{[jl]}^{ik}$ gives us the irreducible representation $(3, 0) = 10$.

After taking out $T_{jl}^{[ik]}$ and $T_{[jl]}^{ik}$ from T_{jl}^{ik} , what is left is then symmetric in its upper indices and in its lower indices, and hence furnishes the irreducible representation $(2, 2)$, which from (1) we learn is the 27.

Thus, we obtain

$$(1, 1) \otimes (1, 1) = (2, 2) \oplus (3, 0) \oplus (0, 3) \oplus (1, 1) \oplus (1, 1) \oplus (0, 0) \quad (9)$$

and

$$8 \otimes 8 = 27 \oplus 10 \oplus 10^* \oplus 8 \oplus 8 \oplus 1 \quad (10)$$

Again, we accounted for everybody,⁵ of course: $8 \cdot 8 = 64 = 27 + 10 + 10 + 8 + 8 + 1$.

Multiplication rule for SU (3)

How do we multiply (m, n) and (m', n') ?

After all the examples I went through, I am joking only slightly if I say that the most difficult part of obtaining the general multiplication rule for $SU(3)$ is to invent a suitable notation for the beasts we encounter on our way to the answer. If the following gets hard to follow, simply refer back to the $(1, 1) \otimes (1, 1)$ example we just did.

The product of (m, n) and (m', n') is a tensor with $m + m'$ upper indices and $n + n'$ lower indices.

Let us denote by $(m, n; m', n')$ the traceless tensor with $m + m'$ upper indices, symmetric in the first m indices and symmetric in the second m' indices, and with $n + n'$ lower indices, symmetric in the first n indices and symmetric in the second n' indices. In other words, $(m, n; m', n')$ denotes what remains of the product of (m, n) and (m', n') after all traces have been taken out.

Note that $(m, n; m', n')$, since it is defined to be traceless, is certainly not $(m, n) \otimes (m', n')$, which has no reason to be traceless. Let us record the process of taking out traces:

$$\begin{aligned}
 (m, n) \otimes (m', n') &= (m, n; m', n') \\
 &\quad \oplus (m - 1, n; m', n' - 1) \oplus (m, n - 1; m' - 1, n') \\
 &\quad \oplus (m - 1, n - 1; m' - 1, n' - 1) \oplus (m - 2, n; m', n' - 2) \\
 &\quad \oplus (m - 2, n - 1; m' - 1, n' - 2) \\
 &\quad \oplus \cdots ||
 \end{aligned} \tag{11}$$

Note that (11) is merely saying that we first obtain $(m - 1, n; m', n' - 1)$ from $(m, n) \otimes (m', n')$ by contracting an upper index from the first m upper indices with a lower index from the second n' lower indices, and then removing all traces.

Since (m, n) is already traceless, there is no point in contracting an upper index from the first m upper indices with a lower index from the first n lower indices.

And since (m, n) and (m', n') are both respectively symmetric in their upper indices and in their lower indices, it does not matter which upper index and which lower index we choose to contract.

By the way, all this is harder to say than to understand.

Next, we obtain $(m, n-1; m'-1, n')$ by contracting an upper index from the second m' upper indices with a lower index from the first n' lower indices, and then removing all traces.

We keep repeating this process.

The notation $\dots$ in (11) indicates that the process stops when we run out of indices to contract.

After doing all this, we end up with a bunch of reducible representations furnished by objects $(m-p, n-q; m'-q, n'-p)$.

Note the word “reducible.”

As in our $8 \otimes 8$ example, we next take an index from the $(m-p)$ upper indices and an index from the $(m'-q)$ upper indices, and antisymmetrize.

The claim, as in the explicit example, is that the result is already symmetric in the lower indices.

Similarly, we can exchange the words “upper” and “lower” in the preceding paragraph.⁶ It is important to emphasize that $(m, n; m', n')$ (and $(m-p, n-q; m'-q, n'-p)$ in particular) is neither symmetric in all its upper indices nor symmetric in all its lower indices, and hence is certainly not $(m + m', n + n')$.

Thus, $(m, n; m', n')$ furnishes a reducible representation, not an irreducible one.

Let us illustrate this procedure with our $8 \otimes 8$ example:

$$(1, 1) \otimes (1, 1) = (1, 1; 1, 1) \oplus (0, 1; 1, 0) \oplus (1, 0; 0, 1) \oplus (0, 0; 0, 0) \quad (12)$$

followed by

$$\begin{aligned} (1, 1; 1, 1) &= (2, 2) \oplus (3, 0) \oplus (0, 3) \\ (0, 1; 1, 0) &= (1, 1) \\ (1, 0; 0, 1) &= (1, 1) \\ (0, 0; 0, 0) &= (0, 0) \end{aligned} \quad (13)$$

Putting (12) and (13) together, we recover (9).

From the Λ to the Eightfold Way

At this point, for pedagogical clarity, it would be best to interject some particle physics, which I recount in as abridged and cartoonish a form as possible.

The Λ baryon was discovered in 1950 by a team of Australians.

Its mass, ≈ 1115 MeV, is not far from the mass of the proton and the neutron, ≈ 939 MeV, and its other properties (for instance, it also has spin $1/2$) are also similar to that of the proton p and the neutron n .

This naturally led S. Sakata in 1956 to generalize Heisenberg's $SU(2)$ to $SU(3)$, proposing that p , n , and Λ furnish the fundamental or defining representation of $SU(3)$.

But Nature was unkind to Sakata. Soon, other baryons (collectively known as hyperons, but never mind) were discovered: Σ^+ , Σ^0 , and Σ^- , with masses around 1190 MeV, and Ξ^- and Ξ^0 , with masses around 1320 MeV. They form an isospin triplet and an isospin doublet, respectively. Also, four K mesons, denoted by K^+ , K^0 and $\bar{K}^0$, K^- , with properties similar to the three pions, were discovered.¹⁰ One troubling feature was that their masses, around 495 MeV, were quite different from the pion masses, around 138 MeV.

That there were seven pseudoscalar spin 0 mesons (pseudoscalar because they were odd under parity, that is, under spatial reflection) plus the fact that some eminent theorists presented a series of arguments that the Λ and Σ^0 had opposite parities and hence could not belong together, sent a number of theorists on a search¹¹ for a symmetry group with a 7-dimensional irreducible representation.

But eventually another pseudoscalar spin 0 meson, the η with mass around 550 MeV, was discovered. Furthermore, better measurements established that the Λ and Σ^0 do in fact have the same parity, namely, the parity of the proton and the neutron.

Gell-Mann,¹² and independently Ne'eman, then proposed that the eight spin 0 mesons¹³ and the eight spin $\frac{1}{2}$ baryons furnish the 8-dimensional adjoint representation of $SU(3)$, namely $(1, 1)$ in the tensor notation.

As explained earlier, Gell-Mann then made the striking prediction that the nine known baryon resonances (recall the N^{*++} from that chapter) belong to a 10-dimensional irreducible representation (namely, $(3, 0)$) of $SU(3)$. The missing particle, named the Ω^- , was soon found. Taking his inspiration from Buddhism,¹⁴ Gell-Mann referred to his scheme as the Eightfold Way in honor of the 8-dimensional adjoint representation.

If $SU(2)$ was regarded as an approximate symmetry, then $SU(3)$ is a badly broken symmetry. We would expect the predictions to hold to an accuracy of at most $\sim 20\text{--}30\%$, as measured by, say, the fractional mass difference¹⁵ between the Ξ s and the nucleons. We will discuss later what group theory has to say about baryon and meson masses.

Quarks and triality

You might have noticed that the irreducible representations (m, n) actually observed experimentally, namely, the octet $8 = (1, 1)$ and the decuplet $10 = (3, 0)$, have $m - n = 0 \bmod 3$. It was soon understood that the irreducible representations of $SU(3)$ may be classified according to their triality, which we define to be equal to $(m - n) \bmod 3$.

Recall that, earlier, we defined the center of a group as the set of all elements that commute with all the other elements. Also, we have already discussed the center of $SU(N)$. Still, it is worthwhile to discuss here the physical meaning of the center of $SU(3)$.

What is the center of $SU(3)$? Consider the set of all unitary 3-by-3 matrices with unit determinant. To commute with all these matrices, a matrix has to be proportional to the identity matrix and to have unit determinant. Thus, the proportionality constant has to be a multiple of $e^{2\pi i/3}$. The center of $SU(3)$ is the group Z_3 , consisting of the three elements I , z , and z^2 , where

$$z \equiv \begin{pmatrix} e^{2\pi i/3} & 0 & 0 \\ 0 & e^{2\pi i/3} & 0 \\ 0 & 0 & e^{2\pi i/3} \end{pmatrix} = e^{2\pi i/3} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{14}$$

(I wrote this out in two different forms for emphasis.) Acting on the fundamental representation 3, this “triality matrix” z multiplies the representation by the phase $e^{2\pi i/3}$. Acting on the 3^* , z multiplies it by the phase $e^{-2\pi i/3}$. More generally, acting on the tensor (m, n) , z multiplies it by $e^{2\pi i/3}$ for each upper index and by $e^{-2\pi i/3}$ for each lower index, that is, by $e^{2\pi i(m-n)/3}$. This explains the origin of triality.

Historically, the mysterious fact that the experimentally observed irreducible representations are precisely those with triality 0 provided another important clue to the underlying theory of the strong interaction.

Indeed, at this point, you might well ask,* “Where is the fundamental representation 3?” Gell-Mann soon worked out the properties of the particles furnishing the 3 and named them quarks,[†] specifically the up quark u , the down quark d , and the strange quark s . The conjugate representation 3^* is then furnished by the antiquarks, $\bar{u}$, $\bar{d}$, and $\bar{s}$.

Group theory tells us, very nicely, how the observed strongly interacting particles are built up from quarks and antiquarks. More specifically, each upper index in the tensor furnishing a given irreducible representation corresponds to a quark and each lower index to an antiquark. For example, $3 \otimes 3^*$ contains the 8 according to (3), and the observed octet of pseudoscalar mesons are now known to be bound states of a quark and an antiquark: $\bar{u}u, \bar{d}u, \bar{u}s$, and so on. And $3 \otimes 3 \otimes 3$ also contains the 8 according to (8), and so the protons and neutrons and their cousins are bound states of 3 quarks, uud, udd, uus , and so on. Interestingly, since $3 \otimes 3 \otimes 3$ also contains the 10 according to (8), the celebrated Ω^- is also a 3-quark bound state, namely, sss .

You see that triality is just the number of quarks minus the number of antiquarks mod 3.

Experimentalists have searched for quarks in vain. It is now believed that quarks, and more generally, particles with nonzero triality, are confined; that is, they cannot be produced experimentally. A major challenge for the theory of the strong interaction is to prove confinement.¹⁶

You shall know the whole by its parts: The analog of crystal field splitting

You learned earlier that an irreducible representation r of a group G will, in general (when we restrict ourselves to a subgroup H of G), break up into a bunch of irreducible representations of H .

As explained there, this makes total sense, because H has fewer transformations than G does.

Fewer transformations are available to transform the members of the irreducible representation r into linear combinations of one another, and so these guys can segregate themselves into different cliques.

In everyday life, you might discover how something is put together by taking it apart.

Historically, one way that particle physicists became acquainted with $SU(3)$ was to ask how the irreducible representations of $SU(3)$ break up upon restriction of $SU(3)$ to $SU(2)$.

By the late 1950s, particle physicists were well acquainted with Heisenberg's isospin and hence with the group $SU(2)$.

Mathematically, the problem of breaking up the irreducible representations of $SU(3)$ is almost trivial, at least in hindsight, using the method of tensors: we simply divide the index set $i = \{1, 2, 3\}$ into two sets $\{1, 2\}$ and $\{3\}$. Introduce indices $a, b, \dots$ that are allowed to range over $\{1, 2\}$.

Given a tensor, replace the index i by either a or 3. This is more easily explained with the aid of a few examples than with a bunch of words. Consider the fundamental representation 3 of $SU(3)$. We separate $\psi^i, i = 1, 2, 3$ into $\psi^a, a = 1, 2$, and ψ^3 . The $SU(2)$ subgroup of $SU(3)$ acts on the index a but leaves the index 3 alone. This is a long-winded way of stating the obvious: upon restriction¹⁷ of $SU(3)$ to $SU(2)$, the irreducible representation 3 of $SU(3)$ decomposes as

$$3 \rightarrow 2 \oplus 1 \tag{15}$$

Physically, $SU(3)$ transforms the three quarks, u , d , and s , into linear combinations of one another, while its subgroup $SU(2)$, namely, Heisenberg's isospin, transforms u and d but leaves s alone. Under isospin, the up and down quarks form a doublet, while the strange quark is a singlet.

We can convey more information by going to $SU(2) \otimes U(1)$, the maximal subgroup of $SU(3)$. Here $U(1)$ denotes the group consisting of the $SU(3)$ elements $e^{i\theta Y}$ with the hypercharge matrix

$$Y = \frac{1}{3} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix} \quad (16)$$

Note that Y is hermitean and traceless, so that $e^{i\theta Y}$ is unitary and has unit determinant and hence belongs to $SU(3)$. In other words, the up and down quarks have hypercharge $\frac{1}{3}$, while the strange quark has hypercharge $-\frac{2}{3}$. Thus, upon restriction of $SU(3)$ to $SU(2) \otimes U(1)$, we have, instead of (15), the more informative $3 \rightarrow (2, 1) \oplus (1, -2)$. Here, in $(2, 1)$, for example, the first number indicates the dimension of the isospin representation, while the second number is $3Y$. (I put in the 3 to avoid writing $\frac{1}{3}$ constantly.) The compact notation I_{3Y} , with the hypercharge (multiplied by 3) written as a subscript to the isospin, is often more convenient. For example, we write the decomposition just given as

$$3 \rightarrow 2_1 \oplus 1_{-2} \quad (17)$$

The decomposition of the fundamental or defining representation specifies how the subgroup H is embedded in G .

Since all representations may be built up as products of the fundamental representation, once we know how the fundamental representation decomposes, we know how all representations decompose.

Conjugating (17), we have $3^* \rightarrow 2_{-1} + 1_2$. Notice that it is not necessary to write 2^* (and certainly not necessary to write 1^*). Given these two basic facts, we can then work out how a higher-dimensional irreducible representation of $SU(3)$ decomposes by multiplying representations together.

For example, the multiplication law $3 \otimes 3^* = 8 \oplus 1$ in $SU(3)$ becomes in $SU(2)$ (using (15))

$$\begin{aligned}
 3 \otimes 3^* &= 8 \oplus 1 \\
 &\rightarrow (2 \oplus 1) \otimes (2 \oplus 1) \\
 &= (2 \otimes 2) \oplus (2 \otimes 1) \oplus (1 \otimes 2) \oplus (1 \otimes 1) \\
 &= 3 \oplus 1 \oplus 2 \oplus 2 \oplus 1
 \end{aligned} \tag{18}$$

Thus, we learn that

$$8 \rightarrow 3 \oplus 1 \oplus 2 \oplus 2 \tag{19}$$

Group theory tells us that the octet of mesons should consist of an isospin triplet (the π^+ , π^0 , π^-), two isospin doublets (K^+ , K^0 and $\bar{K}^0$, K^-), and an isospin singlet (the η), exactly as was observed experimentally. Similarly, the octet of baryons consists of an isospin triplet (the Σ^+ , Σ^0 , Σ^-), two isospin doublets (our beloved proton and neutron, and Ξ^- , Ξ^0), and an isospin singlet (the Λ that we started our discussion with). Note that the decomposition (19) is a property of the irreducible representation 8, regardless of whether we thought of it as being contained in $3 \otimes 3^*$ or in $3 \otimes 3 \otimes 3$.

We can readily include hypercharge. Using (17) and its conjugate $3^* \rightarrow 2_{-1} \oplus 1_2$, and noting that hypercharges simply add (since the group is simply $U(1)$), we obtain a more informative version of (19):

$$8 \rightarrow 3_0 \oplus 1_0 \oplus 2_3 \oplus 2_{-3} \quad (20)$$

(This follows because the first two terms, for example, come from $2_1 \otimes 2_{-1}$.) Recall that the subscript is actually $3Y$. Thus, in the octet 8, the isospin triplet has $Y = 0$; the two isospin doublets $Y = \pm 1$, respectively; and the isosinglet $Y = 0$.

Alternatively, we can simply inspect the tensors involved. For example, the adjoint 8 decomposes like (on splitting the indices i and j into a, b , and 3, as explained earlier)

$$\varphi_j^i = \{\bar{\varphi}_b^a, \varphi_3^a, \varphi_a^3, \varphi_3^3\} \quad (21)$$

where the bar on $\bar{\varphi}_b^a$ reminds us that it is traceless. This corresponds precisely to (19).

In chapter IX.2, we will decompose various representations of $SU(5)$. Everything we do there will simply be somewhat more elaborate versions of what we have done here.

The electric charge

Theorists may think of isospin and hypercharge, but experimentalists identify particles by their electric charges. Back in chapter V.1, we saw that the proton has $I_3 = \frac{1}{2}$ but electric charge $Q = 1$, while the π^+ has both I_3 and Q equal to 1. In the early 1950s, physicists proposed the empirical formula $Q = I_3 + \frac{Y}{2}$, but they were deeply puzzled by this “dislocation” between Q and I_3 , and there was certainly no understanding of this mysterious hypercharge.

But within $SU(3)$ and the Eightfold Way, this formula suddenly makes sense: the photon couples to a particular linear combination of two $SU(3)$ generators, but group theory tells us how I_3 and $\frac{Y}{2}$ are correlated. For instance, (20) says that the isodoublet (consisting of the proton and neutron) has $Y = 1$ and hence charge $Q = (\frac{1}{2} + \frac{1}{2}, -\frac{1}{2} + \frac{1}{2}) = (1, 0)$, while the isotriplet (the pions) has $Y = 0$ and hence charge $Q = (1, 0, -1)$.

But this understanding also led to widespread skepticism of the notion of quarks. According to (17), that is, $3 \rightarrow 2_1 \oplus 1_{-2}$, the up quark must have electric charge

$$Q = \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{3} = \frac{2}{3} \quad (22)$$

This was quite a shock to physicists used to seeing integral electric charges! The down quark has charge $-\frac{1}{3}$. (One way of seeing this instantly is to note $Q_{\text{up}} - Q_{\text{down}} = 1$.) The strange quark has charge $Q = 0 + \frac{1}{2} \cdot \frac{1}{3} \cdot (-2) = -\frac{1}{3}$. The early searches for quarks often involved looking for these peculiar fractional charges.

The Lie Algebra of $SU(3)$ and Its Root Vectors

The way the Lie algebra of $SU(3)$ works out is not only important for its own sake, but it also indicates to us how to analyze a general Lie algebra.

The eight Gell-Mann matrices

To start with, we can readily write down the eight traceless hermitean matrices that represent the algebra of $SU(3)$, as already explained earlier.

They generalize the Pauli matrices and are known as Gell-Mann matrices in particle physics:

$$\begin{aligned}
\lambda_1 &= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, & \lambda_2 &= \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\
\lambda_4 &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, & \lambda_5 &= \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix} \\
\lambda_6 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, & \lambda_7 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} \\
\lambda_3 &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, & \lambda_8 &= \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}
\end{aligned} \tag{1}$$

These matrices are normalized by $\text{tr } \lambda_a \lambda_b = 2\delta_{ab}$. It would not have escaped your notice that the three Gell-Mann matrices λ_1 , λ_2 , and λ_3 correspond to the three Pauli matrices σ_1 , σ_2 , and σ_3 , respectively, with a third row and column of 0s added. This evidently specifies how the $SU(2)$ algebra is embedded in the $SU(3)$ algebra, as noted earlier.

You would also have noticed that I have separated the eight matrices into four groups as follows:

- (a) λ_3 and λ_8 are the two diagonal guys;
- (b) λ_1 and λ_2 , together with λ_3 , form an $SU(2)$ subalgebra;
- (c) λ_4 and λ_5 are basically the same as λ_1 and λ_2 , except that they live in the “1-3 sector,” while λ_1 and λ_2 live in the “1-2 sector”;
- (d) λ_6 and λ_7 are also basically the same as λ_1 and λ_2 , except that they live in the “2-3 sector”.

Next, note that

$$[\lambda_4, \lambda_5] = \left[\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix} \right] = 2i \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} \equiv 2i\lambda_{[4,5]} = i(\lambda_3 + \sqrt{3}\lambda_8) \quad (2)$$

You understand that the matrix I named $\lambda_{[4,5]}$ is the analog of λ_3 living in the 1-3 sector, and together with λ_4 and λ_5 , it generates an $SU(2)$ subalgebra.

Similarly, λ_6 and λ_7 , together with a matrix we might call $\lambda_{[6,7]}$ (which you should guess and then work out), the analog of λ_3 living in the 2-3 sector, generate an $SU(2)$ subalgebra.

Thus, not surprisingly, $SU(3)$ contains three overlapping $SU(2)$ subalgebras. They overlap because there are only two traceless hermitean matrices in the $SU(3)$ algebra, namely, λ_3 and λ_8 . (Of course, the three $SU(2)$ subalgebras have to overlap, since otherwise $SU(3)$ would be $SU(2) \otimes SU(2) \otimes SU(2)$, which it sure isn't.¹⁾ Three $SU(2)$ s have to share two J_3 s, so to speak.

In the defining or fundamental representation, the generators are represented by²
 $T^a = \frac{1}{2}\lambda_a$.

As explained in chapter I.3, the generators of any Lie algebra satisfy

$$[T^a, T^b] = if^{abc}T^c \quad (3)$$

with some structure constants f^{abc} characteristic of the algebra. Since (3) holds for any representation of the algebra, we can work out the f^{abc} s for $SU(3)$ by explicitly commuting the Gell-Mann matrices with one another. The result is:

$$\begin{aligned} f^{123} &= 1 \\ f^{147} &= -f^{156} = f^{246} = f^{257} = f^{345} = -f^{367} = \frac{1}{2} \\ f^{458} &= f^{678} = \frac{\sqrt{3}}{2} \end{aligned} \quad (4)$$

I have already worked out f^{453} and f^{458} in (2) for you. You should check the rest, or at least a few of them. Recall that f^{abc} is totally antisymmetric, a fact that we will prove in later. Thus, $f^{453} = f^{345}$, which you should also check explicitly.

It is important to realize that while we determined f^{abc} using the matrices in the fundamental representation, Lie's relation (3) holds for the generators in any representation.

There is, of course, absolutely no point in memorizing these structure constants; you will see that, even when we actually calculate using $SU(3)$, we will not need the explicit values given here.

It is, however, essential that you understand what the structure constants embody about the algebra.

From ladders to jungle gyms

For what to do next, we seek inspiration from the ladder approach used for $SU(2)$ and try to generalize it to $SU(3)$. Recall that, of the three generators J_x , J_y , and J_z of $SU(2)$, we diagonalized J_z and defined $J_{\pm} \equiv J_x \pm i J_y$. This enabled us to write the $SU(2)$ algebra as

$$[J_z, J_{\pm}] = \pm J_{\pm}, \quad [J_+, J_-] = 2J_z \quad (5)$$

Using Dirac's bra and ket notation and thinking of the matrix J_z as an operator, we wrote the eigenvector of J_z with eigenvalue m as $|m\rangle$: $J_z |m\rangle = m |m\rangle$. Manipulating (5), we deduced that $J_+ |m\rangle = c_{m+1} |m+1\rangle$ and $J_- |m\rangle = c_m^* |m-1\rangle$, with c_m a normalization constant that we determined by a clever use of the algebra (following Dirac!). We thus visualized the states $\dots, |m-1\rangle, |m\rangle, |m+1\rangle, \dots$ as rungs on a ladder, and thought of J_+ and J_- as a raising operator and a lowering operator, respectively, that enable us to climb up and down the ladder. The next step was to show that the ladder terminates, and that the maximum and minimum values of m are given by $\pm j$, with j an integer for $SO(3)$, and an integer or half-integer for $SU(2)$.

It is crucial for you to be thoroughly familiar with what we did for $SU(2)$, as we will be simply generalizing the reasoning involved in that case to $SU(3)$.

From the explicit forms of λ_3 and λ_8 in (1), we see that the generators T^3 and T^8 commute: $[T^3, T^8] = 0$, and thus we can simultaneously diagonalize them. At this point, there is a trivial complication: history. We now rename T^3 as the third component of isospin I_3 , and the T^8 as hypercharge Y up to the normalization factor $\frac{\sqrt{3}}{2}$. In a given irreducible representation, the states are characterized by the eigenvalue of I_3 and the eigenvalue of Y . In other words, we use the eigenstates of I_3 and Y to represent the algebra $SU(3)$, namely, the kets $|i_3, y\rangle$ such that $I_3 |i_3, y\rangle = i_3 |i_3, y\rangle$ and $Y |i_3, y\rangle = y |i_3, y\rangle$. The states $|i_3, y\rangle$ generalize the states $|m\rangle$ for $SU(2)$.

The states in an irreducible representation of $SU(3)$ are then labeled by two numbers and hence may be arranged in a 2-dimensional lattice rather than on a ladder.

The number of generators that can be simultaneously diagonalized is known as the rank of the Lie algebra. The rank of $SU(2)$ is 1: we can only diagonalize J_3 . The rank of $SU(3)$, in contrast, is 2. The states in an irreducible representation of a Lie algebra of rank[†] l are labeled by l numbers and hence may be arranged in an l -dimensional lattice. We will refer to the 2-dimensional lattice for $SU(3)$ as a jungle gym.

What about the other $6 = 8 - 2$ generators besides I_3 and Y ? We use $SU(2)$ as a guide and follow our noses, so to speak. Define

$$I_{\pm} = T_1 \pm iT_2$$

$$U_{\pm} = T_6 \pm iT_7$$

$$V_{\pm} = T_4 \pm iT_5$$

$$I_3 = T_3$$

$$Y = \frac{2}{\sqrt{3}}T_8$$

Generalizing isospin, particle physicists refer to the $SU(2)$ subalgebra generated by $I_{\pm}$, $U_{\pm}$, and $V_{\pm}$ as I spin, U spin, and V spin, respectively.*

Then, generalizing (5), we rewrite the commutation relations (3) as

$$[I_3, I_{\pm}] = \pm I_{\pm}, \quad [I_3, U_{\pm}] = \mp \frac{1}{2} U_{\pm}, \quad [I_3, V_{\pm}] = \pm \frac{1}{2} V_{\pm} \quad (7)$$

$$[Y, I_{\pm}] = 0, \quad [Y, U_{\pm}] = \pm U_{\pm}, \quad [Y, V_{\pm}] = \pm V_{\pm} \quad (8)$$

and

$$[I_+, I_-] = 2I_3 \quad (9)$$

$$[U_+, U_-] = \frac{3}{2}Y - I_3 = \sqrt{3}T_8 - T_3 \equiv 2U_3 \quad (10)$$

$$[V_+, V_-] = \frac{3}{2}Y + I_3 = \sqrt{3}T_8 + T_3 \equiv 2V_3 \quad (11)$$

$$[I_+, V_-] = -U_- \quad (12)$$

$$[I_+, U_+] = V_+, \quad (13)$$

$$[U_+, V_-] = T_-, \quad (14)$$

$$[I_+, V_+] = 0, \quad (15)$$

$$[I_+, U_-] = 0, \quad (16)$$

$$[U_+, V_+] = 0 \quad (17)$$

Other commutation relations can be obtained from the ones listed here by hermitean conjugation, for example, $([I_+, V_+])^{\dagger}$ gives $[I_-, V_-]$.

Clearly, with eight generators instead of three, we have many more commutation relations in $SU(3)$ than in $SU(2)$, but we will soon make sense of them. In the meantime, you might want to focus on the three commutation relations $[I_+, I_-] = 2I_3$, $[U_+, U_-] = 2U_3$, and $[V_+, V_-] = 2V_3$, with U_3 and V_3 defined above. These commutation relations, as remarked earlier, show that $SU(3)$ may be pictured as three overlapping $SU(2)$ s, with $I_\pm$, $U_\pm$, and $V_\pm$ the raising and lowering operators for the three $SU(2)$ algebras, respectively.

Climbing around on a jungle gym

The ladder for $SU(2)$ generalizes to a kind of 2-dimensional jungle gym for $SU(3)$ (and in general, an l -dimensional jungle gym). Instead of each state or ket corresponding to a rung on a ladder, each ket $|i_3, y\rangle$ now corresponds to a node or point on a jungle gym or lattice, laid down on a plane with one axis labeled by i_3 and the other by y . The six generators, $I_\pm$, $U_\pm$, and $V_\pm$, play the role of the raising and lowering operators $J_\pm$. Instead of taking us from one rung of the ladder to another, they now take us from one node on the jungle gym to another.

Since the commutation relations between I_3 , $I_\pm$ are exactly the same as those between J_3 , $J_\pm$, we have $I_\pm |i_3, y\rangle \propto |i_3 \pm 1, y\rangle$, exactly as in (IV.2.5) and (IV.2.6): $I_\pm$ increases and decreases i_3 by one unit, leaving y untouched. (This time we won't bother to write down and determine the proportionality constants, the analogs of the c_m s. I leave it to you to work it out.)

Next, what does $U_\pm$ do to $|i_3, y\rangle$? Find out by acting with I_3 and Y on $U_\pm |i_3, y\rangle$ and using (7) and (8). We obtain

$$\begin{aligned}
I_3 U_{\pm} |i_3, y\rangle &= \left(U_{\pm} I_3 \mp \frac{1}{2} U_{\pm} \right) |i_3, y\rangle = U_{\pm} \left(I_3 \mp \frac{1}{2} \right) |i_3, y\rangle \\
&= \left(i_3 \mp \frac{1}{2} \right) U_{\pm} |i_3, y\rangle
\end{aligned} \tag{18}$$

and

$$\begin{aligned}
Y U_{\pm} |i_3, y\rangle &= (U_{\pm} Y \pm U_{\pm}) |i_3, y\rangle = U_{\pm} (Y \pm 1) |i_3, y\rangle \\
&= (y \pm 1) U_{\pm} |i_3, y\rangle
\end{aligned} \tag{19}$$

We conclude that $U_{\pm} |i_3, y\rangle \propto |i_3 \mp \frac{1}{2}, y \pm 1\rangle$. (In this discussion, the $\pm$ signs are clearly correlated.)

Going through the analogous manipulations for $V_{\pm}$, you would conclude that $V_{\pm} |i_3, y\rangle \propto |i_3 \pm \frac{1}{2}, y \pm 1\rangle$.

Root vectors

Thus, the operators $I_{\pm}$, $U_{\pm}$, and $V_{\pm}$ move us around this lattice. Each of them can be characterized by a vector specifying the change in location $(\Delta i_3, \Delta y)$. From the calculation we just did, we see that

$I_{\pm}$ moves us by $(\pm 1, 0)$

$U_{\pm}$ moves us by $(\mp \frac{1}{2}, \pm 1)$

$V_{\pm}$ moves us by $(\pm \frac{1}{2}, \pm 1)$ (20)

in the (i_3, y) plane. The vectors $(\pm 1, 0)$, $(\mp \frac{1}{2}, \pm 1)$, and $(\pm \frac{1}{2}, \pm 1)$ associated with $I_{\pm}$, $U_{\pm}$, and $V_{\pm}$, respectively, are known as root vectors of the Lie algebra, or roots for short.

What can we say about the lengths of the root vectors and the angles between them?

But wait, the question has no meaning unless the two axes are normalized in the same way.* Earlier, we made sure that the Gell-Mann matrices λ_a were normalized in the same way; referring to (1), we see that, in the fundamental representation, while the third component of isospin is given by $I_3 = \frac{1}{2}\lambda_3$, the hypercharge, given by $Y = \frac{1}{\sqrt{3}}\lambda_8$, is normalized differently for “historical” reasons (namely the Gell-Mann-Nishijima formula mentioned in chapter V.1).

Physicists typically use I_3 and Y , bowing to history, but the great art of mathematics could care less about how physicists fumbled around in the 1950s. Thus, if we hope to discover any universal truths about root vectors, we better go back to $T^3 = I_3$ and $T^8 = \frac{\sqrt{3}}{2}Y$. In other words, we need to multiply the second component of the root vectors listed above by $\frac{\sqrt{3}}{2}$.

Denote the correctly normalized root vectors using a self-evident notation, and write

$$\begin{aligned}\vec{I}_{\pm} &= (\pm 1, 0) \\ \vec{U}_{\pm} &= \left(\mp \frac{1}{2}, \pm \frac{\sqrt{3}}{2}\right) \\ \vec{V}_{\pm} &= \left(\pm \frac{1}{2}, \pm \frac{\sqrt{3}}{2}\right)\end{aligned}\tag{21}$$

Remembering your equilateral triangle, you see that, interestingly, the root vectors are of equal length (length squared $= 1^2 + 0^2 = (\frac{1}{2})^2 + (\frac{\sqrt{3}}{2})^2 = 1$), and the angles between them are all equal to either 60° or 120° . See figure 1, known as a root diagram. Indeed, the 60° between a neighboring pair of root vectors merely reflects democracy: $\vec{I}_\pm$, $\vec{U}_\pm$, and $\vec{V}_\pm$ have exactly the same status. The root vectors $\vec{I}_\pm$ appears to be more special than $\vec{U}_\pm$ and $\vec{V}_\pm$ merely because of physics history.

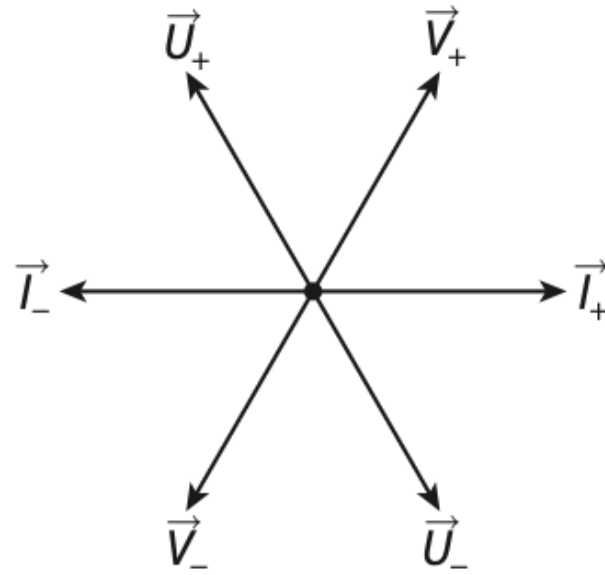


Figure 1

In this language, the root vectors of $SU(2)$ live in one dimension. A 1-dimensional vector is of course just a number.

Working with roots

From the root diagram, we can read off various commutation relations. At least we can see which ones ought to vanish. For example, $[I_-, U_+] = 0$.

From the root diagram, we can read off various commutation relations. At least we can see which ones ought to vanish. For example, $[I_-, U_+] = 0$.

To show this, let us first introduce some useful notation to save writing. Write the ket $|i_3, y\rangle$ as $|\vec{\omega}\rangle$; that is, define the vector $\vec{\omega} = (i_3, y)$. Then $I_- U_+ |\vec{\omega}\rangle \propto I_- |\vec{\omega} + \vec{U}_+\rangle \propto |\vec{\omega} + \vec{U}_+ + \vec{I}_-\rangle$. Similarly, $I_- U_+ |\vec{\omega}\rangle \propto I_- |\vec{\omega} + \vec{U}_+\rangle \propto |\vec{\omega} + \vec{I}_- + \vec{U}_+\rangle$. Thus, $[I_-, U_+] |\vec{\omega}\rangle \propto |\vec{\omega} + \vec{I}_- + \vec{U}_+\rangle$.

The proportionality factor could well vanish. If it does not vanish, then the root vector associated with $[I_-, U_+]$ would be pointing in the direction corresponding to $\vec{I}_- + \vec{U}_+$. But, referring to figure 1, we see that there isn't a root vector in that direction. Thus, $[I_-, U_+] = 0$.

With some practice, you can see almost instantly from the root diagram that certain commutators must vanish.

In contrast, the vector $\vec{I}_+ + \vec{U}_+$ points in the direction of $\vec{V}_+$, and sure enough, $[I_+, U_+] \propto V_+$.

Some “empirical facts”

From this example of $SU(3)$, we observe several “empirical facts.”

All the root vectors have equal length.

The angles between them have “nice” values, that is, not equal to something like 17° .

The sum of two roots is sometimes equal to another root and sometimes not.

But the sum of a root with itself does not exist; in other words, if $\vec{\alpha}$ is a root, then $2\vec{\alpha}$ (for example, $2\vec{V}_+ = (1, \sqrt{3})$) is not a root.

Positive and simple roots

Two notions about roots also naturally suggest themselves. Here we have six roots living in a 2-dimensional space. We ought not to have to specify all six; specifying two roots ought to suffice. To begin with, three of the six roots are just given by (-1) times the other three.

To define the notion of positive roots, we first have to agree to some ordering of the coordinate axes; here we put T^3 before T^8 . Then a root is said to be positive if its first nonzero component is positive. Referring to (21), we see that $\vec{I}_+$, $\vec{U}_-$, and $\vec{V}_+$ are positive.

Colloquially, the roots that point “mostly east” are defined to be positive. Note that which roots are positive depends on which axis we favor. If we had favored T^8 over T^3 , then $\vec{I}_+$, $\vec{U}_+$, and $\vec{V}_+$ would be called positive.

Of the three positive roots, $\vec{I}_+$, $\vec{U}_-$, and $\vec{V}_+$, any one of them can be written as a linear combination of the other two. Hence we can choose any two of the three positive roots as basic, but one particular choice recommends itself: speaking colloquially, we see that $\vec{I}_+$ is flanked by $\vec{U}_-$ and $\vec{V}_+$, that is, it lies inside the wedge or cone defined by $\vec{U}_-$ and $\vec{V}_+$. Note that $\vec{I}_+ = \vec{U}_- + \vec{V}_+$. The following more mathematical definition thus emerges naturally.

Given a set of positive roots in l -dimensional space, a subset is said to be simple if any of the positive roots can be written as a linear combination of the simple roots with non-negative coefficients. Thus, $\vec{U}_-$ and $\vec{V}_+$ are the two simple roots of $SU(3)$.

Thus far, I have carefully distinguished an operator, for example U_+ , from the root vector $\vec{U}_+$ associated with it. Now that you understand these concepts, I will often drop the arrow.

Weight diagrams

For a given irreducible representation $\mathcal{R}$ of $SU(3)$, we can plot each state in the representation as a point in the $(i_3, i_8) = (i_3, \frac{\sqrt{3}}{2}y)$ plane. The collection of points form a lattice known as the weight diagram for $\mathcal{R}$, the appropriate generalization of the ladder for $SU(2)$. The number of points in the weight diagram corresponds to the dimension of $\mathcal{R}$. (Recall that the ladder for the irreducible representation j of $SU(2)$ has $2j + 1$ rungs.)

The root vectors take us from one point in the weight diagram to another. They generalize the raising and lowering $J_{\pm}$ that take us from one rung of the ladder to another. Since we are talking about a finite-dimensional representation, a sequence of root vectors acting on a given state either takes us out of the weight diagram (that is, yields 0) or might eventually take us back to our starting point.

It is important to note what should be self-evident for most readers: for a given Lie algebra, there is only one root diagram, but there are many weight diagrams, one for each irreducible representation.

Let's see how this works for a few typical representations. Consider 3, the defining representation, and for convenience, use particle physics labels for the states in the representation. The u quark is located at $(i_3, \frac{\sqrt{3}}{2}y) = (\frac{1}{2}, \frac{1}{2\sqrt{3}})$, the d quark at $(-\frac{1}{2}, \frac{1}{2\sqrt{3}})$, and the s quark at $(0, -\frac{1}{\sqrt{3}})$. The weight diagram consists of three points forming an inverted equilateral triangle. See figure 2.

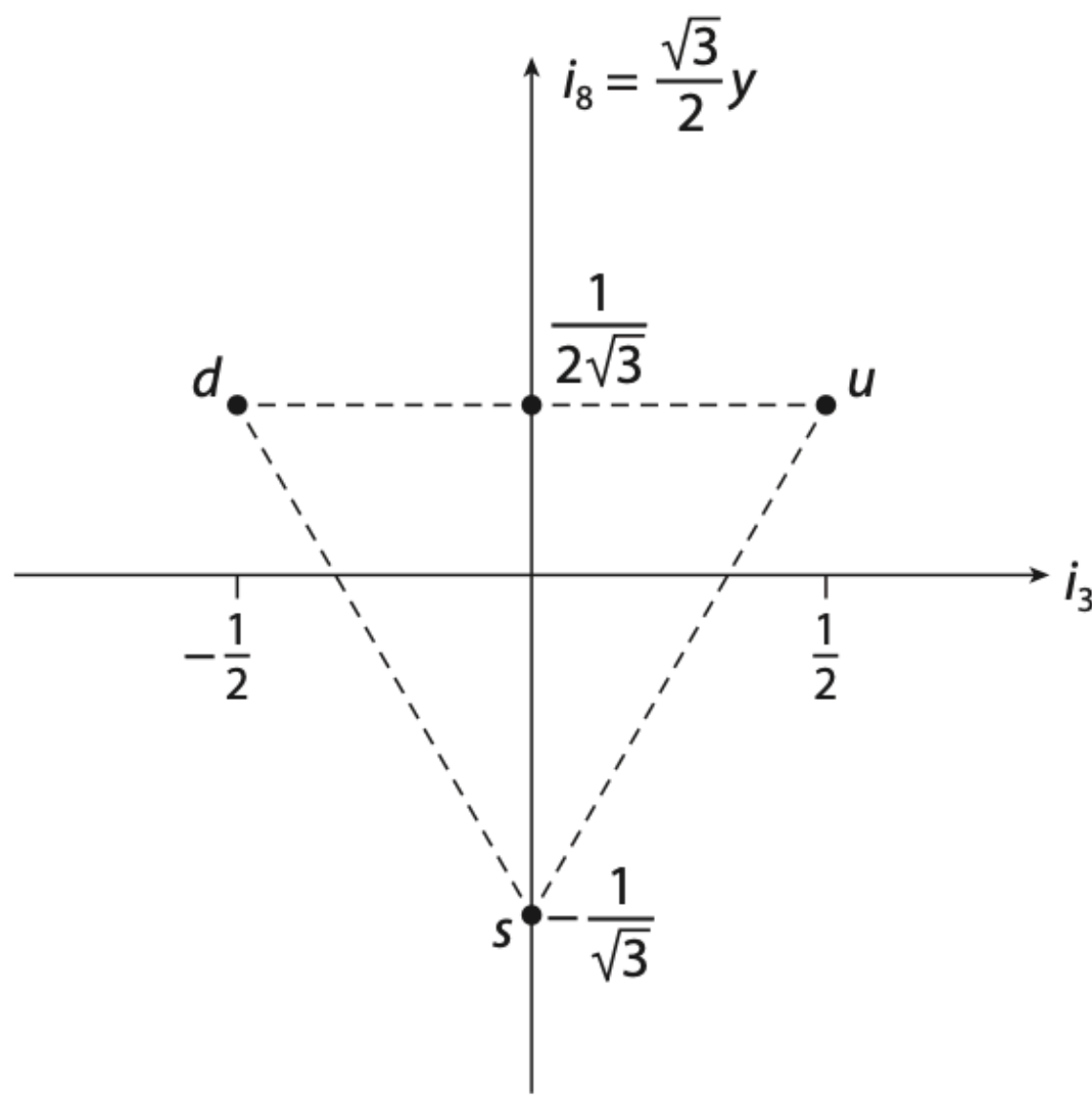


Figure 2

The roots listed in (21) take us from one point in the weight diagram to another point, just as $J_{\pm}$ take us from one rung on the ladder to the next in the $SU(2)$ case. For example,

$$\left(0, -\frac{1}{\sqrt{3}}\right) + \vec{V}_+ = \left(0, -\frac{1}{\sqrt{3}}\right) + \left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right) = \left(\frac{1}{2}, \frac{1}{2\sqrt{3}}\right) \quad (22)$$

In other words, $\vec{V}_+$ takes the s quark to the u quark.

Indeed, you can verify that the requirement that the roots in (21) connect three points in the $(i_3, \frac{\sqrt{3}}{2}y)$ plane fixes the positions of the 3 points and the weight diagram. Analogously, in $SU(2)$, if we require the ladder to terminate, we cannot start at an arbitrary value of m : it has to be integral or half-integral. Indeed, since $SU(3)$ contains three overlapping $SU(2)$ algebras, the argument is basically a direct repeat of the argument we went through for $SU(2)$: I spin seems privileged compared to U spin and V spin merely because we chose the horizontal axis in our plot to describe i_3 .

But $U_{\pm}$ and $V_{\pm}$ are in no way second class citizens compared to $I_{\pm}$. The attentive reader might have noticed that I have quietly defined, in (10) and (11), $[U_+, U_-] = 2U_3 \equiv \frac{3}{2}Y - I_3 = \sqrt{3}T_8 - T_3$ and $[V_+, V_-] = 2V_3 \equiv \frac{3}{2}Y + I_3 = \sqrt{3}T_8 + T_3$. You can immediately verify, as you would expect, that the d and s quarks form a U spin doublet.

As an exercise, work out the weight diagram of 3^* .

We can do another example. Requiring that the roots connect eight points fixes the weight diagram for the famous octet into which Gell-Mann placed the nucleons and their cousins. Alternatively, start with the known position of the proton $(\frac{1}{2}, \frac{\sqrt{3}}{2})$. We already know that $\vec{I}_-$ takes the proton to the neutron at $(-\frac{1}{2}, \frac{\sqrt{3}}{2})$. Let us act with $\vec{U}_-$ on the proton: $(\frac{1}{2}, \frac{\sqrt{3}}{2}) + (\frac{1}{2}, -\frac{\sqrt{3}}{2}) = (1, 0)$. We get to the Σ^+ hyperon. (See figure 3.)

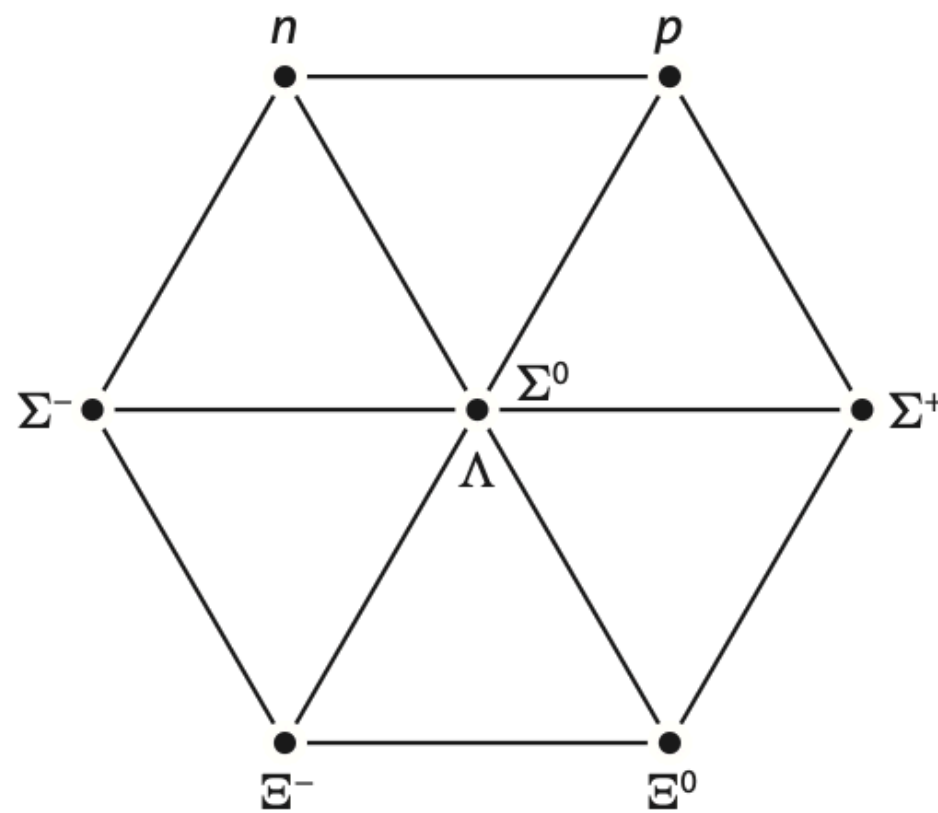


Figure 3

Suppose we apply $\vec{U}_-$ again, now on the Σ^+ hyperon. We get 0, just like applying the generator J_- of $SU(2)$ to the bottom rung of the ladder produces 0. (Again, see figure 3.) On the other hand, applying $\vec{V}_-$ to Σ^+ gets us to $(1, 0) + (-\frac{1}{2}, -\frac{\sqrt{3}}{2}) = (\frac{1}{2}, -\frac{\sqrt{3}}{2})$, which we recognize as the Ξ^0 . In this way, we built the 8 of the baryons proposed by Gell-Mann.

As an exercise, you should now work out the weight diagram of 10 using the root vectors. See figure 4.

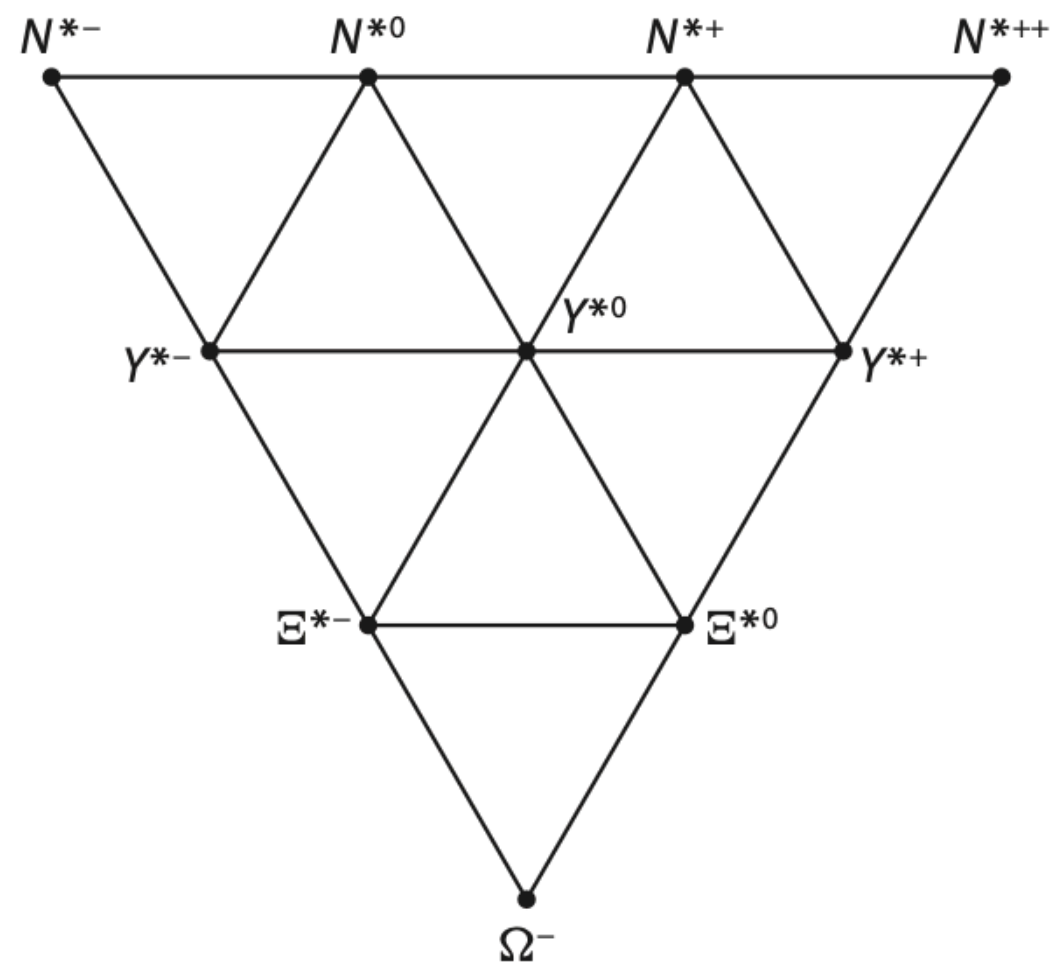


Figure 4

Theorems about weight diagrams

Clearly, due to the hexagonal symmetry of the root vectors, the weight diagrams all exhibit hexagonal symmetry.

It is possible to prove various theorems about the shape of the weight diagrams.

For example, the boundary shown schematically in figure 5 is not possible.

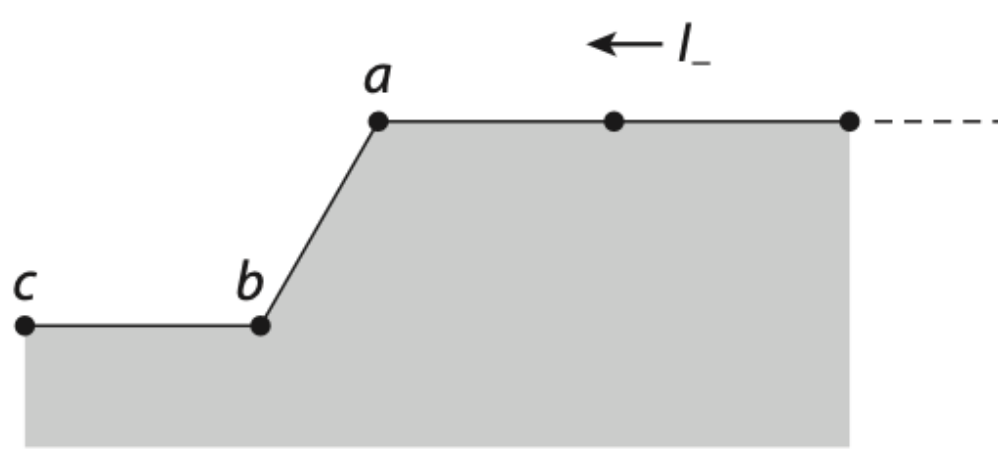


Figure 5

Roughly speaking, weight diagrams have to be convex outward.

To prove this, label the states as shown in the figure. (The shaded region in figure 5 indicates the interior of the weight diagram.)

Along the top of the weight diagram, we keep lowering with I_- until we get to the state $|a\rangle$ when the I spin ladder terminates: $I_- |a\rangle = 0$. As shown, $V_- |a\rangle = \beta |b\rangle$ and $I_- |b\rangle = \gamma |c\rangle$, with β and γ two Clebsch-Gordan coefficients. Applying I_- to $V_- |a\rangle = \beta |b\rangle$ then gives

$$I_- V_- |a\rangle = \beta I_- |b\rangle = \beta \gamma |c\rangle = ([I_-, V_-] + V_- I_-) |a\rangle = [I_-, V_-] |a\rangle = 0 \quad (23)$$

The last equality follows from (15). Hence either β or γ vanishes, or both vanish, which shows that the figure as drawn is not allowed. In the 10, for example, the state $|b\rangle$ does not exist; the boundary continues in the U_- direction. In contrast, in the 8, $|b\rangle$ exists but not $|c\rangle$; at $|b\rangle$ the boundary continues also in the U_- direction.

Clearly, we can also learn a lot about the weight diagrams of $SU(3)$ from the fact that the algebra consists of three overlapping $SU(2)$ algebras. For example, slicing the weight diagram along the U direction, we should see the states lining up neatly into U spin multiplets. Thus, as we have already seen, in the 8 of baryons, the proton and the Σ^+ form a U spin doublet.

Group Theory Guides Us into the Microscopic World

To a generation who grew up with superstrings and the multiverse, even quarks and gluons may sound old hat. But not that long ago, they were totally unknown, and utter confusion reigned in the subnuclear world of the hadrons. The strong interaction was gradually unveiled only through decades of assiduous experimental and theoretical work. In this endeavor, group theory was an indispensable guide.

We now know that $SU(3)$ symmetry, as explained earlier, is due to the approximate invariance of the strong interaction under a unitary transformation of the up u , down d , and strange s quarks into one another, approximate because these three quarks have different masses. But none of that was known around 1960.

Here we discuss one particular success of $SU(3)$ and its breaking, namely, the Gell-Mann-Okubo mass formula.¹ To focus on $SU(3)$ breaking, we assume that Heisenberg's $SU(2)$ is not broken. Readers totally unfamiliar with quantum mechanics might wish to skip this part.

SU(3) breaking

In a world with exact $SU(3)$ symmetry, the π , K , and η mesons would all have the same mass. In the language of chapter III.1, they would be degenerate. But in fact their masses are quite different. Our world is only roughly $SU(3)$ invariant.

Without a full understanding of the strong interaction, we had (and still have) no way of calculating these meson masses analytically. Instead, what particle physicists accomplished in the 1960s amounted to an impressive mix of science and art, of group theory and inspired guesses.

For readers not about to specialize in particle physics (which may well be most readers), the details of meson mass splitting are not important. The important point to grasp is that what we are discussing here is conceptually similar to other symmetry breaking phenomena, such as crystal field splitting discussed in interlude IV.i2. As another example, for the triangular “molecule” consisting of masses tied together by springs discussed in chapter III.2, by making one spring slightly different from the other two, we break the symmetry group, and the shifts in the eigenfrequencies can be studied using the methods discussed here. As we shall see, Gell-Mann and Okubo, by making a simplifying* assumption about the nature of the symmetry-breaking interaction, were able to go further than just figuring out the pattern of how the 8-dimensional irreducible representation of $SU(3)$ breaks upon restriction to $SU(2)$.

The meson octet

Let us start with the eight mesons, π^+ , π^0 , π^- , K^+ , K^0 , K^- , $\bar{K}^0$, and η , whose acquaintance we made earlier. They were postulated to populate the adjoint representation 8 of $SU(3)$.

Since we are assuming that the $SU(2)$ isospin subgroup is unbroken, π^+ , π^0 , and π^- have the same mass. Similarly, K^+ and K^0 have the same mass, and K^- and $\bar{K}^0$ also. However, since K^- is the antiparticle of K^+ , they have the same mass, and thus also K^0 and $\bar{K}^0$. The data thus consists of three masses, and so if group theory restricts the dynamics sufficiently to reduce the number of unknown constants to less than three, we would have a prediction, or strictly, a postdiction.

We will use the tensor and effective field theory approach to discuss meson masses. By now, you know well that the adjoint is given by a traceless tensor Φ^i_j , $i, j = 1, 2, 3$ and can thus be written as a traceless 3-by-3 matrix:

$$\Phi = \frac{1}{\sqrt{2}} \sum_{a=1}^8 \phi_a \lambda_a = \begin{pmatrix} \frac{1}{\sqrt{2}}\pi^0 + \frac{1}{\sqrt{6}}\eta & \pi^+ & K^+ \\ \pi^- & -\frac{1}{\sqrt{2}}\pi^0 + \frac{1}{\sqrt{6}}\eta & K^0 \\ K^- & \bar{K}^0 & -\frac{2}{\sqrt{6}}\eta \end{pmatrix} \quad (1)$$

In the first equality, we merely expand Φ in terms of the eight Gell-Mann matrices λ_a (as given in (V.3.1)). The eight fields ϕ_a may be thought of as the Cartesian components of Φ . In the second equality, the fields are identified by their street names, thus $\pi^0 = \phi_3$, $\eta = \phi_8$, and so on.

In the first equality, we merely expand Φ in terms of the eight Gell-Mann matrices λ_a (as given in (V.3.1)). The eight fields ϕ_a may be thought of as the Cartesian components of Φ . In the second equality, the fields are identified by their street names, thus $\pi^0 = \phi_3$, $\eta = \phi_8$, and so on.

Note that the 2-by-2 submatrix contained in the northwest corner of Φ , with η removed, is precisely $\frac{1}{\sqrt{2}}$ times the matrix φ discussed in chapter V.1 when we discussed Heisenberg's $SU(2)$. In particular, the relative factor of $\sqrt{2}$ between π^0 and π^+ was explained there.

Let us warm up by imagining a world with exact $SU(3)$ symmetry and write down the effective Lagrangian

$$\mathcal{L} = \frac{1}{2} \text{tr} \left((\partial_\mu \Phi)^2 - m_0^2 \Phi^2 \right) = \frac{1}{2} \sum_{a=1}^8 \left((\partial_\mu \phi_a)^2 - m_0^2 \phi_a^2 \right) \quad (2)$$

Here we used the normalization $\text{tr} \lambda_a \lambda_b = 2\delta_{ab}$ adopted in chapter V.3 so that, for example, $\text{tr} \Phi^2 = \text{tr} \left(\frac{1}{\sqrt{2}} \sum_{a=1}^8 \phi_a \lambda_a \right)^2 = \frac{1}{2} \sum_{a=1}^8 \sum_{b=1}^8 \phi_a \phi_b (\text{tr} \lambda_a \lambda_b) = \sum_{a=1}^8 \phi_a^2$. In this imaginary world, the eight mesons would be equally massive, with mass squared $= m_0^2$.

But $SU(3)$ is broken in our world.

We now have to appeal to basic quantum mechanics. Suppose we have found the eigenstates and eigenvalues of a Hamiltonian H_0 . Consider another Hamiltonian $H = H_0 + \alpha H_1$. In general, finding the eigenstates and eigenvalues of H would pose an entirely different problem. However, if α is known to be small, then to leading order in α , we would expect the eigenstates and eigenvalues of H to be closely approximated by the eigenstates and eigenvalues of H_0 . As some readers may know, this is known as perturbation theory: αH_1 is treated as a small perturbation on H_0 . One important result of perturbation theory is that to order α , the eigenvalue of the eigenstate $|s\rangle$ of H_0 is shifted by the amount $\langle s | \alpha H_1 | s \rangle$. (To this order, knowing the eigenstate $|s\rangle$ of H_0 suffices; we do not have to worry about how $|s\rangle$ is shifted by the perturbation.)

Indeed, I have already alluded to this perturbative approach to atomic spectrum in chapter IV.3 and mentioned the Wigner-Eckart theorem. If we follow this approach, then we will have to calculate quantities like $\langle 8, a | H_1 | 8, b \rangle$, with $|8, b\rangle$ the eigenstates in an $SU(3)$ symmetric world. Even to this day, we are far from being able to calculate such quantities starting from first principles; instead, we rely on group theory to tell us something. As already mentioned, rather than deriving the analog of the Wigner-Eckart theorem for $SU(3)$, we follow the effective Lagrangian approach.

The Gell-Mann-Okubo mass formula for mesons

You might question whether first order perturbation theory is applicable at all: back in chapter V.2, we learned that $SU(3)$ is badly broken, with the corresponding α at least as large as $\sim 0.2\text{--}0.3$. Nevertheless, Gell-Mann (and others at that time) bravely pushed ahead.

To apply group theory, we have to specify how the $SU(3)$ breaking interaction transforms under $SU(3)$. In $\mathcal{L}$, the mass term is quadratic in Φ ; it transforms like the product of two Φ s. But we have already worked out, in chapter V.2, that $8 \otimes 8 = 27 \oplus 10 \oplus 10^* \oplus 8 \oplus 8 \oplus 1$. However, we had multiplied two different traceless tensors together. Here we are multiplying Φ by itself, and so the antisymmetric part of the product just given drops out. If you go through the multiplication we did, you would see² that the symmetric part of $8 \otimes 8$ is given by

$$(8 \otimes 8)_s = 27 \oplus 8 \oplus 1 \tag{3}$$

while the antisymmetric part is given by

$$(8 \otimes 8)_a = 10 \oplus 10^* \oplus 8 \tag{4}$$

One easy check is that the counting works out: $8 \cdot (8 + 1)/2 = 36 = 27 + 8 + 1$, while $8 \cdot (8 - 1)/2 = 28 = 10 + 10 + 8$.

Therefore, there are two possible $SU(3)$ -breaking terms, one transforming like the 27 and the other like the adjoint 8. (The term transforming like 1 in (3) does not break $SU(3)$, of course.)

Boldly, Gell-Mann and Okubo (working independently) proposed to throw the 27 out, as Feynman (later) advised Zweig. Thus, to break $SU(3)$, we add the term* $\text{tr } \Phi^2 \lambda_8$ to the effective Lagrangian $\mathcal{L}$.

It is straightforward to multiply three 3-by-3 matrices together and then take the trace to obtain $\text{tr } \Phi^2 \lambda_8$. But with some forethought, we can minimize our arithmetical exertions. Observe that $\lambda_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}$ can be written as a linear combination of the 3-by-3 identity matrix and the matrix $\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. The contribution of the identity matrix to $\text{tr } \Phi^2 \lambda_8$ (namely, $\text{tr } \Phi^2$) can be absorbed into the m_0^2 term in $\mathcal{L}$. Thus, the $SU(3)$ -breaking term in $\mathcal{L}$ is effectively given by

$$\begin{aligned} \text{tr } \Phi^2 \lambda_8 &\rightarrow \text{tr } \Phi \Phi \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \text{tr} \begin{pmatrix} X & X & X \\ X & X & X \\ K^- & \bar{K}^0 & -\frac{2}{\sqrt{6}}\eta \end{pmatrix} \begin{pmatrix} 0 & 0 & K^+ \\ 0 & 0 & K^0 \\ 0 & 0 & -\frac{2}{\sqrt{6}}\eta \end{pmatrix} \\ &= K^- K^+ + K^0 \bar{K}^0 + \frac{2}{3}\eta^2 \end{aligned} \tag{5}$$

Here X denotes those entries we won't bother to write down, since we intend to take the trace after multiplying the two matrices.

The mass of the K and of the η are shifted, but not that of the π . To get the normalization right, it is easiest to compare with the $SU(3)$ invariant mass term $\text{tr } \Phi^2 = (\pi^0)^2 + \eta^2 + 2K^-K^+ + \dots$. Since we know that this term implies that the K meson and the η meson have equal mass, we have all together $m_\pi^2 = m_0^2$, $m_K^2 = m_0^2 + \frac{1}{2}\kappa$, and $m_\eta^2 = m_0^2 + \frac{2}{3}\kappa$, where κ is an unknown constant.

Happy are we. With three masses and two unknown constants, we obtain one relation, namely, the Gell-Mann Okubo mass formula for mesons,

$$4m_K^2 = 3m_\eta^2 + m_\pi^2 \tag{6}$$

You can plug in the measured masses given in chapter V.2 and verify that this mass formula is satisfied to within the expected accuracy.

Note that if we had included the 27 as well as the 8, we would have an additional unknown constant and thus lose the mass formula (6). In 1961–1962, the proposal that the $SU(3)$ -breaking term transforms purely like the 8 was just a guess. As I said earlier, physics at the cutting edge often involves inspired guesses, and is art as well as science.

This guess is now understood in terms of quantum chromodynamics, in which $SU(3)$ breaking is entirely due to the strange quark being much more massive than the up and down quarks. Since these three quarks transform as the defining 3, their mass terms transform as $3^* \otimes 3 = 1 \oplus 8$, as was worked out earlier.

Baryon mass splitting

We next turn to the eight baryons, p , n , Σ^+ , Σ^0 , Σ^- , Λ^0 , Ξ^- , and Ξ^0 , which we met earlier, and which form the baryon octet. As in (1), the eight baryon fields³ can be arranged in a 3-by-3 traceless matrix:

$$B = \begin{pmatrix} \frac{1}{\sqrt{2}}\Sigma^0 + \frac{1}{\sqrt{6}}\Lambda^0 & \Sigma^+ & p \\ \Sigma^- & -\frac{1}{\sqrt{2}}\Sigma^0 + \frac{1}{\sqrt{6}}\Lambda^0 & n \\ \Xi^- & \Xi^0 & -\frac{2}{\sqrt{6}}\Lambda^0 \end{pmatrix} \quad (7)$$

Now that we have worked through the derivation of the mass formula for the mesons, we should be able to breeze through a similar derivation for the baryons. However, there are some differences.

One difference is that, unlike the mesons, which carry spin 0, the baryons carry spin $\frac{1}{2}$ and thus transform under the Lorentz group (which we will study in part VII). Thus, as already mentioned in chapter V.1, the hermitean conjugate of B transforms differently from B and is denoted by

$$\bar{B} = \begin{pmatrix} \frac{1}{\sqrt{2}}\bar{\Sigma}^0 + \frac{1}{\sqrt{6}}\bar{\Lambda}^0 & \bar{\Sigma}^- & \bar{\Xi}^- \\ \bar{\Sigma}^+ & -\frac{1}{\sqrt{2}}\bar{\Sigma}^0 + \frac{1}{\sqrt{6}}\bar{\Lambda}^0 & \bar{\Xi}^0 \\ \bar{p} & \bar{n} & -\frac{2}{\sqrt{6}}\bar{\Lambda}^0 \end{pmatrix} \quad (8)$$

the

the

the

the

the

the

the

Consequently, we now have two possible $SU(3)$ -breaking terms that transform as 8 (rather than the single term in the meson case), namely,

$$\text{tr } \bar{B} B \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \text{tr } \bar{B} \begin{pmatrix} 0 & 0 & p \\ 0 & 0 & n \\ 0 & 0 & -\frac{2}{\sqrt{6}}\Lambda^0 \end{pmatrix} = (\bar{p}p + \bar{n}n + \frac{2}{3}\bar{\Lambda}^0\Lambda^0) \quad (9)$$

and

$$\text{tr } \bar{B} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} B = \text{tr } \begin{pmatrix} 0 & 0 & -\bar{\Xi}^- \\ 0 & 0 & \bar{\Xi}^0 \\ 0 & 0 & -\frac{2}{\sqrt{6}}\bar{\Lambda}^0 \end{pmatrix} B = (\bar{\Xi}^- \Xi^- + \bar{\Xi}^0 \Xi^0 + \frac{2}{3}\bar{\Lambda}^0\Lambda^0) \quad (10)$$

We see that (i) the two terms have different effects and (ii) the symmetry breaking does not affect the Σ s. Again, we calculate $\text{tr } \bar{B} B$ to make sure that things are properly normalized. To save computational effort, we can set the fields Σ, n, Ξ^0 to zero in light of the preceding remark and since we are not breaking isospin; then $\text{tr } \bar{B} B = \bar{p}p + \bar{\Lambda}^0\Lambda^0 + \bar{\Xi}^- \Xi^- + \dots$.

Denoting the unknown coefficients of the two terms in (9) and (10) by κ and ζ , we thus obtain $M_N = M_0 + \kappa$, $M_\Lambda = M_0 + \frac{2}{3}(\kappa + \zeta)$, $M_\Xi = M_0 + \zeta$, $M_\Sigma = M_0$. This time around there are three unknowns, but we have four masses. So we still get one mass relation, namely,*

$$3M_\Lambda + M_\Sigma = 2(M_N + M_\Xi) \quad (11)$$

The attentive reader will have noticed that I wrote down masses in (11) but masses squared in (6). You will see in chapter VII.3 that in the field theory Lagrangian for spin $\frac{1}{2}$ particles, mass appears, but for spin 0 particles, mass squared appears (as was written in chapter IV.9 generalizing the harmonic oscillator). Actually, since we are working to linear order in the symmetry breaking, we have no right to quibble about the quadratic versus the linear mass formula. Expanding $m_a^2 = (m_0 + \delta m_a)^2 = m_0^2 + 2m_0\delta m_a + O((\delta m_a)^2)$, we could just as well have written (6) as $4m_K = 3m_\eta + m_\pi$.

Consequences of $SU(3)$

The consequences⁴ of $SU(3)$ were soon worked out in the early 1960s. With the setup given in this chapter, you can join in the fun retroactively. For instance, in chapter V.1, we worked out the pion-nucleon couplings. You can now work out the couplings of the meson octet to the baryon octet. Two terms are allowed by $SU(3)$:

$$g_1 \text{tr } \bar{B} B \Phi + g_2 \text{tr } \bar{B} \Phi B \quad (12)$$

In principle, 17 coupling constants (for example, the $\bar{N} K \Sigma$ coupling) are measurable. That they can be expressed in terms of two unknowns⁵ is an impressive achievement of group theory.

Our faith in Nature's simplicity

We close with a story. George Zweig, who discovered quarks independently⁶ of Gell-Mann, recalled that he thought that the weak interaction currents of the strongly interacting particles should also be classified in representations of $SU(3)$, and that both the 8- and 27-dimensional representations were to be used. At the time, a decay process requiring the presence of the 27 had been seen by an experimentalist of high reputation⁷ working with a strong team using a well understood technique.

Zweig went to talk to Feynman, who liked the idea of applying $SU(3)$ to the decay, but kept saying “Throw away the 27!”

Feynman turned out to be right; the experimentalists were mistaken.

We need to use group theory in physics, but somehow we never have to mess with anything beyond the simplest irreducible representations. Another example verifying the principle of Nature's kindness to theoretical physicists!