

Symmetry in Physics

5CCP2322

Lecturer: Prof. Sarben Sarkar
sarben.sarkar@kcl.ac.uk
Office: S7.19
Department of Physics
King's College London

The material in this and subsequent sections can be considered classic and is widely available. These notes make no claim to originality but I hope the notes can serve as a summary of the material that the student will be expected to know at the end of the course.

Motivation and introduction

Common Parlance:

Symmetry \leftrightarrow Harmony of proportions

Geometrical objects:

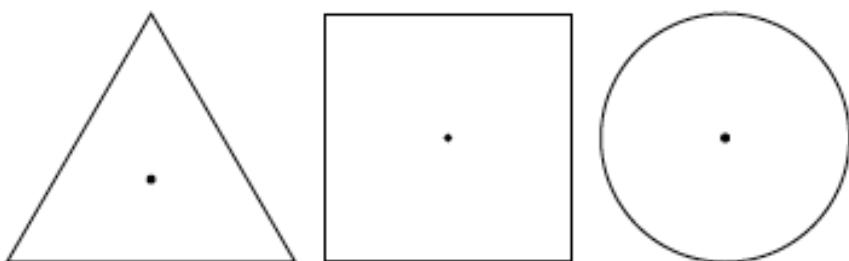


Figure 1: Geometrical objects: equilateral triangle, square and circle

Where ‘•’ represents an axis coming out of the paper. Triangle is symmetric under rotations of $2\pi/3$ about the axis. Square is symmetric under rotations of $2\pi/4 = \pi/2$ about the axis. Hexagon is symmetric under $2\pi/6$ rotations. For a regular polygon with n vertices the polygon is symmetric under rotations by $2\pi/n$. As $n \rightarrow \infty$ the polygon approaches a circle. So circle is symmetric under arbitrary rotation.

Laws of physics can also have symmetries: Consider a 3-dimensional classical wave-equation

$$\frac{1}{c^2} \frac{\partial^2 \phi}{\partial t^2} = \frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} + \frac{\partial^2 \phi}{\partial z^2} \quad (3.1)$$

where $\phi(t, x, y, z)$ is a scalar field. Such equations arise in electromagnetism. Consider the transformations of space-time

$$\begin{aligned} x' &= \gamma(x - vt) \\ y' &= y \\ z' &= z \\ t' &= \gamma\left(t - \frac{v}{c^2}x\right) \end{aligned} \quad (3.2)$$

where $\gamma = \sqrt{1 - v^2/c^2}$.

Let

$$\phi'(t', x', y', z') = \phi(t, x, y, z) \quad (3.3)$$

then

$$\frac{1}{c^2} \frac{\partial^2 \phi'}{\partial t'^2} = \frac{\partial^2 \phi'}{\partial x'^2} + \frac{\partial^2 \phi'}{\partial y'^2} + \frac{\partial^2 \phi'}{\partial z'^2} \quad (3.4)$$

The transformations (3.2) are known as Lorentz transformations, and the wave equation is covariant under the transformation.

Quantum mechanical example:

Time-independent Schrödinger equation for the one-dimensional motion of a particle of mass m in a potential $V(x)$:

$$\left[-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + V(x) \right] \psi(x) = E\psi(x) \quad (3.5)$$

where ψ is the wavefunction and E is the energy of the particle. Suppose under $x \rightarrow -x$ (inversion),

$$V(x) = V(-x)$$

and of course

$$\frac{d^2}{d(-x)^2} = \frac{d^2}{dx^2}.$$

Examples of an even $V(x)$:

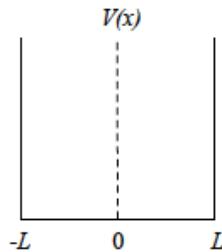


Figure 2: Infinite square-well potential

Another example is

$$V(x) = \frac{1}{2}kx^2, \quad k > 0 \quad (3.6)$$

For even-potentials

$$\left[-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + V(x) \right] \psi(-x) = E\psi(-x) \quad (3.7)$$

A property of the Schrödinger equation is that the eigenvalues are non-degenerate, so

$$\psi(-x) = \alpha\psi(x) \quad (3.8)$$

where α is a (complex) constant. Under 2 inversions

$$\psi(x) = \alpha^2\psi(x) \quad (3.9)$$

so $\alpha^2 = 1$ which leads to $\alpha = \pm 1$.

\therefore wavefunction $\psi(x)$ under inversion are either even

$$\psi(-x) = \psi(x) \quad (3.10)$$

or odd

$$\psi(-x) = -\psi(x) \quad (3.11)$$

Why should we be concerned with symmetry? In the first place, symmetry is fascinating to the human mind, and everyone likes objects or patterns that are in some way symmetrical. It is an interesting fact that nature often exhibits certain kinds of symmetry in the objects we find in the world around us. Perhaps the most symmetrical object imaginable is a sphere, and nature is full of spheres—stars, planets, water droplets in clouds. The crystals found in rocks exhibit many different kinds of symmetry, the study of which tells us some important things about the structure of solids. Even the animal and vegetable worlds show some degree of symmetry, although the symmetry of a flower or of a bee is not as perfect or as fundamental as is that of a crystal.

But our main concern here is not with the fact that the *objects* of nature are often symmetrical. Rather, we wish to examine some of the even more remarkable symmetries of the universe—the symmetries that exist in the *basic laws themselves* which govern the operation of the physical world.

First, what *is* symmetry? How can a physical *law* be “symmetrical”? The problem of defining symmetry is an interesting one and we have already noted that Weyl gave a good definition, the substance of which is that a thing is symmetrical if there is something we can do to it so that after we have done it, it looks the same as it did before. For example, a symmetrical vase is of such a kind that if we reflect or turn it, it will look the same as it did before. The question we wish to consider here is what we can do to physical phenomena, or to a physical situation in an experiment, and yet leave the result the same. A list of the known operations under which various physical phenomena remain invariant is shown in the table below:

Symmetry Operations

Translation in space
Translation in time
Rotation through a fixed angle
Uniform velocity in a straight line (Lorentz transformation)
Reversal of time
Reflection of space
Interchange of identical atoms or identical particles
Quantum-mechanical phase
Matter-antimatter (charge conjugation)

Symmetry in space and time

The first thing we might try to do, for example, is to *translate* the phenomenon in space. If we do an experiment in a certain region, and then build another apparatus at another place in space (or move the original one over) then, whatever went on in one apparatus, in a certain order in time, will occur in the same way if we have arranged the same condition, with all due attention to the restrictions that we mentioned before: that all of those features of the environment which make it not behave the same way have also been moved over—we talked about how to define how much we should include in those circumstances, and we shall not go into those details again.

In the same way, we also believe today that *displacement in time* will have no effect on physical laws. (That is, *as far as we know today*—all of these things are as far as we know today!) That means that if we build a certain apparatus and start it at a certain time, say on Thursday at 10:00 a.m., and then build the same apparatus and start it, say, three days later in the same condition, the two apparatuses will go through the same motions in exactly the same way as a function of time no matter what the starting time, provided again, of course, that the relevant features of the environment are also modified appropriately in *time*. That symmetry means, of course, that if one bought General Motors stock three months ago, the same thing would happen to it if he bought it now!

We have to watch out for geographical differences too, for there are, of course, variations in the characteristics of the earth's surface. So, for example, if we measure the magnetic field in a certain region and move the apparatus to some other region, it may not work in precisely the same way because the magnetic field is different, but we say that is because the magnetic field is associated with the earth. We can imagine that if we move the whole earth and the equipment, it would make no difference in the operation of the apparatus.

Another thing that we can discuss is rotation in space: if we turn an apparatus at an angle it works just as well, provided we turn everything else that is relevant along with it.

On a more advanced level we had another symmetry—the symmetry under uniform velocity in a straight line. That is to say—a rather remarkable effect—that if we have a piece of apparatus working a certain way and then take the same apparatus and put it in a car, and move the whole car, plus all the relevant surroundings, at a uniform velocity in a straight line, then so far as the phenomena inside the car are concerned there is no difference: all the laws of physics appear the same. We even know how to express this more technically, and that is that the mathematical equations of the physical laws must be unchanged under a *Lorentz transformation*. As a matter of fact, it was a study of the relativity problem that concentrated physicists' attention most sharply on symmetry in physical laws.

Now the above-mentioned symmetries have all been of a *geometrical* nature, time and space being more or less the same, but there are other symmetries of a different kind. For example, there is a symmetry which describes the fact that we can replace one atom by another of the same kind; to put it differently, there *are* atoms of the same kind. It is possible to find groups of atoms such that if we change a pair around, it makes no difference—the atoms are identical. Whatever one atom of oxygen of a certain type will do, another atom of oxygen of that type will do. One may say, "That is ridiculous,

that is the *definition* of equal types!" That may be merely the definition, but then we still do not know whether there *are* any "atoms of the same type"; the *fact* is that there are many, many atoms of the same type. Thus it does mean something to say that it makes no difference if we replace one atom by another of the same type. The so-called elementary particles of which the atoms are made are also identical particles in the above sense—all electrons are the same; all protons are the same; all positive pions are the same; and so on.

After such a long list of things that can be done without changing the phenomena, one might think we could do practically anything; so let us give some examples to the contrary, just to see the difference. Suppose that we ask: "Are the physical laws symmetrical under *a change of scale*?" Suppose we build a certain piece of apparatus, and then build another apparatus five times bigger in every part, will it work exactly the same way? The answer is, in this case, *no!* The wavelength of light emitted, for example, by the atoms inside one box of sodium atoms and the wavelength of light emitted by a gas of sodium atoms five times in volume is not five times longer, but is in fact exactly the same as the other. So the ratio of the wavelength to the size of the emitter will change.

Another example: we see in the newspaper, every once in a while pictures of a great cathedral made with little matchsticks—a tremendous work of art by some retired fellow who keeps gluing matchsticks together. It is much more elaborate and wonderful than any real cathedral. If we imagine that this wooden cathedral were actually built on the scale of a real cathedral, we see where the trouble is; it would not last—the whole thing would collapse because of the fact that scaled-up matchsticks are just not strong enough. "Yes," one might say, "but we also know that when there is an influence from the outside, it also must be changed in proportion!" We are talking about the ability of the object to withstand gravitation. So what we should do is first to take the model cathedral of real matchsticks and the real earth, and then we know it is stable. Then we should take the larger cathedral and take a bigger earth. But then it is even worse, because the gravitation is increased still more!

Today, of course, we understand the fact that phenomena depend on the scale on the grounds that matter is atomic in nature, and certainly if we built an apparatus that was so small there were only five atoms in it, it would clearly be something we could not scale up and down arbitrarily. The scale of an individual atom is not at all arbitrary—it is quite definite.

The fact that the *laws of physics are not unchanged under a change of scale* was discovered by Galileo. He realized that the strengths of materials were not in exactly the right proportion to their sizes, and he illustrated this property that we were just discussing, about the cathedral of matchsticks, by drawing two bones, the bone of one dog, in the right proportion for holding up his weight, and the imaginary bone of a "super dog" that would be, say, ten or a hundred times bigger—that bone was a big, solid thing with quite different proportions. We do not know whether he ever carried the argument quite to the conclusion that the laws of nature must have a definite scale, but he was so impressed with this discovery that he considered it to be as important as the discovery of the laws of motion, because he published them both in the same volume, called "On Two New Sciences."

Another example in which the laws are not symmetrical, that we know quite well, is this: *a system in rotation at a uniform angular velocity does not give the same apparent laws as one that is not rotating*. If we make an experiment and then put everything in a space ship and have the space ship spinning in empty space, all alone at a constant angular velocity, the apparatus will not work the same way because, as we know, things inside the equipment will be thrown to the outside, and so on, by the centrifugal or Coriolis forces, etc. In fact, we can tell that the earth is rotating by using a so-called Foucault pendulum, without looking outside.

Next we mention a very interesting symmetry which is obviously false, i.e., *reversibility in time*. The physical laws apparently cannot be reversible in time, because, as we know, all obvious phenomena are irreversible on a large scale: "The moving finger writes, and having writ, moves on." So far as we can tell, this irreversibility is due to the very large number of particles involved, and if we could see the individual molecules, we would not be able to discern whether the machinery was working forward or backwards. To make it more precise: we build a small apparatus in which we know what all the atoms are doing, in which we can watch them jiggling. Now we build another apparatus like it, but which starts its motion in the final condition of the other one, with all the velocities precisely reversed. *It will then go through the same motions, but exactly in reverse.* Putting it another way: if we take a motion

picture, with sufficient detail, of all the inner works of a piece of material and shine it on a screen and run it backwards, no physicist will be able to say, "That is against the laws of physics, that is doing something wrong!" If we do not see all the details, of course, the situation will be perfectly clear. If we see the egg splattering on the sidewalk and the shell cracking open, and so on, then we will surely say, "That is irreversible, because if we run the moving picture backwards the egg will all collect together and the shell will go back together, and that is obviously ridiculous!" But if we look at the individual atoms themselves, the laws look completely reversible. This is, of course, a much harder discovery to have made, but apparently it is true that the fundamental physical laws, on a microscopic and fundamental level, are completely reversible in time!

Symmetry and conservation laws

The symmetries of the physical laws are very interesting at this level, but they turn out, in the end, to be even more interesting and exciting when we come to quantum mechanics. For a reason which we cannot make clear at the level of the present discussion—a fact that most physicists still find somewhat staggering, a most profound and beautiful thing, is that, in quantum mechanics, *for each of the rules of symmetry there is a corresponding conservation law*; there is a definite connection between the laws of conservation and the symmetries of physical laws. We can only state this at present, without any attempt at explanation.

The fact, for example, that the laws are symmetrical for translation in space when we add the principles of quantum mechanics, turns out to mean that *momentum is conserved*.

That the laws are symmetrical under translation in time means, in quantum mechanics, that *energy is conserved*.

Invariance under rotation through a fixed angle in space corresponds to the *conservation of angular momentum*. These connections are very interesting and beautiful things, among the most beautiful and profound things in physics.

Incidentally, there are a number of symmetries which appear in quantum mechanics which have no classical analogue, which have no method of description in classical physics. One of these is as follows: If ψ is the amplitude for some process or other, we know that the absolute square of ψ is the probability that the process will occur. Now if someone else were to make his calculations, not with this ψ , but with a ψ' which differs merely by a change in phase (let Δ be some constant, and multiply $\exp(i\Delta)$ times the old ψ), the absolute square of ψ' , which is the probability of the event, is then equal to the absolute square of ψ :

Therefore the physical laws are unchanged if the phase of the wave function is shifted by an arbitrary constant. That is another symmetry. Physical laws must be of such a nature that a shift in the quantum-mechanical phase makes no difference. As we have just mentioned, in quantum mechanics there is a conservation law for every symmetry. The conservation law which is connected with the quantum-mechanical phase seems to be the *conservation of electrical charge*. This is altogether a very interesting business!

Mirror reflections

Now the next question, which is going to concern us, is the question of symmetry under *reflection in space*. The problem is this: Are the physical laws symmetrical under reflection? We may put it this way: Suppose we build a piece of equipment, let us say a clock, with lots of wheels and hands and numbers; it ticks, it works, and it has things wound up inside. We look at the clock in the mirror. How it *looks* in the mirror is not the question. But let us actually *build* another clock which is exactly the same as the first clock looks in the mirror—every time there is a screw with a right-hand thread in one, we use a screw with a left-hand thread in the corresponding place of the other; clocks, both physical, which bear to each other the relation of an object and its mirror image, although they are both actual, material objects, we emphasize. Now the question is: If the two clocks are started in the same condition, the springs wound to corresponding tightnesses, will the two clocks tick and go around,

We would suspect that, at least in the case of these clocks, reflection in space is one of the symmetries of physical laws, that if we change everything from “right” to “left” and leave it otherwise the same, we cannot tell the difference. Let us, then, suppose for a moment that this is true. If it is true, then it would be impossible to distinguish “right” and “left” by any physical phenomenon, just as it is, for example, impossible to define a particular absolute velocity by a physical phenomenon. So it should be impossible, by any physical phenomenon, to define absolutely what we mean by “right” as opposed to “left,” because the physical laws should be symmetrical.

So we must try to find some phenomenon in which “right hand” is involved fundamentally. The next possibility we discuss is the fact that polarized light rotates its plane of polarization as it goes through, say, sugar water.

That is a way of defining “right-hand,” because we may dissolve some sugar in the water and then the polarization goes to the right. But sugar has come from living things, and if we try to make the sugar artificially, then we discover that it *does not* rotate the plane of polarization! But if we then take that same sugar which is made artificially and which does not rotate the plane of polarization, and put bacteria in it (they eat some of the sugar) and then filter out the bacteria, we find that we still have sugar left (almost half as much as we had before), and this time it does rotate the plane of polarization, but *the other way!* It seems very confusing, but is easily explained.

Take another example: One of the substances which is common to all living creatures and that is fundamental to life is protein. Proteins consist of chains of amino acids. This amino acid is called L-alanine and has a handedness (left say). On the other hand, if we try to make alanine from carbon dioxide, ethane, and ammonia (and we *can* make it, it is not a complicated molecule), we discover that there are equal amounts of L-alanine and D-alanine (right handed). So it looks as though the phenomena of life permit a distinction between “right” and “left”.

If life is entirely a physical and chemical phenomenon, then we can understand that the proteins are all made in the same corkscrew only from the idea that at the very beginning some living molecules, by accident, got started and a few won. Somewhere, once, one organic molecule was lopsided in a certain way, and from this particular thing the “right” happened to evolve in our particular geography; a particular historical accident was one-sided, and ever since then the lopsidedness has propagated itself. Once having arrived at the state that it is in now, of course, it will always continue—all the enzymes digest the right things, manufacture the right things: when the carbon dioxide and the water vapor, and so on, go in the plant leaves, the enzymes that make the sugars make them lopsided because the enzymes are lopsided. If any new kind of virus or living thing were to originate at a later time, it would survive only if it could “eat” the kind of living matter already present. Thus it, too, must be of the same kind.

Broken symmetries

The next question is, what can we make out of laws which are *nearly* symmetrical? The marvelous thing about it all is that for such a wide range of important, strong phenomena—nuclear forces, electrical phenomena, and even weak ones like gravitation—over a tremendous range of physics, all the laws for these seem to be symmetrical. On the other hand, this little extra piece says, “No, the laws are not symmetrical!” How is it that nature can be almost symmetrical, but not perfectly symmetrical? What shall we make of this? First, do we have any other examples? The answer is, we do, in fact, have a few other examples. For instance, the nuclear part of the force between proton and proton, between neutron and neutron, and between neutron and proton, is all exactly the same—there is a symmetry for nuclear forces, a new one, that we can interchange neutron and proton—but it evidently is not a general symmetry, for the electrical repulsion between two protons at a distance does not exist for neutrons. So it is not generally true that we can *always* replace a proton with a neutron, but only to a good approximation. Why *good*? Because the nuclear forces are much stronger than the electrical forces. So this is an “almost” symmetry also. So we do have examples in other things.

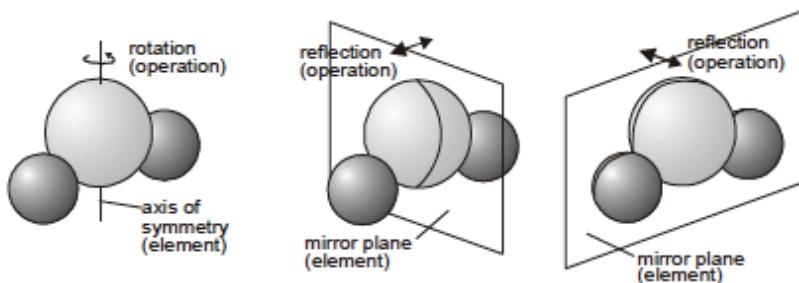
We have, in our minds, a tendency to accept symmetry as some kind of perfection. In fact it is like the old idea of the Greeks that circles were perfect, and it was rather horrible to believe that the planetary orbits were not circles, but only nearly circles. The difference between being a circle and being nearly a

circle is not a small difference, it is a fundamental change so far as the mind is concerned. There is a sign of perfection and symmetry in a circle that is not there the moment the circle is slightly off—that is the end of it—it is no longer symmetrical. Then the question is why it is only *nearly* a circle—that is a much more difficult question. The actual motion of the planets, in general, should be ellipses, but during the ages, because of tidal forces, and so on, they have been made almost symmetrical. Now the question is whether we have a similar problem here. The problem from the point of view of the circles is if they were perfect circles there would be nothing to explain, that is clearly simple. But since they are only nearly circles, there is a lot to explain, and the result turned out to be a big dynamical problem, and now our problem is to explain why they are nearly symmetrical by looking at tidal forces and so on.

So our problem is to explain where symmetry comes from. Why is nature so nearly symmetrical? No one has any idea why. The only thing we might suggest is something like this: There is a gate in Japan, a gate in Neiko, which is sometimes called by the Japanese the most beautiful gate in all Japan; it was built in a time when there was great influence from Chinese art. This gate is very elaborate, with lots of gables and beautiful carving and lots of columns and dragon heads and princes carved into the pillars, and so on. But when one looks closely he sees that in the elaborate and complex design along one of the pillars, one of the small design elements is carved upside down; otherwise the thing is completely symmetrical. If one asks why this is, the story is that it was carved upside down so that the gods will not be jealous of the perfection of man. So they purposely put an error in there, so that the gods would not be jealous and get angry with human beings.

We might like to turn the idea around and think that the true explanation of the near symmetry of nature is this: that God made the laws only nearly symmetrical so that we should not be jealous of His perfection!

Molecular examples:



Each **symmetry operation** has a corresponding **symmetry element**, which is the axis, plane, line or point with respect to which the symmetry operation is carried out. The symmetry element consists of all the points that stay in the same place when the symmetry operation is performed. In a rotation, the line of points that stay in the same place constitute a **symmetry axis**; in a reflection the points that remain unchanged make up a **plane of symmetry**.

The symmetry elements that a molecule may possess are:

1. E - the identity. The identity operation consists of doing nothing, and the corresponding symmetry element is the entire molecule. Every molecule has at least this element.
2. C_n - an n-fold axis of rotation. Rotation by $360^\circ/n$ leaves the molecule unchanged. The H_2O molecule above has a C_2 axis. Some molecules have more than one C_n axis, in which case the one with the highest value of n is called the **principal axis**. Note that by convention rotations are *countrerclockwise* about the axis.
3. σ - a plane of symmetry. Reflection in the plane leaves the molecule looking the same. In a molecule that also has an axis of symmetry, a mirror plane that includes the axis is called a **vertical mirror plane** and is labelled σ_v , while one perpendicular to the axis is called a **horizontal mirror plane** and is labelled σ_h . A vertical mirror plane that bisects the angle between two C_2 axes is called a **dihedral mirror plane**, σ_d .
4. i - a centre of symmetry. Inversion through the centre of symmetry leaves the molecule unchanged. Inversion consists of passing each point through the centre of inversion and out to the same distance on the other side of the molecule. An example of a molecule with a centre of inversion is shown below.



5. S_n - an n-fold improper rotation axis (also called a rotary-reflection axis). The rotary reflection operation consists of rotating through an angle $360^\circ/n$ about the axis, followed by reflecting in a plane perpendicular to the axis. Note that S_1 is the same as reflection and S_2 is the same as inversion. The molecule shown above has two S_2 axes.

The identity E and rotations C_n are symmetry operations that could actually be carried out on a molecule. For this reason they are called **proper symmetry operations**. Reflections, inversions and improper rotations can only be imagined (it is not actually possible to turn a molecule into its mirror image or to invert it without some fairly drastic rearrangement of chemical bonds) and as such, are termed **improper symmetry operations**.

Section 1

We shall discuss mathematical fundamentals.

Finite Groups:

Symmetries of a square: A plane symmetry of a square (or any plane figure F) is a function from the square to itself that preserves distances, i.e. the distances between the images of points P and Q equal the distance between P and Q.

The 8 symmetries of a square are : 4 rotations, flip about horizontal axis, flip about vertical axis, flips about the two diagonals; let us call these

e , the identity or zero rotation, R the clockwise rotation by 90° , similarly R^2 and R^3 , m_2 the flip about the horizontal axis, m_1 flip about the vertical axis, m_3 the flip about the northwest diagonal, m_4 the flip about the northeast diagonal.

A succession these operations is also a symmetry of the square, e.g. $HR_{90} = D.S, R_0S = S = SR_0$ for any symmetry S ,

We can make a (Cayley) multiplication table of such compositions

	e	R	R^2	R^3	m_1	m_2	m_3	m_4
e	e	R	R^2	R^3	m_1	m_2	m_3	m_4
R	R	R^2	R^3	e	m_4	m_3	m_1	m_2
R^2	R^2	R^3	e	R	m_2	m_1	m_4	m_3
R^3	R^3	e	R	R^2	m_3	m_4	m_2	m_1
m_1	m_1	m_3	m_2	m_4	e	R^2	R	R^3
m_2	m_2	m_4	m_1	m_3	R^2	e	R^3	R
m_3	m_3	m_2	m_4	m_1	R^3	R	e	R^2
m_4	m_4	m_1	m_3	m_2	R	R^3	R^2	e

Notice that $Rm_1 = m_4$ and $m_1R = m_3$; so $Rm_1 \neq m_1R$ i.e the composition operation

is said to be non-commutative (non-Abelian). The set of symmetries is closed; we call the set of symmetries D_4 .

Definition of group:

G is a group and $x \in G$ is a group element.

Product operator \cdot on G , $x \cdot y \in G$ for $x \in G$, $\forall x, y \in G$ $x \cdot y$ forms multiplication table of group.

Also $(x \cdot y) \cdot z = x \cdot (y \cdot z)$

Also $I \in G$ such that $I \cdot x = x$ and $x \cdot I = x$, $\forall x \in G$

For each $x \in G$ there exists an element denoted by x^{-1} such that $x^{-1} \cdot x = x \cdot x^{-1} = I$.

In general $x \cdot y \neq y \cdot x$ (e.g. matrix multiplication).

If $x \cdot y = y \cdot x$ for all elements of a group G , then G is called **Abelian** or **commutative**.

If the set of all elements of G is finite then the group is finite.

A subset of a group G whose elements satisfy the group axioms is called a subgroup of G .

From the Cayley table we can see that D_4 is a group, the dihedral group of order 8.

Abelian (commutative) subgroups of D_4 are:

$\{e, R, R^2, R^3\}, \{e\}, \{e, R^2\}, \{e, m_1\}$

$\{e, m_2\}, \{e, m_3\}, \{e, m_4\}$

$\{e, R^2, m_1, m_2\}, \{e, R^2, m_3, m_4\}$

We can generalise to the symmetry group of n -gons which will be denoted by D_n and is of order $2n$.

For even n , the axes of symmetry are lines joining midpoints of opposite sides or lines joining opposite vertices.

For odd n , axes of symmetry are the lines joining a vertex to the midpoint of the opposite side.

Note: a reflection followed by a different reflection is a rotation.

Fields:

A set F for which both addition (+) and multiplication (\cdot) are defined as separate operations is called a field. Informally F can be denoted by $(A, +, \cdot)$.

Addition and multiplication each define their own Abelian group structure.

For addition the identity is denoted by 0; for multiplication the identity is denoted by 1.

For addition the inverse of an element a is $-a$.

For $a \neq 0$ the multiplicative inverse is a^{-1} .

Multiplication is distributive over addition: $a \cdot (b + c) = a \cdot b + a \cdot c$

Integers do not form a field, because the operation of multiplication does not have an inverse.

Rational numbers \mathbb{Q} form a field.

The real numbers \mathbb{R} form a field.

The irrational numbers do not form a field.

The complex numbers \mathbb{C} form a field.

Sets

A set is a collection of objects. For example

$$S = \{a, b, c\}$$

is a set named S . The objects are called its elements or its members. If element a belongs to S , we write $a \in S$ and say “ a is in S ”. The simplest set is the empty set and is denoted by

$$\emptyset$$

The number of elements in a set is called its cardinality or order, and denoted $|S|$, so in the above $|S| = 3$. You are probably familiar with common sets like the set of all integers³

$$\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$$

or the set of real numbers \mathbb{R} . We won't go into the details of the rigorous construction of sets which is a thorny subject. Instead, we will play hard and loose, and endow sets with properties. For example, we can make the set of Natural numbers by

$$\mathbb{N} = \{x | x \in \mathbb{Z}, x > 0\}.$$

In words, you can read $|$ as “such that”, and what follows are termed loosely “properties” – rules which tell us what elements belong to the set. Intuitively, specifying properties like this seems like a perfectly good way to construct a set, and in this case it is. However, using properties alone to construct sets can lead to all sorts of paradoxes (the most famous of which is Russell's Paradox), so mathematicians spend much angst in developing a whole new way of constructing sets. Having said that, we will not worry too much about this and use properties with wild abandon.

While you must be familiar with sets of numbers, the elements in principle can be made out of anything like letters, points in space, cats, or even other sets. There is a big advantage in not forcing sets to contain just numbers or letters – since if we prove a theorem then it will apply to a greater number of situations. This notion of “keeping things as general as possible” is called abstraction. Numbers themselves are abstractions of things – for example, there is no need to develop a theory of counting apples and a theory of counting oranges, we can just develop a general theory of counting and use it where applicable.

Notice that in the specification of \mathbb{Z} above, we have used the dot-dot-dots, and intuitively you have assumed that the first set of dots mean $-3, -4, -5$ etc, and the second set of dots mean $3, 4, 5$ etc. Your

³ \mathbb{Z} is for Zahlen, German for “integers”.

brain has automatically assumed an ordering. Of course, we don't have to, and we can equally well specify

$$\mathbb{Z} = \{\dots, -4123, 69, 794, 0, 66, -23, \dots\}.$$

but now the dots-dots-dots are confusing. To have a notion of ordering, we would need to invent the ideas of $<$, $>$ and $=$. These ideas seems “self-evident”, but let’s see how we can cast them in set-theoretic⁴ language. So clear your head about any pre-conceived notions you have learned from high school in what follows.

Let’s begin with some definitions.

(Definition) Subsets: Suppose A and B are sets, and we say that A is a *subset* of B whenever $x \in A$ then $x \in B$. We write

$$A \subseteq B$$

Note that if $A \subseteq B$ and $B \subseteq C$ then $A \subseteq C$ (which you should try to show by drawing a **Venn Diagram**). We say that \subseteq is *transitive*.

Let’s define something seemingly completely obvious: what do we mean by two sets are “equal”.

(Definition) Equality: Suppose $A \subseteq B$ and $B \subseteq A$, then we say that A and B are *equal* and write $A = B$.

This definition has the obvious consequence that if $A = \{a, b\}$ and $B = \{b, a\}$ then $A = B$ (note that the ordering does not matter). The not so obvious consequence is that

$$\{a, b, c\} = \{a, a, a, b, c, c, a\} = \{a, b, c, c\}$$

and so on. So even if two sets do not “look” the same, they can still be equal. If A and B are not equal, we write $A \neq B$.

Now we are ready to, literally, make numbers out of nothing. What follows is an arcane way of making up sets of numbers which was invented by the great physicist/information theorist/mathematician John Von Neumann.

Natural Numbers (Neumann) \mathbb{N} : Start with nothing, i.e. the empty set \emptyset , and then we can put this empty set into a set, i.e. we construct $\{\emptyset\}$. We can put *that* into another set $\{\emptyset, \{\emptyset\}\}$, and iterate this to make $\{\emptyset, \{\emptyset, \{\emptyset\}\}\}$ etc. We then *name* these sets

$$0 = \emptyset, 1 = \{\emptyset\}, 2 = \{\emptyset, \{\emptyset\}\}, 3 = \{\emptyset, \{\emptyset, \{\emptyset\}\}\}$$

and so on. Notice that we have abstracted numbers themselves as sets!

In this construction, notice that since \emptyset is in $\{\emptyset\}$ this implies that $\emptyset \subseteq \{\emptyset\}$, and also $\{\emptyset\} \subseteq \{\emptyset, \{\emptyset\}\}$ etc. Or in everyday language $0 \leq 1 \leq 2 \dots$, or you can also write $0 < 1 < 2 \dots$ by replacing \subseteq with $<$ (which keeps the proposition still true), and the transitivity property of \subseteq gets imported in to become the ordering. So, like magic, we have pulled rabbits out of thin air and constructed numbers with a natural ordering out of some set-theoretic **axioms** (or definitions). This set of numbers is called **Natural Numbers** and given the mathematical symbol \mathbb{N} .

Now is a good time point out a pedantic (but important) difference between $A \subseteq B$ and $A \subset B$ – which one deserves to be the mathematical symbol for “subset”? The way out of this is to invent a new term, **proper**. For $A \subset B$, we say that A is a **proper subset** of B , i.e. A is a subset of B but that $A \neq B$. So every time when you see the word “proper” attached to any definition, think of the difference between \subseteq and \subset . You will see a lot of these in our lectures.

⁴You may hear some high-brow people often say “set-theoretic”, “field-theoretic” etc, this simply means that we want to discuss something in the language of sets and fields. One of the things that you should start learning early is to look for abstractions – the same set of physical systems can often be recast in different forms. One unstated goal of this course is to encourage you to think of “obvious” ideas in not so obvious ways. Ok, now we have stated it.

A word on jargon: things like $<$, \geq , \subseteq – which allow us to determine the *relationships* between two different sets are called **relations** – we will study this later. This is to distinguish another thing which actually “does stuff” to sets, as we will study next.

2.1.1 Operations on Sets

We can use properties to make new sets, by “doing stuff” to them. Mathematically, “doing stuff” is called “operating” or “acting”. We can invent operations which act on sets.

(Definition) Union: Suppose A and B are sets, the union of two sets is defined to be

$$A \cup B = \{x | x \in A \text{ or } x \in B\}.$$

Notice the English word "or" in the property. So if you have done some logic, you can think of \cup as an "or" operator.

(Definition) Intersection: Suppose A and B are sets, the intersection of two sets is defined to be

$$A \cap B = \{x | x \in A, x \in B\}.$$

You can replace the comma "," with "and", so \cap is an "and" operator.

If $A \cap B = \emptyset$ then we say that A and B are disjoint or not connected. Furthermore, the operations are commutative

$$A \cup B = B \cup A, A \cap B = B \cap A.$$

This looks like the commutativity of the algebra you have seen before $x + y = y + x$ in high school⁵. Can you prove this?

Also, in high school you learned to rewrite commutativity as a "minus" sign -

$$(x + y) - (y + x) = 0.$$

But remember we are doing sets, and we have not invented the notion of the operation of "subtraction". Let's invent it now using properties

(Definition) Quotient: Suppose A and B are sets, then the quotient of two sets is

$$A \setminus B = \{x | x \in A, x \notin B\}.$$

Note that \setminus is not the divide sign $/$! We say $A \setminus B$ "A quotient B". \notin means "not in", i.e. it is the negation of \in .

We can say the property in words: " $A \setminus B$ " is a set which contains elements x which are in A and not in B , i.e. it is a subtraction, so you can also call it $A - B$. To see this, consider two sets $A = \{a, b, c\}$ and $B = \{a\}$. We start with the elements in A , and keep only those that are not in B , i.e. b and c . So $A - B = \{b, c\}$. Question: what if $A = \{a, b, c\}$ and $B = \{a, d\}$?

(Definition) Complement: Suppose U is the set of everything in the Universe and A is a set, then the complement $\bar{A} = U \setminus A = U - A$ is the set of everything not in A .

You are also familiar with the associativity law $(x+y)+z = x+(y+z)$. The set theoretic associative laws are

$$(A \cup B) \cup C = A \cup (B \cup C), (A \cap B) \cap C = A \cap (B \cap C).$$

And the distributive law $x \cdot (y+z) = x \cdot y + x \cdot z$ (where \cdot means multiplication in the usual sense)

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C), A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

Power Set: A more complicated operation is to make sets out of a set is to do the following "Given a set A , form all possible subsets of A , and put these subsets into a new set call the Power set $\mathcal{P}(A)$ ". In mathematical language, this is simply

$$\mathcal{P}(A) = \{x | x \subseteq A\}. \quad \text{So if } A = \{a, b\}, \text{ then } \mathcal{P}(A) = \{\emptyset, \{a\}, \{b\}, \{a, b\}\}.$$

Cartesian Product and Coordinate Systems

Finally, intuitively, we know that pairs of objects exist – say for example a coordinate system (x, y) or $(\text{Justin}, \text{Selena})$ etc. We can form these objects by the operation call "taking the Cartesian product".

(Definition) Cartesian Product: Suppose A and B are sets, then the Cartesian product of A and B is given by

$$A \times B = \{(a, b) | a \in A, b \in B\}$$

and we call (a, b) an ordered pair. If we have two ordered pairs (a, b) and (c, d) then we say that $(a, b) = (c, d)$ when $a = c$ and $b = d$ (this may look pedantic, but we have secretly defined a new notion of equality, one which is intuitive but still new.)

So if you want to make a 2-dimensional continuous coordinate system (x, y) , then you can define

$$\{(x, y) | x \in \mathbb{R}, y \in \mathbb{R}\}.$$

Maps and Functions

Given two sets A and B , we can define a link between them. Mathematically, we say that we want to find a **mapping** between two sets. The thing we use to do this is called a (doh) **map** or a **function**. If you have not thought about functions as maps, now is a time to take a private quiet moment to yourself and think about it. Again there is a very rigorous way of defining maps in mathematics, but we will simply think of them as a set of *rules*. A rule is something that takes in an element of one set, and give you back an element of another set. Let's define it properly:

(Definition) **Map**: Suppose A and B are sets, then a map f defines a link between A and B as follows

$$f : A \rightarrow B; (\text{Rules}).$$

Example: Suppose $A = \mathbb{R}$ and $B = \mathbb{R}$, so

$$f : A \rightarrow B; f : x \mapsto x^2 \forall x \in \mathbb{R}$$

where we have distinguished the arrows \rightarrow to mean “maps to” while \mapsto means “the rule is as follows” (or its **recipe**). Having said that, we will be careless about such arrows from now on and use \rightarrow . The above described, in language you are familiar with, $f(x) = x^2$. (If you have not seen \forall before, it is called “for all”.) So sometimes we write

$$f : A \rightarrow B; f(x) = x^2 \forall x \in \mathbb{R}.$$

Here A and B are morally speaking, in general, *different* sets, even though they are equal to \mathbb{R} – one can imagine them to be two different worlds that look like \mathbb{R} . Of course one can also imagine them to be the special case where they are the same world. Notice however, for any $x \in A$, since $x^2 > 0$, $f(x)$ maps only to the positive definite part of B , plus zero. In addition, both $x = -4$ and $x = 4$ maps to the same point $f(4) = f(-4) = 16$ etc.

The set of inputs A is called the **domain** of the map, while the set where the outputs live B is called the **codomain**. We call them $\text{dom}(f)$ and $\text{cod}(f)$ respectively. We can define also the **image** of f in the following way

(Definition) **Image**: Suppose $f : A \rightarrow B$ is a map, then the image of f is

$$\text{im}(f) = \{x | x \in B, x = f(a) \text{ for some } a \in A\},$$

or even shorter form

$$\text{im}(f) = \{f(a) | a \in A\}.$$

In words, the image of f is the part of the codomain which is the domain is mapped to. So if you pick a point in the codomain which is *not* in $\text{im}(f)$ then you are out of luck as you have no partner in the domain.

One can also construct a very special map, which maps every element of a set into the same element. This is called the **Identity Map** and is defined as follows.

Identity Map: Given any set A ,

$$\text{Id} : A \rightarrow A; \text{Id}(a) = a \forall a \in A.$$

It's clear that an identity map always exists for all sets.

Note that A and B do not have to be “different” sets, they could be (and often is) the “same” set (although mathematically it doesn't matter if it is the “same” or not – one can always say B is a copy of A for example, A won't mind).

Surjective, Injective and Bijective

In the map $f(x) = x^2$ we described above, there exists points in $\text{cod}(f)$ that has no “partner” in $\text{dom}(f)$, i.e. colloquially, no points in A is mapped to negative values. A map which *does* map to the entire codomain, i.e.

$$\text{im}(f) = \text{cod}(f)$$

is called a **surjective** (Latin) or **onto** (English) or **epic** (Hellenic) map.

Also, in the $f(x) = x^2$ map, two points in the domain is mapped to a single point in the co-domain. On the other hand, one can imagine a map which *gives different outputs for different inputs*, i.e. consider a map g

$$g : A \rightarrow B; \quad a, a' \in A \text{ and if } a \neq a' \text{ then } f(a) \neq f(a').$$

Such a map is called an **injective** (Latin) or **one-to-one** (English) or **monic** (Hellenic) map.

Finally if a map is both onto and one-to-one, then it is **bijective**.

Equality of maps: Given these, we say that two maps f and g are equal and write $f = g$ iff (if and only if)

$$\begin{aligned} \text{dom}(f) &= \text{dom}(g), \\ \text{cod}(f) &= \text{cod}(g), \end{aligned}$$

$$\forall x \in \text{dom}(f), \quad f(x) = g(x).$$

The first two conditions tells us that the domains and codomains must be the same, while the third condition ensures that both f and g maps to the same points in the (same) codomain. Note that specifying the 3rd condition alone is not enough! For example, the maps $f(x) = x$ and $g(x) = x$ might look the same but if $\text{dom}(f) = \mathbb{Z}$ while $\text{dom}(g) = \mathbb{R}$, $f \neq g$.

(Definition) Countability: A set S is said to be *countable* if there exists a bijective map $f : S \rightarrow \mathbb{N}$, i.e. a bijective map from S to the set of Natural Number \mathbb{N} (which we constructed early in Chapter 2).

Hence a countable set is something you can “count”, by ticking off the elements of \mathbb{N} i.e. $0, 1, 2, 3, \dots$ and you will be sure that you have not missed any numbers in between⁹. After all, we are taught to count in terms of the Natural numbers!

Continuity vs Infinity: We make a remark here that although S can be infinite, it is not *continuous* so don't mix up the two concepts! Briefly, even though S can be countably infinite, this doesn't mean that there exist a well defined notion that two elements in S is “arbitrarily close together”. Indeed, to talk about “close together” require a notion of “distance between points” – your teacher might not have told you but Calculus is secretly the study of a mathematical structure which possess such a notion of *infinitesimal distances*. In fact, since \mathbb{N} does not possess such a structure, it does not make sense to talk about continuity.

Composition of maps

One way to think about a map is that it is like a vending machine or a blackbox : you feed it something, and something else comes out. You can then take this something else, and feed it into some other blackbox, and some other thing will come out. This series of events is called a **composition**; we have taken two “leaps” across two different codomains.

Suppose we have two maps $f : A \rightarrow B$ and $g : B \rightarrow C$, then the composition map $g \circ f$ is described by

$$g \circ f : A \rightarrow C; \quad g(f(x)) \quad \forall x \in A.$$

So you take x from A , give it to f who will spit out $f(x)$, and then you feed this output to g who will give you back $g(f(x))$.

You can string maps together. For example, given a third map $h : C \rightarrow D$ we can construct

$$h \circ g \circ f : A \rightarrow D; \quad h(g(f(x))) \quad \forall x \in A.$$

We will now state three theorems.

Theorem: *The composition of two injective maps is injective.*

Proof: Suppose $f : A \rightarrow B$ and $g : B \rightarrow C$ are injective maps. Suppose further $a, a' \in A$ and $a \neq a'$. f is injective so $f(a) \neq f(a')$. But since g is also injective this means that $g(f(a)) \neq g(f(a'))$. \square .

(This may be “obvious” to you, but it is good to practice thinking like a mathematician once in a while. Anyhow the little square \square is to denote “as to be demonstrated”, and we like to slap it on the end of any proofs as it is very satisfying. Try it.)

Theorem: *The composition of two surjective maps is surjective.*

Proof: Suppose $f : A \rightarrow B$ and $g : B \rightarrow C$ are surjective maps. We will now work backwards from C . Choose any $c \in C$. g is surjective, so $\exists b \in B$ such that $g(b) = c$. Now f is also surjective, so $\exists a \in A$ such that $f(a) = b$, hence $g(f(a)) = c$, i.e. there exist a map from a to every element in C hence $g \circ f$ is surjective. \square .

(We have sneaked in the symbol \exists , which means “there exists”.)

Corollary: *The composition of two bijective maps is bijective.*

The last theorem is a corollary – i.e. it is an “obvious” result so we won’t prove it. Obviousness is subjective, and corollaries are nice dodges for people who are too lazy to write out proofs (like me).

Inverse: Consider a map $f : X \rightarrow Y$. Now, we want to ask the question: can we construct a map g from Y back to X such that $g(f(x))$ is the unique element x . If f is onto, but not one-to-one, then $\exists y \in Y$ which was originally mapped from more than 1 element in X . On the other hand, if f is one-to-one, but not onto, then $\exists y \in Y$ which was not mapped from any point $x \in X$.

It turns out the crucial property of f that allows the construction of such an inverse map g is that f is bijective. In fact it is more than that: the statement that f is a bijective is equivalent to the statement that there exists an inverse such that $g(f(x)) = \text{Id}_X$ (where we have appended the subscript X to indicate that it is the identity map in X). Formally,

Let $f : X \rightarrow Y$ be a map. Then the following two statements are equivalent:

f is bijective \Leftrightarrow There exists an inverse map $g : Y \rightarrow X$ such that $f \circ g = \text{Id}_Y$ and $g \circ f = \text{Id}_X$.

Notice that g is also the inverse of f , i.e. inverses “go both ways”. One thing we have been a bit careful here to state is the notion of equivalence, which we have used the symbol \Leftrightarrow . Roughly speaking, equivalence means that both statements has the same content, so stating one is sufficient. Compare equivalence with implication with the symbol \Rightarrow , e.g. f is bijective \Rightarrow f is onto, whose reverse is of course not necessarily true.

While there are plenty of mathematical constructs which do not require inverses, inverses are crucial in most mathematics which describe the physical world. In fact, inverses and identities often go hand-in-hand, like ying and yang.

Sets of Maps/Functions

By now we are comfortable with the idea of sets as containers of objects, and maps as devices which allow us to build links between the objects inside sets. However, *maps are objects too*, so we can build sets out of maps. This leap of abstraction will underlie much of our future discussion on groups.

Suppose A and B are sets, then we can define a *set* of maps formally by the following

$$\text{map}_{A,B} = \{f | f : A \rightarrow B\}$$

so $\text{map}_{A,B}$ is a set of maps from A to B . This set is very general – any map from A to B is in it. There exist some interesting subsets, say the subset of all bijective maps

$$\text{bij}_{A,B} = \{g | g \in \text{map}_{A,B}, g \text{ is bijective}\}.$$

We can also construct the set of all onto maps $\text{sur}_{A,B}$ and all one-to-one maps $\text{inj}_{A,B}$. Can you now see why $\text{bij}_{A,B} = \text{sur}_{A,B} \cap \text{inj}_{A,B}$?

Relations, Equivalence Relationships

We will now discuss a concept that is super important, but for some completely strange reason is not widely taught or even understood. This concept is the notion of an “equivalence between objects” (as opposed to equivalence between statements) – in vulgar language we want to define “what is equal”. Let’s start with something simple that you have grasped a long time ago, and work our way up the abstraction ladder.

When we are given a set of objects, they are nothing but a bunch of objects, for example, you are given a set of students in the Symmetry in Physics class, or a set of Natural numbers \mathbb{N} . To give “structure” to the sets so we can do more interesting things with them, we can *define* relations between the objects. Let’s be pedantic for the moment.

(Definition) Relation: Suppose S is a set, then we say $p, q \in S$ are **related** by relation \bowtie and we write $p \bowtie q$. We call \bowtie a *relation on S* .

We have not *specified* what \bowtie means of course, but you have already learned some relations. For $S = \mathbb{N}$ then you already have learned relations $=, <, >$ on \mathbb{N} . So, obviously, $1 = 1$, $2 < 5$, $9 > 5$ etc. However, who says we are stuck with those relations? How do you define relations between members of the set of Symmetry in Physics class?

Surprisingly, it turns out that relations themselves can be divided into *only* a few major properties! Here are three of them which we will be concerned about; suppose \bowtie is a relation on set S , then it is

- **Reflexive:** if $a \bowtie a$ for every $a \in S$
- **Symmetric:** if $a \bowtie b$ then $b \bowtie a$ for $a, b \in S$
- **Transitive:** if $a \bowtie b$, and $b \bowtie c$ then $a \bowtie c$ for $a, b, c \in S$.

Whatever relation we specify, they may have *none*, *some* or *all* of above properties. Consider the relation $=$ on \mathbb{N} – it is reflexive, symmetric and transitive since $1 = 1$, $2 = 2$ etc (can you see why it is transitive?). On the other hand, $<$ on \mathbb{N} is not reflexive nor symmetric, but it is clearly transitive since if $1 < 3$ and $3 < 6$ then $1 < 6$ is true.

Relations allow us to put extra structure on sets. Now consider again our set of Symmetry in Physics students, let’s call it P . Let’s say we want to define an **equivalence relationship**, let’s call it \sim , based on the gender of the students. For example, we want to be able to say things like “Janice is equivalent to Ellie” because both Janice and Ellie are females so $\text{Janice} \sim \text{Ellie}$. Naturally, $\text{Ellie} \sim \text{Janice}$ (symmetric), and furthermore $\text{Anna} \sim \text{Ellie}$ and hence $\text{Janice} \sim \text{Anna}$ (transitive).

(Definition) Equivalence Relation: A relation \bowtie on set S is an *equivalence relation* if it is reflexive, symmetric and transitive. We usually use the symbol \sim for equivalence.

Why go through all this bother? Broadly speaking, equivalence relationships give us a notion of “sameness” within some (to be specified) criteria. A set can, of course, have more than one equivalence relationship (think “long hair” and “short hair” members in the Symmetry of Physics students set).

In addition, it turns out that equivalence relations can **partition** a set into *disjoint* subsets of objects which are equivalent to each other (it is obvious if you spend a minute thinking about it), which we can agree is a useful concept. You have seen “disjoint” before, but what do we mean by *partition*? For example, the “all members of one gender relation” partitions the class into “boys” subset M and “girls” subset F , and it’s clear that $M \cap F = \emptyset$. Let’s now define partitions.

(Definition) Partition: A **partition** of S is a collection C of subsets of S such that (a) $X \neq \emptyset$ whenever $X \in C$, (b) if $X, Y \in C$ and $X \neq Y$ then $X \cap Y = \emptyset$, and (c) the union of *all* of the elements of the partition is S .

The subsets that are partitioned by equivalence relations *must* be disjoint – again think about boys and girls. These disjoint subsets are called **equivalent classes**. Equivalence classes will play an extremely important role in group theory and the study of symmetries in physics.

Algebras

We have studied sets, relations on the sets which tell us how the objects in a set are related to each other, and maps between sets. The final piece of mathematical structure we need to study is its **algebra**.

What is an algebra? In your school days, “algebra” refers to the study of things like $a + a = 2a$, $a \cdot (b + c) = a \cdot b + c \cdot b$ etc. You were taught how to manipulate the addition **operators** $+$ and also the multiplication operators \cdot . However, more abstractly, an **algebra** is a set of elements equipped with rules that tell us how to combine two elements in the set to produce another element of the set. Such rules are called **binary operators**.

Example: Suppose you are given a set $\mathbb{Z}_2 = \{s_1, s_2\}$. We want to invent a set of rule for “multiplication” of two objects in the set, which we write by

or $s_1 * s_2$ or $s_1 \cdot s_2$ or simply $s_1 s_2$.

We say that we want a **product** or **composition** rule⁷.

To invent a rule, we just specify the result of such a composition. There are only two elements, so we need 4 rules for all possible permutations. A possible (by no means unique) set of rules is

$$s_1 s_1 = s_1, \quad s_1 s_2 = s_2, \quad s_2 s_1 = s_2, \quad s_2 s_2 = s_1.$$

Notice that we have specified $s_1 s_2 = s_2$ and $s_2 s_1 = s_2$ separately. In this very simple example, the composition gives the same element s_2 but *in general* they don't have to be the same and in fact often are not the same, i.e. they are **non-commutative**.

In words, we say that “set S with rules Eq. (65) form an algebra.”

Of course, if we have a set with infinite order, for example the set of all real numbers \mathbb{R} , then it will be tedious (or crazy) to specify all the rules for each possible pair of elements. (This is the kind of thing they make you do when you die and go to mathematician hell.) More efficiently, if we invent a rule, we would like it to apply to *all* pairs of elements in the set. This is exactly what we will do, and it turns out that we can classify the type of rules by specifying certain **algebraic axioms**. There are several well known algebras, and we will discuss two of them in the following.

Vector spaces

Spaces of vectors form vector spaces e.g. \mathbb{R}^n n -dimensional vectors with real components. Formally

$$\mathbb{R}^n = \left\{ \mathbf{x} = \begin{pmatrix} x^1 \\ x^2 \\ \vdots \\ x^n \end{pmatrix}, x^i \in \mathbb{R} \right\}$$

Two vectors \mathbf{x} and \mathbf{y} can be concatenated to form a new vector $\mathbf{z} = \mathbf{x} + \mathbf{y}$.

More formally a \mathbb{F} -vector space is a triple $(V, +, \cdot)$ consisting of a set V , a vector addition rule,

$$+ : V \times V \rightarrow V, \quad (\mathbf{x}, \mathbf{y}) \mapsto \mathbf{x} + \mathbf{y}$$

and a rule of multiplication by scalars:

$$F \times V \rightarrow V, \quad (a, \mathbf{v}) \rightarrow a \cdot \mathbf{v} = a\mathbf{v}$$

satisfying the vector space axioms

- the addition of vectors defines an Abelian group
- the neutral element $\mathbf{0}$ is called the null vector
- the inverse element is called the negative vector $-\mathbf{v}$
- scalar multiplication satisfies the following rules
 1. $(a + b)\mathbf{v} = a\mathbf{v} + b\mathbf{v}$
 2. $a(\mathbf{v} + \mathbf{w}) = a\mathbf{v} + a\mathbf{w}$
 3. $(ab)\mathbf{v} = a(b\mathbf{v})$
 4. $1\mathbf{v} = \mathbf{v}$

V may contain a set $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ of n vectors which is complete if

$$\text{span}(S) = V$$

S is linearly independent if any vector \mathbf{v} can be decomposed uniquely using linear combinations in S .

Scalar product in \langle , \rangle in \mathbb{R} -vector space:

$$\langle , \rangle : V \times V \rightarrow \mathbb{R}$$

Properties of \langle , \rangle :

- symmetry $\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle$
- linearity $\langle \alpha\mathbf{v}, \mathbf{w} \rangle = \alpha \langle \mathbf{v}, \mathbf{w} \rangle$ and $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$
- positive definiteness $\langle \mathbf{v}, \mathbf{v} \rangle > 0$ for all $\mathbf{v} \neq \mathbf{0}$

The norm of a vector is defined by $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$.

Complex inner product on \mathbb{C} -vector space

$$\langle , \rangle : V \times V \rightarrow \mathbb{C}$$

Symmetry definition changes to $\langle \mathbf{v}, \mathbf{w} \rangle = \overline{\langle \mathbf{w}, \mathbf{v} \rangle}$.

Linear maps on vector spaces

A map $\phi : V \rightarrow W$ between two vector spaces is called a linear map if

$$\phi(ax + by) = a\phi(x) + b\phi(y).$$

Matrices are linear maps with $V = \mathbb{C}^m$ and $W = \mathbb{C}^n$. If $x = (x^1, x^2, \dots, x^m)$ then

$$(\phi x)^i = \sum \phi_j^i x^j$$

and ϕ_j^i is the $n \times m$ matrix.

Finite Groups

A Few Easy Finite Order Discrete Groups

(Definition) **Order (of Group)**: The *order* of a group G is the total number of elements in the group, and is denoted $|G|$.

The order of a group can range from zero to infinity. **Finite order groups** are groups which possess a finite number of elements $|G| < \infty$.

Order 2 Group Z_2 $P\psi(x) = \psi(-x)$.

Theorem (Uniqueness of order 2 group): *The only possible order two group is Z_2 .*

Proof. Let $G = \{a, b\}$ be an order 2 group. A group possess a binary operator, so we want to find the result for the compositions aa, ab, ba, bb . By the identity axiom, we must have an identity. Let $a = e$ be the identity, then the first three compositions yield e, b, b respectively. The last composition bb must be e or b by the closure axiom. Suppose now that $bb = b$ then $bb = b = be = b(bb^{-1}) = b^2b^{-1}$, the last equality be associativity. But using our supposition, $b^2b^{-1} = bb^{-1} = e$ which means that $b = e$. From our proof previously that the identity has to be unique, this is a contradiction, so the only other possibility is $bb = e$. Since simply by using the group axioms, we have completely determined all the composition laws and recover Z_2 , this means that it is the *unique* order 2 group \square .

Parity: Despite its completely innocuous nature, Z_2 is in fact one of the most important groups in physics! Suppose we have a function $\psi(x)$ where x is the space coordinate with domain $-\infty < x < \infty$ (remember our map language). Now consider the **Parity Operator** P whose action is to flip the sign of the argument of $\psi(x)$, i.e.

We can also consider the “do nothing operator”, let’s call it e ,
 $e\psi(x) = \psi(x)$.

Acting on $\psi(x)$ twice with P we get

$$P(P\psi(x)) = P\psi(-x) = \psi(x)$$

or $P^2 = e$ since the operation is clearly associative. Furthermore, this means that P is its own inverse. And it’s clear that $Pe = eP = P$. Hence the set of two operators $\{P, e\}$ form a group, and by our theorem above it is Z_2 as you can check that the composition laws are the correct one.

Note that we have not said anything about the *symmetry* of $\psi(x)$ – i.e. the symmetry of the object being operated on by the group operators is irrelevant to the group operators. On the other hand, if $\psi(x) = \psi(-x)$, i.e. it is symmetric under reflection around $x = 0$, then the *value* of $P\psi(x) = \psi(-x) = \psi(x)$ and we say that $\psi(x)$ is *invariant* under parity around $x = 0$.

But as we discussed in Chapter 1, once we have the group operators, we do not need the underlying vector space (here it is $\psi(x)$ – it might be hard for you think of a function as a vector space for the moment, technically it is an infinite dimensional vector space) to possess any symmetry. Let’s consider such a case explicitly.

Consider the ubiquitous bit you know and love from computing, which can have two possible states 1 and 0, or \uparrow and \downarrow . The action of a “NOT” operator P operator flips the \uparrow to a \downarrow and vice versa i.e.

$$P \uparrow = \downarrow, \quad P \downarrow = \uparrow$$

and of course the do nothing operator exists and it leaves the bit as it is

$$e \downarrow = \downarrow, \quad e \uparrow = \uparrow.$$

You can check that flipping the bit twice with $PP = e$ gets you back the original bit, and hence P is its own inverse, and associativity is easily checked. So again the set of operators $\{P, e\}$ forms a group and it is Z_2 .

We can represent the two possible states of the bit by a 2 by 1 column matrix

$$\uparrow = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \downarrow = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

then the operators P and e can be represented by 2×2 column matrices

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad e = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

form a Linear Group Representation of Z_2 . We threw in “linear” because it is clear that

$$P(\uparrow + \downarrow) = P\uparrow + P\downarrow \text{ etc. Notice that the matrices } \uparrow \text{ and } \downarrow \text{ form a Vector Space.}$$

Order 3 Group

Consider the order 3 group $Z_3 = \{e, a, b\}$. We now want to show that, like the Z_2 group, the group axioms alone is sufficient to determine all the composition rules and hence the order 3 group is unique. We will now prove this statement and derive the algebra.

(Theorem) Uniqueness of Order 3 Group Z_3 : Z_3 is the unique order 3 group.

Proof: From the identity axiom, $ae = ea = a$, $be = eb = b$. Now we want to derive the compositions ab, ba, bb, aa .

Consider ab , closure means that it can be a, b or e . Suppose $ab = b$, then $(ab)b^{-1} = bb^{-1}$ where we have used the fact that inverses for a, b exist (which we denote as a^{-1}, b^{-1} for the moment and we'll find out what they are later). Then by associativity $abb^{-1} = e$ and this implies $a = e$ which cannot be right since the identity is unique, so $ab \neq b$. The assumption of $ab = a$ leads to $b = e$ so $ab \neq a$, and hence we have found $ab = e$.

Replacing a with b and vice versa in the last paragraph, we obtain $ba = e$. (I.e. the argument is symmetric under the interchange of a and b). This means that a and b are inverses of each other.

Let's now turn to a^2 and b^2 . Consider a^2 , closure implies that it has to be a, b, e . Suppose that $a^2 = e$, then $a^2b = eb = b$, and by associativity $a(ab) = b$ or using our previous results $ab = e$ we get $a = b$. But we need a and b to be distinct, and hence $a^2 \neq e$. Suppose $a^2 = a$, but then $a^2a^{-1} = aa^{-1} = e$, and we get $a = e$ which by uniqueness of identity cannot be true. Hence we find that $a^2 = b$. Using this, we can calculate $b^2 = a^4 = a(a^3) = a(a(a^2)) = a(ab) = a(e) = a$, i.e. $b^2 = a$. Note that for the last composition, you can also invoke the symmetry of the argument under the interchange of a and b .

We have thus found the all the composition laws for the group. The multiplication table looks like

	a	b
a	b	e
b	e	a

Since we have derived the group algebra from the axioms alone, this group must be unique \square .

Cyclic Group Z_n , Isomorphism

Now you must be wondering about the names we have given to the unique order 2 and 3 groups Z_2 and Z_3 . It turns out that they are members of a large class of groups called the **Cyclic Group**. Let's start with some definitions.

(Definition) Generating Sets and Generators: Suppose G is a group. A generator $g \in G$ is an element of G where by repeated application of the group composition with itself or other generators of G , makes (or generates) other elements of G . The minimal number of elements in G which are required to generate all the elements of G is a subset $B \subseteq G$. We call B the generating subset of G .

Generating sets are not unique in general.

Example: Recall the dihedral-4 group D_4 in Chapter 1. We have shown that the generating set $\{R, m_1\}$ generates all the other elements of the group. The group written in terms of generators would be

(Definition) Cyclic Groups Z_n : The Cyclic group Z_n is an order n group generated by a single generator a

$$Z_n = \{e, a, a^2, a^3, \dots, a^{n-1}\}, \quad a^n = e.$$

Since the group is generated by a single elements, the composition law between two elements a^m and a^p for $m, p \in \mathbb{N}$ is simply $a^m a^p = a^k$ where $k = p + m$ modulo n – the group “cycles” back to e . Hence the inverse for any element a^k is then a^{n-k} since $a^k a^{n-k} = e$.

Technically when $n = \infty$, Z_n is an infinite order group.

Z_n is trivially Abelian. Let's define Abelian properly:

(Definition) Abelian and non-Abelian: Suppose G is a group then for any two elements $g, f \in G$, the group is Abelian if

$$gf = fg \quad \text{otherwise it is non-Abelian.}$$

Previously when we discuss the Z_2 group, we say that Z_2 is the unique order 2 group. We then show that the parity operators P and the identity form an order 2 group, and the NOT operator and the identity also form an order two group. Furthermore, in your homework set, you will show that the group $G = \{-1, 1\}$ is a group under the usual rule of multiplication of numbers. All these order two groups have the same composition laws of course but they are in a sense “different” from each other because their elements looks different. So what do we mean by “unique”? Intuitively, although the elements are the different, they are “the same” because their composition laws have the same structure. Let’s formalize this notion.

(Definition) Isomorphism: Suppose we have two groups G_1 and G_2 with composition laws denoted by \cdot and $*$. They are said to be *isomorphic*, and written $G_1 \cong G_2$, if there exist a bijective map $i : G_1 \rightarrow G_2$ such that it preserves the group composition law in the following way

$$i(g_1 \cdot g'_1) = i(g_1) * i(g'_1), \quad \forall g_1, g'_1 \in G_1.$$

Since i is a bijection, the following relation follows

$$i^{-1}(g_2 * g'_2) = i^{-1}(g_2) \cdot i^{-1}(g'_2), \quad \forall g_2, g'_2 \in G_2.$$

In this language, we say that all order 2 groups are *isomorphic* to Z_2 .

If in the definition of isomorphism we relax the requirement that the map i is bijective, the map is called a homomorphism. The subset of G_1 mapped by i into the identity of G_2 is a subgroup of G_1 known as the kernel.

The idea is that the bijection i maps the elements from G_1 to G_2 , but the group composition law in G_2 has the same shape as that of G_1 , and hence it “doesn’t matter” which set of elements you choose to work with. This is clearer with an example.

Example: Let $N_n = \{0, 1, 2, 3, \dots, n - 1\}$ be a set of integers. We equip it with a composition law “addition modulo n ”, and then one can show that this forms a group with the identity 0 (homework). This group is isomorphic to Z_n in the following way which we can prove by finding the bijective map i

$$i : N_n \rightarrow Z_n ; \quad i(k) = a^k. \tag{83}$$

We can check that

- The identity in N_n is mapped to the identity in Z_n : $i(0) = a^0 = e$.
- For $p, k \in N_n$, the group composition law is $p + k \bmod n$. The map $i(p+k) = i(p)i(k) = a^p a^k = a^{p+k \bmod n}$.

hence $N_n \cong Z_n$.

Before we finish with the Cyclic group, let's talk about (yet another) definition of "order".

(Definition) Order of an element: Suppose G is a group and $g \in G$. Then the *smallest* value of n such that $g^n = e$ is called its *order*. The identity element has order 1.

Example: Consider the group $Z_2 = \{e, a\}$. e is trivially of order 1. While $a^2 = e$, so a is of order 2.

Example: Consider the dihedral group D_4 . $R^4 = e$, so R is of order 4. (What is the order of R^2 ?).

Unfortunately, the word "order" is overused in mathematics, like the word "finger", can mean very different things to different people. So you have to be careful when you use it, lest you upset people with a careless application. For example, we say that the order of the group Z_6 is 6, while the order of the element a^3 in Z_6 is 2.

■ EXAMPLE 1 Let G be the real numbers under addition and let \overline{G} be the positive real numbers under multiplication. Then G and \overline{G} are isomorphic under the mapping $\phi(x) = 2^x$. Certainly, ϕ is a function from G to \overline{G} . To prove that it is one-to-one, suppose that $2^x = 2^y$. Then $\log_2 2^x = \log_2 2^y$, and therefore $x = y$. For "onto," we must find for any positive real number y some real number x such that $\phi(x) = y$; that is, $2^x = y$. Well, solving for x gives $\log_2 y$. Finally,

$$\phi(x + y) = 2^{x+y} = 2^x \cdot 2^y = \phi(x)\phi(y)$$

for all x and y in G , so that ϕ is operation-preserving as well. ■

Klein four-group, Product Groups

We have shown that Z_2 and Z_3 are unique in the sense that every order 2 or 3 groups must be isomorphic to them. What about order 4 groups? It is clear that Z_4 is an order 4 group, but (fortunately) the cardinality of the set of elements have reached sufficient size that the group axioms no longer fully determine its structure, so Z_4 is not the unique order 4 group.

As you will be asked to prove in the Homework, there is only *one* other order 4 group other than Z_4 – it is the **Klein four-group** or the *Vierergruppe*¹¹.

Klein four-group V_4 : The Klein four $V_4 = \{e, a, b, c\}$ has the following group composition laws

$$a^2 = b^2 = c^2 = e, \quad ab = ba = c, \quad bc = cb = a, \quad ca = ac = b$$

i.e. it is an Abelian group. It is the also the lowest possible order non-cyclic group. The generators of this group is $\{a, b\}$ (can you see why?).

One thing you might notice is that 4 is also the lowest natural number which is *not a prime number* i.e. $4 = 2 \times 2$. This gives us a way to construct the four-group in a slick way. First we extend the notion of the **Cartesian Product** we studied in Chapter 2 to groups.

(Definition) Products of Groups: Let G_1 and G_2 be groups with composition laws \cdot and $*$. The *product* (or sometimes *Cartesian product*) of G_1 and G_2 is defined to be the set

$$G_1 \times G_2 = \text{The set of all possible pairs } (g_1, g_2) \text{ where } g_1 \in G_1, g_2 \in G_2$$

with the new group composition law

$$(g_1, g_2)(g'_1, g'_2) = (g_1 \cdot g'_1, g_2 * g'_2)$$

and the identity element defined to be (e_1, e_2) where e_1 and e_2 are the identities for G_1 and G_2 respectively. The order of the group is then equal to the all the possible pairings, i.e. $|G_1 \times G_2| = |G_1||G_2|$.

Now we know that order 2 and 3 groups are uniquely Z_2 and Z_3 (Z_1 technically is also a group, but it is a trivial as it consists only of the identity element) – using the above methodology, we can construct groups of any order (which is in general not unique). For an order 4 group, we can product two Z_2 groups together since $|Z_2 \times Z_2| = |Z_2||Z_2| = 4$. We now assert that

$$Z_2 \times Z_2 \cong V_4.$$

Let's prove this statement by explicitly constructing the bijective map. On the left of Eq. (87), suppose the first Z_2 has elements (e_1, μ_1) and the second Z_2 has elements (e_2, μ_2) , then the product group $Z_2 \times Z_2$ possess all the possible pairings

$$Z_2 \times Z_2 = \{(e_1, e_2), (e_1, \mu_2), (\mu_1, e_2), (\mu_1, \mu_2)\}.$$

let's construct a multiplication table

(e_1, e_2)	(e_1, μ_2)	(μ_1, e_2)	(μ_1, μ_2)
(e_1, μ_2)	(e_1, e_2)	(μ_1, μ_2)	(μ_1, e_2)
(μ_1, e_2)	(μ_1, μ_2)	(e_1, e_2)	(e_1, μ_2)
(μ_1, μ_2)	(μ_1, e_2)	(e_1, μ_2)	(e_1, e_2)

which you should check that is correct. Now consider the bijective map i

$$i : Z_2 \times Z_2 \rightarrow V_4 ; (e_1, e_1) \mapsto e , (e_1, \mu_2) \mapsto a , (\mu_1, e_2) \mapsto b , (\mu_1, \mu_2) \mapsto c$$

then you can check that the group composition of the elements in Eq. (90) is the same as the V_4 composition laws Eq. (84), e.g. $(e_1, \mu_2)(\mu_1, e_2) = (\mu_1, \mu_2) \Leftrightarrow ab = c$ etc. Thus $Z_2 \times Z_2 \cong V_4 \square$.

You can have fun constructing groups of higher order than 4 by making product groups. In fact, there is nothing to stop you from making product groups with more than 2 groups. For example, one can make a triple product group $G_1 \times G_2 \times G_3$, where the group laws will have the form $(g_1, g_2, g_3)(g'_1, g'_2, g'_3) = (g_1g'_1, g_2g'_2, g_3g'_3)$ etc. We have thus taken the first steps in parsing the statement “The Standard Model is described by the group $SU(3) \times SU(2) \times U(1)$ ”.

Subgroups

Suppose G is a group, and $H \subseteq G$ is a subset of G . Let $h_1, h_2 \in H$ be elements that belong to H (and also G of course). We can use the group composition law to calculate $h_1h_2 = g$ where $g \in G$. In general g does not have to belong to H . However, suppose that we can find a subset H where $h_1h_2 \in H \forall h_1, h_2 \in H$, then H is a group which “lives” in G , and we call H a **subgroup** of G .

In other words, if we take all possible pairs of the elements of a subset $H \subseteq G$, the composition of them results in another element in H , then H is a **subgroup**¹² of G . A subgroup is a group, hence it obeys all the group axioms as usual. Let’s now be precise.

(Definition) Subgroup: Let G be a group. A subset $H \subseteq G$ is a *subgroup* of G if

- The identity of G is in H . (Can you see why?)
- For every $h_1, h_2 \in H$, $h_1h_2 \in H$. We say that H is *closed* under the group law of G .
- If h is in H , then its inverse $h^{-1} \in H$. We say that H is *closed* under inversion. This is actually *not* a redundant requirement since closure under group law does not cover this possibility.

For any group G , there is always the trivial subgroup consisting of nothing but the identity of G , $\{e\}$. Also, since $G \subseteq G$, G is a subgroup of itself. So since mathematicians are a prickly precise bunch, when you ask one of them “what are all the subgroups of G ”, they have to grumpily state the trivial subgroup, G (and then all the other subgroups besides that). To bypass this annoyance, mathematicians define a

¹²If you are high-brow mathematician, you can also say that H embeds in G . “Embeds” means, roughly, a subset of which “preserves the structure”. In group theory, structure means the group laws, but you will see other applications of embedding when you study general relativity for example.

new term, **proper subgroups** of G , which is the set of all possible subgroups of G minus the trivial subgroup and G .

With this definition, Z_2 then has no proper subgroups (but two subgroups). How about Z_3 and Z_4 ? We can check by brute force in the following way.

Example: Given $Z_3 = \{e, a, a^2\}$, we want to see if we can find a proper subgroup $H \subseteq G$. We start by putting e in H . Now since H is a proper subgroup, $|H| = 2$, so there is only one other element which can either be a or a^2 . Suppose $H = \{e, a\}$, using the group laws $ea = a$ but $aa = a^2 \notin H$, so $\{e, a\}$ cannot be a subgroup of G . Suppose $H = \{e, a^2\}$, then $ea^2 = a^2$ and $a^2a^2 = a(a^3) = ae = a \notin \{e, a^2\}$ hence Z_3 has no proper subgroups.

Example: Given $Z_4 = \{e, a, a^2, a^3\}$. Any proper subgroup of Z_4 must be of order 2 or 3. For order 3, $H = \{e, a, a^2\}$, $H = \{e, a, a^3\}$, $H = \{e, a^2, a^3\}$ – which you can quickly rule out using the requirement that all the elements must be closed under inversion. For order 2, we have $H = \{e, a\}$, $H = \{e, a^2\}$, $H = \{e, a^3\}$. But closure under inversion means that the element other than the identity must be its own inverse, and only a^2 has this property $a^2a^2 = e$, so the only proper subgroup of Z_4 is $H = \{e, a^2\}$.

You will be asked to find the proper subgroups of V_4 in the homework.

3.2 Properties and Actions of Groups

In this section, we continue our general trend of introducing properties of groups as we discuss new classes of groups. In addition we will further develop the important conceptual idea of “group operators doing stuff to objects”. In the introduction Chapter 1, we started with the idea that the operators of groups act (“do stuff”) to some other objects which abstractly have the mathematical structure of a vector space. We have not discussed it, but it turns out that *vector spaces are also sets*, hence we can define actions of group operators on sets. In fact, since the group elements themselves are contained in sets, we can define the action of group operators *on themselves*.

3.2.1 Action of Group Operators on Set

Let’s begin by introducing the notion of action of group operators on some set. We already know that we can think of group operators as “doing stuff to objects”, indeed we start our study of group theory with doing things to the square by rotating and reflecting it. So *actions are simply maps*. We now return to this notion.

Suppose $G = \{e, g_1, g_2, \dots\}$ is a group and $X = \{x_1, x_2, x_3, \dots\}$ is some set (which may or may not be a group for the moment). Suppose that each element of G maps X back to X , i.e.

$$\begin{aligned} e : X &\rightarrow X \\ g_1 : X &\rightarrow X \\ g_2 : X &\rightarrow X \\ &\vdots \end{aligned}$$

Note that each of the maps g_i are in principle *different*, so if we want to completely define the **action of G on X** , we will have to individually specify what each element of g does to every element of X . But this is not as bad as it sounds, since G is a group, so the group elements must obey the group laws. Indeed, once we have specified the action of the *generators* of G on X , we can construct all the other actions of the elements of G on X . Finally, since inverses of g_i must exist according to the Group Axioms, and the inverse $g_i^{-1} : X \rightarrow X$ also, this mean that the maps $g_i : X \rightarrow X$ must be *bijective*.

Instead of the clunky $g_1 : X \rightarrow X$ notation we used above, we are also familiar with writing the maps in the usual shorthand way $g_1(x)$, $g_2(x)$ for $x \in X$, or even shorter hand simply $g_i(x)$, $i = 0, 1, 2, \dots$,

which is a perfectly fine notation. Instead of labeling with i , we write

$$\text{Acting on single element } g_i(x) = L_{g_i}(x) \text{ or } g(x) = L_g(x)$$

where we drop the redundant subscript i in the right hand side. We will see both kinds of notation in these lectures – often it will be clear in context.

Since G is a group, we can compose the two operators together in the usual way, dropping the x as this is always true regardless of the set X

$$g_1g_2 = L_{g_1}L_{g_2} = L_{g_1g_2}.$$

Sometimes, instead of just an element of x , we can act “wholesale” on a set of elements. Let $Y = \{x_1, x_2\} \subseteq X$, then we can act on Y with L_g . The notation is then

$$\text{Acting on Sets of elements } L_g(Y) = \{L_g(x_1), L_g(x_2)\}.$$

The extra L seems to be a step back, since we have to write more stuff, there are two reasons why we attach this extra term. Firstly, sometimes a group can have *different actions on different sets* so we can distinguish them with L_g, P_g, R_g etc. Secondly, more importantly *a group can act on other groups, in particular, it can act on itself since G is also a set*, so gG can sometimes be confusing (although still used). The action of groups on itself is a very powerful idea, and we will devote the next section to study it.

We close this section with a useful definition.

(Definition) Orbit: Suppose G is a group and X is a set, and $x \in X$ is some point or element. Then the *orbit* of the action of G on x is the set of all points which can be reached by acting on x with G , i.e.

$$O_x = \{L_g(x), \forall g \in G\}.$$

The name “orbit” brings into mind the Earth going around the Sun etc., and here you can think of the Sun as the action of x and the “orbit” being all the possible points that can be reached with a single action. Different actions have difference orbits.

Action of Group on Itself, Partitioning

Suppose G is a group and L_g is the action of the group on the group itself, so for each element g , this element can act on another element g' and we write

$$L_g(g').$$

Since the target of the action is the same group as the operators themselves, we can now define what L_g is explicitly. Let’s do a few important ones.

Left Action/Multiplication: This is simply

$$L_g(g') = gg'.$$

With left, there is right:

Right action/Multiplication: This is, giving it a new name just to be clear

$$R_g(g') = g'g.$$

In particular, since the action of the group on itself are constructed using the group elements and the group law, the result of an action is *always another element of the group* by closure rules. Hence the group operators maps the group back to itself, i.e.

$$L_g : G \rightarrow G.$$

If we are given a subset of group elements (which may or may not be a subgroup) $H \subseteq G$, acting on H with a group operator g will create a new subset which is also a subset of G (by closure of G)

$$L_g(H) = H' \subseteq G,$$

where H' is some other subset of G . H' is also termed the **orbit** of the element g on H . As an explicit example, suppose $G = \{e, g_1, g_2, \dots\}$ and $H = \{g_2, g_3\}$ then $L_{g_1}(H) = \{g_1g_2, g_1g_3\}$ and of course since by closure $g_1g_2, g_1g_3 \in G$, $H' \subseteq G$.

Now if H itself is not only a subset of G but is also a **subgroup** of G , then we can show that left (and also right) action of G on H **partitions** the group into **cosets** in the following important way.

(Definition) Cosets: Given a subgroup $H = \{e, h_1, h_2, \dots\}$ of G , and then for any given element $g \in G$ (which may be also in H) we can construct a new set called a **left coset** by acting on the set H with left action of g , i.e.

$$L_g(H) = gH \equiv \{ge, gh_1, gh_2, \dots\}.$$

Note that we have *defined* the name of the coset to be gH (we could have called it Bob too). Similarly we can define a **right coset** of g by acting from the *right* viz

$$R_g(H) = Hg \equiv \{eg, h_1g, h_2g, \dots\}.$$

If G is Abelian, then $gH = Hg$ (can you see why?)

Now let's make an assertion: *the cosets partition G* . Recall from Chapter 2, *partitioning* means that the cosets are **disjoint** from each other or they are the same coset. Let's make this more precise (we will discuss left cosets – the right cosets is similar)

- Suppose g_1H and g_2H are cosets, then $g_1H \cap g_2H = \emptyset$ or $g_1H \cap g_2H = g_1H = g_2H$, i.e. two cosets are either disjoint or identical. *Proof:* Suppose g_1H and g_2H have at least one element in common, let's call this element $g_1h_1 = g_2h_2$ respectively for $h_1, h_2 \in H$. Since from closure of H , $h_2h_1^{-1} \in H$, then $g_1H = g_2h_2h_1^{-1}H = g_2h_3H = g_2H$ since h_3H will result back in the same set H by closure of H . Hence if there exist a common element $g_1H = g_2H$ is the same coset \square .
- $g_1H = g_2H$ iff $g_1^{-1}g_2 \in H$. *Proof:* Let $g_1^{-1}g_2 = h$, or $g_2 = g_1h$ then $g_2H = g_1hH = g_1H$. Conversely, if $g_1^{-1}g_2 \in H$ then $g_1H = g_2H$, or in words g_1 and g_2 are in the same *orbit* of H \square .
- Every element $g \in G$ belongs to some coset. *Proof:* Since $e \in H$, and every g must be in a coset gH \square .

In Chapter 2, we say that an **equivalence relation** \sim partitions a set. What is the equivalence relation that partitions G ? The completely unhelpful (but correct) answer is “being in a coset”. The better answer is to remember that elements which are “equivalent” share a property – and in this case the property is that belong to a coset gH associated with the element g . So if you like, an element in a coset g_1H share the property g_1 -ness.

We can collect all the cosets into a set (see how useful sets are), and such a set is called a **Coset Space**:

(Definition) Coset Space: G/H the set of all left¹³ cosets of the subgroup H of G . The number of cosets is called its **index** and written as $|G/H|$.

A note on notation and jargon: You may have remembered we defined the “quotient” of two sets A and B as A/B back in Chapter 2. The notation here means a different thing (although we will come back

¹³The right coset space is, funnily, H/G , but we will not discuss it in these lectures.

and confuse you further later) – some mathematicians like to use $(G : H)$ instead to distinguish the two but no physicists do that so we won't use it.

We can now state a very important and famous Theorem.

(Theorem) Lagrange: *Let G be a finite order group, and H is a subgroup of G . Then*

$$|G| = |G/H||H|.$$

Proof: Since $|G|$ is finite, $|H|$ must also be finite. Since $|G/H|$ is the number of disjoint sets (i.e. the left cosets), each with size $|H|$, then

$$|G| = \text{Total Disjoint cosets } |G/H| \times \text{Cardinality of each coset } |H|$$

and we are done \square .

Don't let this innocuous and simple result fool you: *it puts a very strong restriction on the order of all possible subgroups* since the order of any subgroup *must be divisor of $|G|$* . In other words, only subgroups of certain sizes are allowed.

Example: How many proper subgroups are there in an order 7 Cyclic group Z_7 ? Answer: since $|Z_7| = 7$, this is a prime number and the only integers that subdivide 7 is 1 or 7, which gives the trivial subgroup and Z_7 itself. These are not proper subgroups, so the total number of proper subgroups of Z_7 is zero.

We can immediately get two corollaries for free.

(Corollary): If $|G|$ is a prime number n , then its has no proper subgroups.

(Corollary): The order of any element $g \in G$ divides $|G|$. (You will be asked to prove this in the homework).

Example: Recall the Dihedral D_4 group that we studied in Chapter 1. The elements of the group are $D_4 = \{e, R, R^2, R^3, m_1, m_2, m_3, m_4\}$, so $|D_4| = 8$. Now Lagrange's Theorem tells us that any subgroup of D_4 must be have an order that divides 8. The subgroups of D_4 are the cyclic group $Z_4 = \{e, R, R^2, R^3\}$, with $|Z_4| = 4$, and the Z_2 groups consisting of a reflection and the identity i.e. $Z_2 = \{e, m_4\}$, with $|Z_2| = 2$. Clearly, both 4 and 2 subdivides 8.

Fermat's Little Theorem

For every integer a and every prime p , $a^p \bmod p = a \bmod p$.

PROOF By the division algorithm, $a = pm + r$, where $0 \leq r < p$. Thus, $a \bmod p = r$, and it suffices to prove that $r^p \bmod p = r$. If $r = 0$, the result is trivial, so we may assume that $r \in U(p)$. [Recall that $U(p) = \{1, 2, \dots, p - 1\}$ under multiplication modulo p .] Then, by the preceding corollary, $r^{p-1} \bmod p = 1$ and, therefore, $r^p \bmod p = r$. ■

Classification of Groups of Order $2p$

Let G be a group of order $2p$, where p is a prime greater than 2. Then G is isomorphic to Z_{2p} or D_p .

PROOF We assume that G does not have an element of order $2p$ and show that $G \approx D_p$. We begin by first showing that G must have an element of order p . By our assumption and Lagrange's Theorem, any nonidentity element of G must have order 2 or p . Thus, to verify our assertion, we may assume that every nonidentity element of G has order 2. In this case, we have for all a and b in the group $ab = (ab)^{-1} = b^{-1}a^{-1} = ba$, so that G is Abelian. Then, for any nonidentity elements $a, b \in G$ with $a \neq b$, the set $\{e, a, b, ab\}$ is closed and therefore is a subgroup of G of order 4. Since this contradicts Lagrange's Theorem, we have proved that G must have an element of order p ; call it a .

Now let b be any element not in $\langle a \rangle$. Then by Lagrange's Theorem and our assumption that G does not have an element of order $2p$, we have that $|b| = 2$ or p . Because $|\langle a \rangle \cap \langle b \rangle|$ divides $|\langle a \rangle| = p$ and $\langle a \rangle \neq \langle b \rangle$ we have that $|\langle a \rangle \cap \langle b \rangle| = 1$. But then $|b| = 2$, for otherwise, by Theorem 7.2 $|\langle a \rangle \langle b \rangle| = |\langle a \rangle||\langle b \rangle| = p^2 > 2p = |G|$, which is impossible. So, any element of G not in $\langle a \rangle$ has order 2.

Next consider ab . Since $ab \notin \langle a \rangle$, our argument above shows that $|ab| = 2$. Then $ab = (ab)^{-1} = b^{-1}a^{-1} = ba^{-1}$. Moreover, this relation completely determines the multiplication table for G . [For example, $a^3(ba^4) = a^2(ab)a^4 = a^2(ba^{-1})a^4 = a(ab)a^3 = a(ba^{-1})a^3 = (ab)a^2 = (ba^{-1})a^2 = ba$.] Since the multiplication table for all noncyclic groups of order $2p$ is uniquely determined by the relation $ab = ba^{-1}$, all noncyclic groups of order $2p$ must be isomorphic to each other. But of course, D_p , the dihedral group of order $2p$, is one such group. ■

Conjugacy Classes and Normal Subgroups

There is another action of a group on itself which partitions the group, called **conjugation**, and is defined as follows.

(Definition) Conjugation: Suppose G is a group and $g, g' \in G$. The *conjugation action* of element g on element g' is defined to be

$$\text{Conjugation } L_g(g') \equiv gg'g^{-1}.$$

(Definition) Conjugacy Equivalence Relation: Given two elements $g_1, g_2 \in G$. g_1 and g_2 are said to be **conjugate** to each if there exist *any* $g \in G$ such that

$$L_g(g_1) = g_2, \text{i.e. } g_2 = gg_1g^{-1} \text{ for any } g$$

and we write $g_1 \sim g_2$.

Colloquially speaking, a “conjugacy operation of g on g_1 , $L_g(g_1)$ ” does the following: We take an element g_1 , and instead of doing the operation g_1 (say rotating like an element in D_4), we first “transform to a new location” by doing a g^{-1} operation, “do the g_1 operation there”, and then transform back to the original location by undoing the g^{-1} operation.

We can quickly prove that conjugacy is an equivalent relationship in the following way

- Reflexive: if $g_1 \sim g_1$ always since we can choose $g = e$.
- Symmetric: if $g_1 \sim g_2 \Rightarrow g_2 \sim g_1$ we have inverses.
- Transitive: if $g_1 \sim g_2$ and $g_2 \sim g_3$ then $g_1 \sim g_3$. Check: given $g_1 = gg_2g^{-1}$ and $g_2 = g'g_3g'^{-1}$, then $g_1 = g(g'g_3g'^{-1})g^{-1} = g''g_3g''^{-1}$, where g, g', g'' are any elements of G .

Since conjugation is an equivalent relationship, it partitions the group into disjoint sets, called **Conjugacy Classes** – every element in each class is conjugate to each other. For Abelian groups, the conjugacy classes are particularly easy to understand – every element is its own conjugate classes since all the elements commute (you should work this out to convince yourself). Also, the identity element is its own conjugacy class for all groups.

Some unions of conjugacy classes themselves are also subgroups of G . These subgroups are called **Normal Subgroups**.

(Definition) Normal Subgroups: A *normal subgroup* H of G is a subgroup whose elements are conjugate to each other (not all unions form a subgroup though). To be specific, a group H is *normal* if for any pair of $h \in H$ and $g \in G$, there exist $h' \in H$ such that

$$L_g(h) = ghg^{-1} = h'.$$

Another way of stating this condition is to say that the left and right cosets of g are equal, viz.

$$gH = Hg$$

A **proper normal subgroup** is a normal subgroup which is neither G or the trivial subgroup $\{e\}$.

If a group possess a proper normal subgroup, it is in a sense “not fundamental” as it means that the group can be “factored into smaller groups”. We use quotations on “factor” because those are just words and it is not clear at this moment what we mean by that. An analogy is the notion of prime numbers – a prime number x cannot be “factored” into integers y and z i.e. $x = yz$ unless either y or z is 1. Another way of saying is that there exist no divisor for a prime number except 1. To complete this analogy in group theory, we need to define a notion of what is a divisor – which is called quotient¹⁴

(Definition) Quotient Group: If H is a proper normal subgroup of G , then we can from the set of left cosets

$$G/H = \{gH; g \in G\}.$$

The left coset G/H is a group under the group laws of G . We call G/H the **Quotient Group** of G for H , and pronounce it “ G quotient H ”.

Proof: Let $g_1, g_2 \in G$, so $g_1H, g_2H \in G/H$. Now, we can show that the coset obey all the group laws – note that we will make frequent use of the relation $gH = Hg$ which is a property of normal subgroups H .

- (Closure): By cleverly using associativity (inheritance from G), we can rewrite

$$\begin{aligned} (g_1H)(g_2H) &= g_1(Hg_2)H \\ &= g_1(g_2H)H \\ &= (g_1g_2)\underbrace{HH}_H \\ &= (g_1g_2)H \\ &= g_3H \in G/H \end{aligned}$$

- (Identity): (eH) is the identity. *Proof:* $(eH)(g_1H) = (Hg_1)H = g_1HH = g_1H$.
- (Inverse): The inverse for (g_1H) is $(g_1^{-1}H)$. *Proof:* $(g_1^{-1}H)(g_1H) = (Hg_1^{-1})(g_1H) = (eH)$.

Quotient groups are extremely important in particle physics and the idea of **spontaneously broken symmetries** which you will encounter when you study Quantum Field Theory and/or Particle Physics next term.

Now, like the fact that prime number occupy a very special place in number theory, groups which cannot be factored (i.e. has no proper normal subgroups hence possess no quotient groups) occupy a very special place in group theory. These groups are called **Finite Simple Groups**. The existence of simple groups, and full *classification* of them is one of the great achievements of modern mathematics – the program is completed only in the last 30-40 years. This means that we know of every single finite simple group exists – an amazing and remarkable fact. Indeed, there is a deep connection between Finite Simple groups and prime number – via Lagrange's Theorem, it is clear that all prime-ordered cyclic group Z_n with n prime is a Simple Group. Another crazy fact is that we also know of a special kind of finite simple group called **sporadic groups** with the biggest order, the so-called **Fisher-Griess Monster Group** M whose order is¹⁵

$$|M| = 2^{46} \cdot 3^{20} \cdot 5^9 \cdot 7^6 \cdot 11^2 \cdot 13^3 \cdot 17 \cdot 19 \cdot 23 \cdot 29 \cdot 31 \cdot 41 \cdot 47 \cdot 59 \cdot 71.$$

Definitions Permutation of A, Permutation Group of A

A **permutation** of a set A is a function from A to A that is both one-to-one and onto. A **permutation group** of a set A is a set of permutations of A that forms a group under function composition.

Although groups of permutations of any nonempty set A of objects exist, we will focus on the case where A is finite. Furthermore, it is customary, as well as convenient, to take A to be a set of the form $\{1, 2, 3, \dots, n\}$ for some positive integer n . Unlike in calculus, where most functions are defined on infinite sets and are given by formulas, in algebra, permutations of finite sets are usually given by an explicit listing of each element of the domain and its corresponding functional value. For example, we define a permutation α of the set $\{1, 2, 3, 4\}$ by specifying

$$\alpha(1) = 2, \quad \alpha(2) = 3, \quad \alpha(3) = 1, \quad \alpha(4) = 4.$$

A more convenient way to express this correspondence is to write α in array form as

$$\alpha = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 1 & 4 \end{bmatrix}.$$

Here $\alpha(j)$ is placed directly below j for each j . Similarly, the permutation β of the set $\{1, 2, 3, 4, 5, 6\}$ given by

$$\beta(1) = 5, \quad \beta(2) = 3, \quad \beta(3) = 1, \quad \beta(4) = 6, \quad \beta(5) = 2, \quad \beta(6) = 4$$

is expressed in array form as

$$\beta = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 5 & 3 & 1 & 6 & 2 & 4 \end{bmatrix}.$$

Composition of permutations expressed in array notation is carried out from right to left by going from top to bottom, then again from top to bottom. For example, let

$$\sigma = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 4 & 3 & 5 & 1 \end{bmatrix}$$

and

$$\gamma = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 4 & 1 & 2 & 3 \end{bmatrix};$$

then

$$\gamma\sigma = \left[\begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \\ \downarrow & & & & \\ 5 & 4 & 1 & 2 & 3 \end{array} \right] \left[\begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \\ \downarrow & & & & \\ 2 & 4 & 3 & 5 & 1 \end{array} \right] = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 2 & 1 & 3 & 5 \end{bmatrix}$$

On the right we have 4 under 1, since $(\gamma\sigma)(1) = \gamma(\sigma(1)) = \gamma(2) = 4$, so $\gamma\sigma$ sends 1 to 4. The remainder of the bottom row $\gamma\sigma$ is obtained in a similar fashion.

We are now ready to give some examples of permutation groups.

■ EXAMPLE 1 Symmetric Group S_3 Let S_3 denote the set of all one-to-one functions from $\{1, 2, 3\}$ to itself. Then S_3 , under function composition, is a group with six elements. The six elements are

$$\varepsilon = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{bmatrix}, \quad \alpha = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{bmatrix}, \quad \alpha^2 = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{bmatrix},$$

$$\beta = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{bmatrix}, \quad \alpha\beta = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{bmatrix}, \quad \alpha^2\beta = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix}.$$

Note that $\beta\alpha = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} = \alpha^2\beta \neq \alpha\beta$, so that S_3 is non-Abelian. ■

The relation $\beta\alpha = \alpha^2\beta$ can be used to compute other products in S_3 without resorting to the arrays. For example, $\beta\alpha^2 = (\beta\alpha)\alpha = (\alpha^2\beta)\alpha = \alpha^2(\beta\alpha) = \alpha^2(\alpha^2\beta) = \alpha^4\beta = \alpha\beta$.

Example 1 can be generalized as follows.

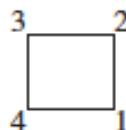
EXAMPLE 2 Symmetric Group S_n Let $A = \{1, 2, \dots, n\}$. The set of all permutations of A is called the *symmetric group of degree n* and is denoted by S_n . Elements of S_n have the form

$$\alpha = \begin{bmatrix} 1 & 2 & \dots & n \\ \alpha(1) & \alpha(2) & \dots & \alpha(n) \end{bmatrix}.$$

It is easy to compute the order of S_n . There are n choices of $\alpha(1)$. Once $\alpha(1)$ has been determined, there are $n - 1$ possibilities for $\alpha(2)$ [since α is one-to-one, we must have $\alpha(1) \neq \alpha(2)$]. After choosing $\alpha(2)$, there are exactly $n - 2$ possibilities for $\alpha(3)$. Continuing along in this fashion, we see that S_n has $n(n - 1) \cdots 3 \cdot 2 \cdot 1 = n!$ elements. We leave it to the reader to prove that S_n is non-Abelian when $n \geq 3$ (Exercise 43). ■

The symmetric groups are rich in subgroups. The group S_4 has 30 subgroups, and S_5 has well over 100 subgroups.

EXAMPLE 3 Symmetries of a Square As a third example, we associate each motion in D_4 with the permutation of the locations of each of the four corners of a square. For example, if we label the four corner positions as in the figure below and keep these labels fixed for reference, we may describe a 90° counterclockwise rotation by the permutation



$$\rho = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{bmatrix},$$

whereas a reflection across a horizontal axis yields

$$\phi = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{bmatrix}.$$

These two elements generate the entire group (that is, every element is some combination of the ρ 's and ϕ 's).

When D_4 is represented in this way, we see that it is a subgroup of S_4 . ■

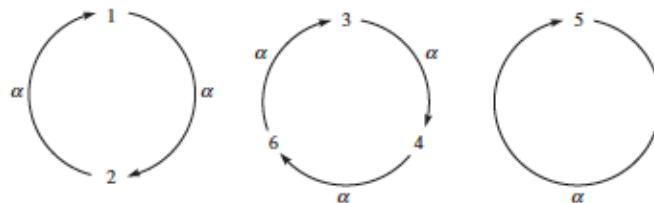
Cycle Notation

There is another notation commonly used to specify permutations. It is called *cycle notation* and was first introduced by the great French mathematician Cauchy in 1815. Cycle notation has theoretical advantages in that certain important properties of the permutation can be readily determined when cycle notation is used.

As an illustration of cycle notation, let us consider the permutation

$$\alpha = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 1 & 4 & 6 & 5 & 3 \end{bmatrix}.$$

This assignment of values could be presented schematically as follows.



Although mathematically satisfactory, such diagrams are cumbersome. Instead, we leave out the arrows and simply write $\alpha = (1, 2)(3, 4, 6)(5)$. As a second example, consider

$$\beta = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 5 & 3 & 1 & 6 & 2 & 4 \end{bmatrix}.$$

In cycle notation, β can be written $(2, 3, 1, 5)(6, 4)$ or $(4, 6)(3, 1, 5, 2)$, since both of these unambiguously specify the function β . An expression of the form (a_1, a_2, \dots, a_m) is called a *cycle of length m* or an *m-cycle*.

A multiplication of cycles can be introduced by thinking of a cycle as a permutation that fixes any symbol not appearing in the

cycle. Thus, the cycle $(4, 6)$ can be thought of as representing the permutation $\begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 2 & 3 & 6 & 5 & 4 \end{bmatrix}$. In this way, we can multiply cycles by thinking of them as permutations given in array form. Consider the following example from S_8 . Let $\alpha = (13)(27)(456)(8)$ and $\beta = (1237)(648)(5)$. (When the domain consists of single-digit integers, it is common practice to omit the commas between the digits.) What is the cycle form of $\alpha\beta$? Of course, one could say that $\alpha\beta = (13)(27)(456)(8)(1237)(648)(5)$, but it is usually more desirable to express a permutation in a *disjoint* cycle form (that is, the various cycles have no number in common). Well, keeping in mind that function composition is done from right to left and that each cycle that does not contain a symbol fixes the symbol, we observe that (5) fixes 1; (648) fixes 1; (1237) sends 1 to 2; (8) fixes 2; (456) fixes 2; (27) sends 2 to 7; and (13) fixes 7. So the net effect of $\alpha\beta$ is to send 1 to 7. Thus, we begin $\alpha\beta = (17 \cdots) \cdots$. Now, repeating the entire process beginning with 7, we have, cycle by cycle, right to left,

$$7 \rightarrow 7 \rightarrow 7 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 3,$$

so that $\alpha\beta = (173 \cdots) \cdots$. Ultimately, we have $\alpha\beta = (1732)(48)(56)$. The important thing to bear in mind when multiplying cycles is to “keep moving” from one cycle to the next from right to left. (Warning: Some authors compose cycles from left to right. When reading another text, be sure to determine which convention is being used.)

To be sure you understand how to switch from one notation to the other and how to multiply permutations, we will do one more example of each.

If array notations for α and β , respectively, are

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 1 & 3 & 5 & 4 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 4 & 1 & 2 & 3 \end{bmatrix},$$

then, in cycle notation, $\alpha = (12)(3)(45)$, $\beta = (153)(24)$, and $\alpha\beta = (12)(3)(45)(153)(24)$.

To put $\alpha\beta$ in disjoint cycle form, observe that (24) fixes 1; (153) sends 1 to 5; (45) sends 5 to 4; and (3) and (12) both fix 4. So, $\alpha\beta$ sends 1 to 4. Continuing in this way we obtain $\alpha\beta = (14)(253)$.

One can convert $\alpha\beta$ back to array form without converting each cycle of $\alpha\beta$ into array form by simply observing that (14) means 1 goes to 4 and 4 goes to 1; (253) means $2 \rightarrow 5, 5 \rightarrow 3, 3 \rightarrow 2$.

One final remark about cycle notation: Mathematicians prefer not to write cycles that have only one entry. In this case, it is understood that any

missing element is mapped to itself. With this convention, the permutation α above can be written as $(12)(45)$. Similarly,

$$\alpha = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 2 & 4 & 1 & 5 \end{bmatrix}$$

can be written $\alpha = (134)$. Of course, the identity permutation consists only of cycles with one entry, so we cannot omit all of these! In this case, one usually writes just one cycle. For example,

$$\varepsilon = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 3 & 4 & 5 \end{bmatrix}$$

can be written as $\varepsilon = (5)$ or $\varepsilon = (1)$. Just remember that missing elements are mapped to themselves.

Properties of Permutations

We are now ready to state several theorems about permutations and cycles. The proof of the first theorem is implicit in our discussion of writing permutations in cycle form.

Products of Disjoint Cycles

Every permutation of a finite set can be written as a cycle or as a product of disjoint cycles.

PROOF Let α be a permutation on $A = \{1, 2, \dots, n\}$. To write α in disjoint cycle form, we start by choosing any member of A , say a_1 , and let

$$a_2 = \alpha(a_1), \quad a_3 = \alpha(\alpha(a_1)) = \alpha^2(a_1),$$

and so on, until we arrive at $a_1 = \alpha^m(a_1)$ for some m . We know that such an m exists because the sequence $a_1, \alpha(a_1), \alpha^2(a_1), \dots$ must be finite; so there must eventually be a repetition, say $\alpha^i(a_1) = \alpha^j(a_1)$ for some i and j with $i < j$. Then $a_1 = \alpha^m(a_1)$, where $m = j - i$. We express this relationship among a_1, a_2, \dots, a_m as

$$\alpha = (a_1, a_2, \dots, a_m) \cdots .$$

The three dots at the end indicate the possibility that we may not have exhausted the set A in this process. In such a case, we merely choose any element b_1 of A not appearing in the first cycle and proceed to create a new cycle as before. That is, we let $b_2 = \alpha(b_1), b_3 = \alpha^2(b_1)$, and so on, until we reach $b_1 = \alpha^k(b_1)$ for some k . This new cycle will have no elements in

Product of 2-Cycles

Every permutation in S_n , $n > 1$, is a product of 2-cycles.

PROOF First, note that the identity can be expressed as $(12)(12)$, and so it is a product of 2-cycles. we know that every permutation can be written in the form

$$(a_1 a_2 \cdots a_k)(b_1 b_2 \cdots b_l) \cdots (c_1 c_2 \cdots c_s).$$

A direct computation shows that this is the same as

$$\begin{aligned} & (a_1 a_k)(a_1 a_{k-1}) \cdots (a_1 a_2)(b_1 b_l)(b_1 b_{l-1}) \cdots (b_1 b_2) \\ & \quad \cdots (c_1 c_s)(c_1 c_{s-1}) \cdots (c_1 c_2). \end{aligned}$$

This completes the proof. ■

The permutation group will turn out to be a fundamental group in the study of finite groups as first pointed out by Cayley.

Definition Even and Odd Permutations

A permutation that can be expressed as a product of an even number of 2-cycles is called an *even* permutation. A permutation that can be expressed as a product of an odd number of 2-cycles is called an *odd* permutation.

For the identity ε

If $\varepsilon = \beta_1\beta_2 \cdots \beta_r$, where the β 's are 2-cycles, then r is even.

PROOF Clearly, $r \neq 1$, since a 2-cycle is not the identity. If $r = 2$, we are done. So, we suppose that $r > 2$, and we proceed by induction. Suppose that the rightmost 2-cycle is (ab) . Then, since $(ij) = (ji)$, the product $\beta_{r-1}\beta_r$ can be expressed in one of the following forms shown on the right:

$$\begin{aligned}\varepsilon &= (ab)(ab), \\ (ab)(bc) &= (ac)(ab), \\ (ac)(cb) &= (bc)(ab), \\ (ab)(cd) &= (cd)(ab).\end{aligned}$$

If the first case occurs, we may delete $\beta_{r-1}\beta_r$ from the original product to obtain $\varepsilon = \beta_1\beta_2 \cdots \beta_{r-2}$, and therefore, by the Second Principle of Mathematical Induction, $r - 2$ is even. In the other three cases, we replace the form of $\beta_{r-1}\beta_r$ on the right by its counterpart on the left to obtain a new product of r 2-cycles that is still the identity, but where the rightmost occurrence of the integer a is in the second-from-the-rightmost 2-cycle of the product instead of the rightmost 2-cycle. We now repeat the procedure just described with $\beta_{r-2}\beta_{r-1}$, and, as before, we obtain a product of $(r - 2)$ 2-cycles equal to the identity or a new product of r 2-cycles, where the rightmost occurrence of a is in the third 2-cycle from the right. Continuing this process, we must obtain a product of $(r - 2)$ 2-cycles equal to the identity, because otherwise we have a product equal to the identity in which the only occurrence of the integer a is in the leftmost 2-cycle, and such a product does not fix a , whereas the identity does. Hence, by the Second Principle of Mathematical Induction, $r - 2$ is even, and r is even as well. ■

Always Even or Always Odd

If a permutation α can be expressed as a product of an even (odd) number of 2-cycles, then every decomposition of α into a product of 2-cycles must have an even (odd) number of 2-cycles. In symbols, if

$$\alpha = \beta_1\beta_2 \cdots \beta_r \quad \text{and} \quad \alpha = \gamma_1\gamma_2 \cdots \gamma_s,$$

where the β 's and the γ 's are 2-cycles, then r and s are both even or both odd.

PROOF Observe that $\beta_1\beta_2 \cdots \beta_r = \gamma_1\gamma_2 \cdots \gamma_s$ implies

$$\begin{aligned}\varepsilon &= \gamma_1\gamma_2 \cdots \gamma_s\beta_r^{-1} \cdots \beta_2^{-1}\beta_1^{-1} \\ &= \gamma_1\gamma_2 \cdots \gamma_s\beta_r \cdots \beta_2\beta_1,\end{aligned}$$

since a 2-cycle is its own inverse. Thus, the lemma on page 103 guarantees that $s + r$ is even. It follows that r and s are both even or both odd. ■

The set of even permutations in S_n forms a subgroup of S_n .

Definition Alternating Group of Degree n

The group of even permutations of n symbols is denoted by A_n and is called the *alternating group of degree n* .

For $n > 1$, A_n has order $n!/2$.

PROOF For each odd permutation α , the permutation $(12)\alpha$ is even and, by the cancellation property in groups, $(12)\alpha \neq (12)\beta$ when $\alpha \neq \beta$. Thus, there are at least as many even permutations as there are odd ones. On the other hand, for each even permutation α , the permutation $(12)\alpha$ is odd and $(12)\alpha \neq (12)\beta$ when $\alpha \neq \beta$. Thus, there are at least as many odd permutations as there are even ones. It follows that there are

equal numbers of even and odd permutations. Since $|S_n| = n!$, we have $|A_n| = n!/2$. ■

Geometric interpretation of symmetric groups:

Rotations of a Tetrahedron

The 12 rotations of a regular tetrahedron can be conveniently described with the elements of A_4 . The top row of Figure 5.1 illustrates the identity and three 180° “edge” rotations about axes joining midpoints of two

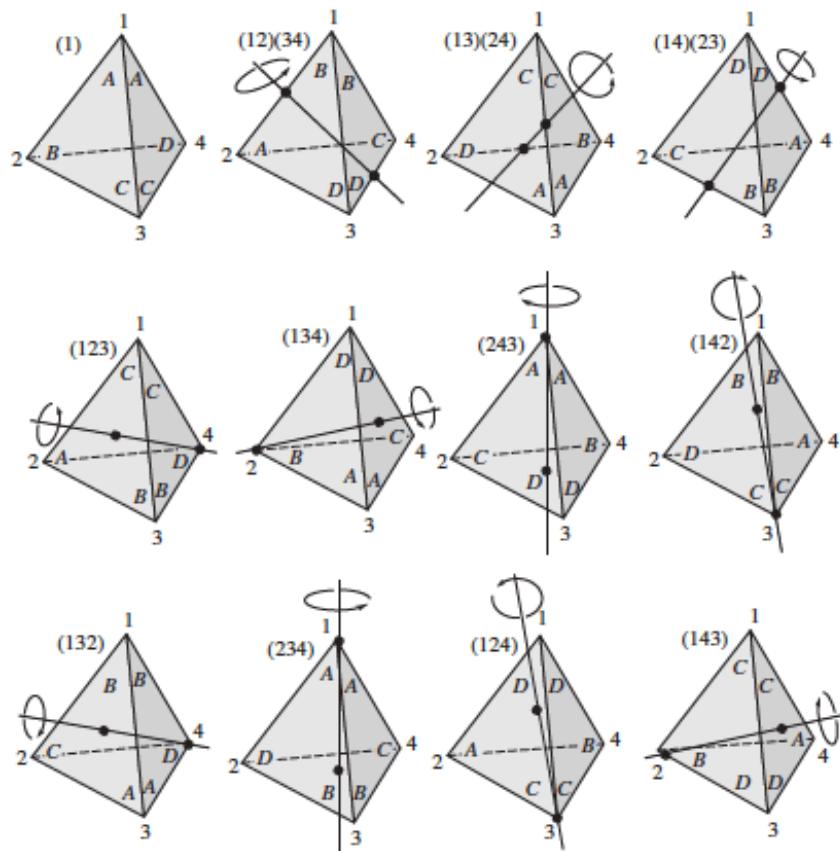


Figure 5.1 Rotations of a regular tetrahedron.

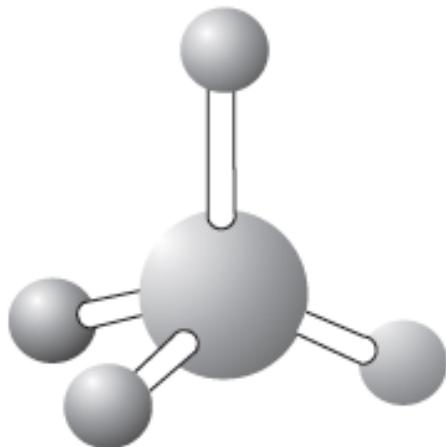
edges. The second row consists of 120° “face” rotations about axes joining a vertex to the center of the opposite face. The third row consists of -120° (or 240°) “face” rotations. Notice that the four rotations in the second row can be obtained from those in the first row by left-multiplying the four in the first row by the rotation (123) , whereas those in the third row can be obtained from those in the first row by left-multiplying the ones in the first row by (132) . ■

The Alternating Group A_4 of Even Permutations of $\{1, 2, 3, 4\}$

(In this table, the permutations of A_4 are designated as $\alpha_1, \alpha_2, \dots, \alpha_{12}$ and an entry k inside the table represents α_k . For example, $\alpha_3 \alpha_8 = \alpha_6$)

	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9	α_{10}	α_{11}	α_{12}
$(1) = \alpha_1$	1	2	3	4	5	6	7	8	9	10	11	12
$(12)(34) = \alpha_2$	2	1	4	3	6	5	8	7	10	9	12	11
$(13)(24) = \alpha_3$	3	4	1	2	7	8	5	6	11	12	9	10
$(14)(23) = \alpha_4$	4	3	2	1	8	7	6	5	12	11	10	9
$(123) = \alpha_5$	5	8	6	7	9	12	10	11	1	4	2	3
$(243) = \alpha_6$	6	7	5	8	10	11	9	12	2	3	1	4
$(142) = \alpha_7$	7	6	8	5	11	10	12	9	3	2	4	1
$(134) = \alpha_8$	8	5	7	6	12	9	11	10	4	1	3	2
$(132) = \alpha_9$	9	11	12	10	1	3	4	2	5	7	8	6
$(143) = \alpha_{10}$	10	12	11	9	2	4	3	1	6	8	7	5
$(234) = \alpha_{11}$	11	9	10	12	3	1	2	4	7	5	6	8
$(124) = \alpha_{12}$	12	10	9	11	4	2	1	3	8	6	5	7

Many molecules with chemical formulas of the form AB_4 , such as methane (CH_4) and carbon tetrachloride (CCl_4), have A_4 as their symmetry group. Figure 5.2 shows the form of one such molecule.



Cayley's Theorem (1854)

"As for everything else, so for a mathematical theory: beauty can be perceived but not explained." (Arthur Cayley)

Every group is isomorphic to a group of permutations.

PROOF To prove this, let G be any group. We must find a group \overline{G} of permutations that we believe is isomorphic to G . Since G is all we have to work with, we will have to use it to construct \overline{G} . For any g in G , define a function T_g from G to G by

$$T_g(x) = gx \quad \text{for all } x \text{ in } G.$$

(In words, T_g is just multiplication by g on the left.) We leave it as an exercise (Exercise 35) to prove that T_g is a permutation on the set of elements of G . Now, let $\overline{G} = \{T_g \mid g \in G\}$. Then, \overline{G} is a group under the operation of function composition. To verify this, we first observe that for any g and h in G we have $T_g T_h(x) = T_g(T_h(x)) = T_g(hx) = g(hx) = (gh)x = T_{gh}(x)$, so that $T_g T_h = T_{gh}$. From this it follows that T_e is the identity and $(T_g)^{-1} = T_{g^{-1}}$ (see Exercise 9). Since function composition is associative, we have verified all the conditions for \overline{G} to be a group.

The isomorphism ϕ between G and \overline{G} is now ready-made. For every g in G , define $\phi(g) = T_g$. If $T_g = T_h$, then $T_g(e) = T_h(e)$ or $ge = he$. Thus, $g = h$ and ϕ is one-to-one. By the way \overline{G} was constructed, we see that ϕ is onto. The only condition that remains to be checked is that ϕ is operation-preserving. To this end, let a and b belong to G . Then

$$\phi(ab) = T_{ab} = T_a T_b = \phi(a)\phi(b).$$

The group \overline{G} constructed previously is called the *left regular representation of G* .

Properties of Isomorphisms Acting on Groups

Suppose that ϕ is an isomorphism from a group G onto a group \bar{G} . Then

1. ϕ^{-1} is an isomorphism from \bar{G} onto G .
2. G is Abelian if and only if \bar{G} is Abelian.
3. G is cyclic if and only if \bar{G} is cyclic.
4. If K is a subgroup of G , then $\phi(K) = \{\phi(k) \mid k \in K\}$ is a subgroup of \bar{G} .
5. If \bar{K} is a subgroup of \bar{G} , then $\phi^{-1}(\bar{K}) = \{g \in G \mid \phi(g) \in \bar{K}\}$ is a subgroup of G .
6. $\phi(Z(G)) = Z(\bar{G})$.

PROOF Properties 1 and 4 are left as exercises (Exercises 33 and 34). Properties 2 and 6 are a direct consequence of property 3 of Theorem 6.2. Property 3 follows from property 4 of Theorem 6.2 and property 1 of Theorem 6.3. Property 5 follows from properties 1 and 4. ■

Theorems 6.2 and 6.3 provide several convenient ways to prove that groups G and \bar{G} are not isomorphic.

1. Observe that $|G| \neq |\bar{G}|$.
2. Observe that G or \bar{G} is cyclic and the other is not.
3. Observe that G or \bar{G} is Abelian and the other is not.
4. Show that largest order of any element in G is not the same as the largest order of any element in \bar{G} .
5. Show that the number of elements of some specific order in G (the smallest order greater than 1 is often the good choice) is not the same as the number of elements of that order in \bar{G} .

Consider these three groups of order 12: Z_{12} , D_6 and A_4 .

A quick check shows that the largest order of any element in the three are 12, 6 and 3, respectively. So no two are isomorphic. Alternatively, the number of elements of order 2 in each is 1, 7, and 3. ■

An Application of Cosets to Permutation Groups

Lagrange's Theorem and its corollaries dramatically demonstrate the fruitfulness of the coset concept. We next consider an application of cosets to permutation groups.

Definition Stabilizer of a Point

Let G be a group of permutations of a set S . For each i in S , let $\text{stab}_G(i) = \{\phi \in G \mid \phi(i) = i\}$. We call $\text{stab}_G(i)$ the *stabilizer of i in G* .

The student should verify that $\text{stab}_G(i)$ is a subgroup of G .

Definition Orbit of a Point

Let G be a group of permutations of a set S . For each s in S , let $\text{orb}_G(s) = \{\phi(s) \mid \phi \in G\}$. The set $\text{orb}_G(s)$ is a subset of S called the *orbit of s under G* . We use $|\text{orb}_G(s)|$ to denote the number of elements in $\text{orb}_G(s)$.

■ EXAMPLE

Let

$$G = \{(1), (132)(465)(78), (132)(465), (123)(456), (123)(456)(78), (78)\}.$$

Then,

$$\begin{aligned}\text{orb}_G(1) &= \{1, 3, 2\}, & \text{stab}_G(1) &= \{(1), (78)\}, \\ \text{orb}_G(2) &= \{2, 1, 3\}, & \text{stab}_G(2) &= \{(1), (78)\}, \\ \text{orb}_G(4) &= \{4, 6, 5\}, & \text{stab}_G(4) &= \{(1), (78)\}, \\ \text{orb}_G(7) &= \{7, 8\}, & \text{stab}_G(7) &= \{(1), (132)(465), (123)(456)\}.\end{aligned}$$

Orbit-Stabilizer Theorem

Let G be a finite group of permutations of a set S . Then, for any i from S , $|G| = |\text{orb}_G(i)| |\text{stab}_G(i)|$.

Definition Automorphism

An isomorphism from a group G onto itself is called an *automorphism* of G .

Let $\mathbf{R}^2 = \{(a, b) \mid a, b \in \mathbf{R}\}$. Then $\phi(a, b) = (b, a)$ is an automorphism of the group \mathbf{R}^2 under componentwise addition. Geometrically, ϕ reflects each point in the plane across the line $y = x$. More generally, any reflection across a line passing through the origin or any rotation of the plane about the origin is an automorphism of \mathbf{R}^2 . ■

Definition **Inner Automorphism Induced by a**

Let G be a group, and let $a \in G$. The function ϕ_a defined by $\phi_a(x) = axa^{-1}$ for all x in G is called the *inner automorphism of G induced by a* .

The action of the inner automorphism of D_4 induced by R_{90} is given in the following table.

x	$\xrightarrow{\phi_{R_{90}}} R_{90}xR_{90}^{-1}$
R_0	$R_{90}R_0R_{90}^{-1} = R_0$
R_{90}	$R_{90}R_{90}R_{90}^{-1} = R_{90}$
R_{180}	$R_{90}R_{180}R_{90}^{-1} = R_{180}$
R_{270}	$R_{90}R_{270}R_{90}^{-1} = R_{270}$
H	$R_{90}HR_{90}^{-1} = V$
V	$R_{90}VR_{90}^{-1} = H$
D	$R_{90}DR_{90}^{-1} = D'$
D'	$R_{90}D'R_{90}^{-1} = D$

Let G be the rotation group of a cube. Label the six faces of the cube 1 through 6. Since any rotation of the cube must carry each face of the cube to exactly one other face of the cube and different rotations induce different permutations of the faces, G can be viewed as a group of permutations on the set $\{1, 2, 3, 4, 5, 6\}$. Clearly, there is some rotation about a central horizontal or vertical axis that carries face number 1 to any other face, so that $|\text{orb}_G(1)| = 6$. Next, we consider $\text{stab}_G(1)$. Here, we are asking for all rotations of a cube that leave face number 1

where it is. Surely, there are only four such motions—rotations of 0° , 90° , 180° , and 270° —about the line perpendicular to the face and passing through its center (see Figure 7.2). Thus, by Theorem 7.4, $|G| = |\text{orb}_G(1)| |\text{stab}_G(1)| = 6 \cdot 4 = 24$. ■

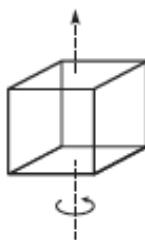


Figure 7.2 Axis of rotation of a cube.

Now that we know how many rotations a cube has, it is simple to determine the actual structure of the rotation group of a cube. Recall that S_4 is the symmetric group of degree 4.

The group of rotations of a cube is isomorphic to S_4

PROOF Since the group of rotations of a cube has the same order as S_4 , we need only prove that the group of rotations is isomorphic to a subgroup of S_4 . To this end, observe that a cube has four diagonals and that the rotation group induces a group of permutations on the four diagonals. But we must be careful not to assume that different rotations correspond to different permutations. To see that this is so, all we need do is show that all 24 permutations of the diagonals arise from rotations. Labeling the consecutive diagonals 1, 2, 3, and 4, it is obvious that there is a 90° rotation that yields the permutation $\alpha = (1234)$; another 90° rotation about an axis perpendicular to our first axis yields the permutation $\beta = (1423)$. See Figure 7.3. So, the group of permutations induced by the rotations contains the eight-element subgroup $\{\varepsilon, \alpha, \alpha^2, \alpha^3, \beta^2, \beta^2\alpha, \beta^2\alpha^2, \beta^2\alpha^3\}$ (see Exercise 63) and $\alpha\beta$, which has order 3. Clearly, then, the rotations yield all 24 permutations, since the order of the rotation group must be divisible by both 8 and 3. ■

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.

What Is a representation?

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

$$D(g_1)D(g_2) = D(g_1g_2)$$

for any two group elements g_1 and g_2 . The matrix² $D(g)$ 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

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

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 in chapter I.1, 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.

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 $4 \otimes 4$ 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-by-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 in chapter I.1.

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.

We shall first introduce matrix and continuous groups.

An Introduction to Continuous Groups

Proposition: The set of Real Numbers \mathbb{R} forms a group under addition.

Proof: We check the group axioms as follows:

- Identity: the element 0 is the identity since $a + 0 = 0 + a = a \forall a \in \mathbb{R}$.
- Closure: $a + b \in \mathbb{R}$.
- Associativity: $a + (b + c) = (a + b) + c$.
- Inverse: For every $a \exists (-a)$ such that $a + (-a) = 0$.

and hence we are done \square . Furthermore since $a + b = b + a$, this is an *abelian* group.

Note that \mathbb{R} is technically *not* a group under multiplication since the element 0 has no inverse as $1/0$ (which you will call ∞) is not an element of \mathbb{R} . However $1/0$ is a tricky customer and causes much angst and requires special care, and often we can define the group $\mathbb{R}^* = \mathbb{R} - \{0\}$ (i.e. real numbers minus the zero) which then forms a group under multiplication¹⁸. What we have not discuss is *why* \mathbb{R} is continuous. The study of continuity is a whole subfield in mathematics called **Point Set Topology**, and digressing there will lead us way too far afield. In these lectures, we will *assert* the following

\mathbb{R} is continuous hence any set or group that can be parameterized by \mathbb{R}^n , or some compact subset, will be continuous. n is the number of dimensions of this group.

What is a Lie group?

The most general definition of a Lie group G is a group that is also a *smooth manifold*. That is, the group “product” and “inverse” operations are smooth functions on the manifold G .

[* An aside on manifolds:

Loosely manifolds are topological spaces that look locally like Euclidean space.

A little more precisely it is a space *together with* a way of identifying it locally with a Euclidean space which is compatible on overlaps.

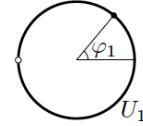
Local description is through maps into Euclidean space on patches which cover the manifold and are said to form an atlas A .

Example 1 (Open Subset of \mathbb{R}^n) Any open subset, \mathcal{O} , of \mathbb{R}^n is a manifold of dimension n . One possible atlas is $A = \{(\mathcal{O}, \varphi_{id})\}$, where φ_{id} is the identity map. That is, $\varphi_{id}(\mathbf{x}) = \mathbf{x}$. Of course one possible choice of \mathcal{O} is \mathbb{R}^n itself.

Example 2 (The Circle) The circle $S^1 = \{ (x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1 \}$ is a manifold of dimension one. One possible atlas is $\mathfrak{A} = \{(U_1, \varphi_1), (U_1, \varphi_2)\}$ where

$$U_1 = S^1 \setminus \{(-1, 0)\} \quad \varphi_1(x, y) = \arctan \frac{y}{x} \text{ with } -\pi < \varphi_1(x, y) < \pi$$

$$U_2 = S^1 \setminus \{(1, 0)\} \quad \varphi_2(x, y) = \arctan \frac{y}{x} \text{ with } 0 < \varphi_2(x, y) < 2\pi$$



Example 3 (S^n) The n -sphere $S^n = \{ \mathbf{x} = (x_1, \dots, x_{n+1}) \in \mathbb{R}^{n+1} \mid x_1^2 + \dots + x_{n+1}^2 = 1 \}$ is a manifold of dimension n . One possible atlas is $\mathfrak{A}_1 = \{ (U_i, \varphi_i), (V_i, \psi_i) \mid 1 \leq i \leq n+1 \}$ where, for each $1 \leq i \leq n+1$,

$$U_i = \{ (x_1, \dots, x_{n+1}) \in S^n \mid x_i > 0 \} \quad \varphi_i(x_1, \dots, x_{n+1}) = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_{n+1})$$

$$V_i = \{ (x_1, \dots, x_{n+1}) \in S^n \mid x_i < 0 \} \quad \psi_i(x_1, \dots, x_{n+1}) = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_{n+1})$$

So both φ_i and ψ_i project onto \mathbb{R}^n , viewed as the hyperplane $x_i = 0$. Another possible atlas is

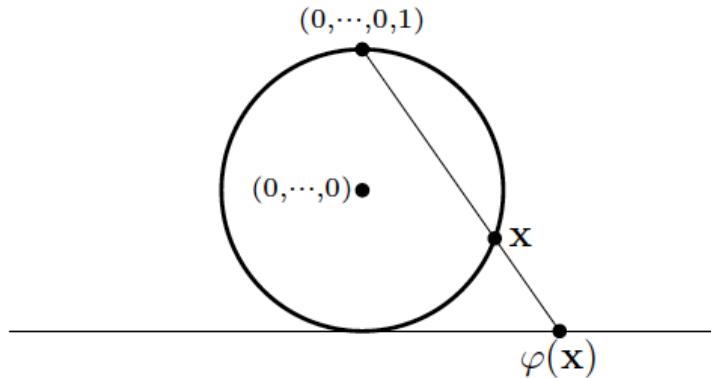
$$\mathfrak{A}_2 = \{ (S^n \setminus \{(0, \dots, 0, 1)\}, \varphi), (S^n \setminus \{(0, \dots, 0, -1)\}, \psi) \}$$

where

$$\varphi(x_1, \dots, x_{n+1}) = \left(\frac{2x_1}{1-x_{n+1}}, \dots, \frac{2x_n}{1-x_{n+1}} \right)$$

$$\psi(x_1, \dots, x_{n+1}) = \left(\frac{2x_1}{1+x_{n+1}}, \dots, \frac{2x_n}{1+x_{n+1}} \right)$$

are the stereographic projections from the north and south poles, respectively.



Both φ and ψ have range \mathbb{R}^n . So we can think of S^n as \mathbb{R}^n plus an additional single “point at infinity”.

Example 4 (Surfaces) Any smooth n -dimensional surface in \mathbb{R}^{n+m} is an n -dimensional manifold. When we say that M is an n -dimensional surface in \mathbb{R}^{n+m} , we mean that M is a subset of \mathbb{R}^{n+m} with the property that for each $\mathbf{z} \in M$, there are

- o a neighbourhood $U_{\mathbf{z}}$ of \mathbf{z} in \mathbb{R}^{n+m}
- o n integers $1 \leq j_1 < j_2 < \dots < j_n \leq n+m$
- o and $m C^\infty$ functions $f_k(\mathbf{x}_{j_1}, \dots, \mathbf{x}_{j_n})$, $k \in \{1, \dots, n+m\} \setminus \{j_1, \dots, j_n\}$

such that the point $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_{n+m}) \in U_{\mathbf{z}}$ is in M if and only if $\mathbf{x}_k = f_k(\mathbf{x}_{j_1}, \dots, \mathbf{x}_{j_n})$ for all $k \in \{1, \dots, n+m\} \setminus \{j_1, \dots, j_n\}$. That is, we may express the part of M that is near \mathbf{z} as

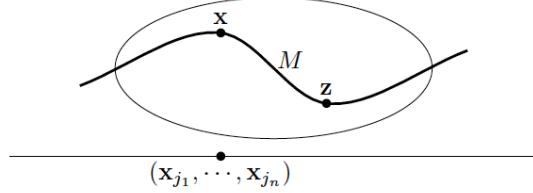
$$\mathbf{x}_{i_1} = f_1(\mathbf{x}_{j_1}, \mathbf{x}_{j_2}, \dots, \mathbf{x}_{j_n})$$

$$\mathbf{x}_{i_2} = f_2(\mathbf{x}_{j_1}, \mathbf{x}_{j_2}, \dots, \mathbf{x}_{j_n})$$

$$\mathbf{x}_{i_m} = f_m(\mathbf{x}_{j_1}, \mathbf{x}_{j_2}, \dots, \mathbf{x}_{j_n})$$

where $\{i_1, \dots, i_m\} = \{1, \dots, n+m\} \setminus \{j_1, \dots, j_n\}$

for some C^∞ functions f_1, \dots, f_m . We may use $\mathbf{x}_{j_1}, \mathbf{x}_{j_2}, \dots, \mathbf{x}_{j_n}$ as coordinates for M in $M \cap U_{\mathbf{z}}$. Of course, an atlas is $\mathcal{A} = \{ (U_{\mathbf{z}}, \varphi_{\mathbf{z}}) \mid \mathbf{z} \in M \}$, with each $\varphi_{\mathbf{z}}(\mathbf{x}) = (\mathbf{x}_{j_1}, \dots, \mathbf{x}_{j_n})$.



Equivalently, M is an n -dimensional surface in \mathbb{R}^{n+m} , if, for each $\mathbf{z} \in M$, there are

- o a neighbourhood $U_{\mathbf{z}}$ of \mathbf{z} in \mathbb{R}^{n+m}
- o and m C^∞ functions $g_k : U_{\mathbf{z}} \rightarrow \mathbb{R}$, such that the vectors $\{ \nabla g_k(\mathbf{z}) \mid 1 \leq k \leq m \}$ are linearly independent

such that the point $\mathbf{x} \in U_{\mathbf{z}}$ is in M if and only if $g_k(\mathbf{x}) = 0$ for all $1 \leq k \leq m$. To get from the implicit equations for M given by the g_k 's to the explicit equations for M given by the f_k 's one need only invoke (possible after renumbering the components of \mathbf{x}) the

Implicit Function Theorem

Let $m, n \in \mathbb{N}$ and let $U \subset \mathbb{R}^{n+m}$ be an open set. Let $\mathbf{g} : U \rightarrow \mathbb{R}^m$ be C^∞ with $\mathbf{g}(\mathbf{z}) = 0$ for some $\mathbf{z} \in U$. Assume that $\det \left[\frac{\partial \mathbf{g}_i}{\partial \mathbf{x}_{n+j}}(\mathbf{z}) \right]_{1 \leq i,j \leq m} \neq 0$. Write $\mathbf{a} = (\mathbf{z}_1, \dots, \mathbf{z}_n)$ and $\mathbf{b} = (\mathbf{z}_{n+1}, \dots, \mathbf{z}_{n+m})$. Then there exist open sets $V \subset \mathbb{R}^{n+m}$ and $W \subset \mathbb{R}^n$ with $\mathbf{a} \in W$ and $\mathbf{z} = (\mathbf{a}, \mathbf{b}) \in V$ such that

for each $\mathbf{x} \in W$, there is a unique $(\mathbf{x}, \mathbf{y}) \in V$ such that $\mathbf{g}(\mathbf{x}, \mathbf{y}) = 0$.

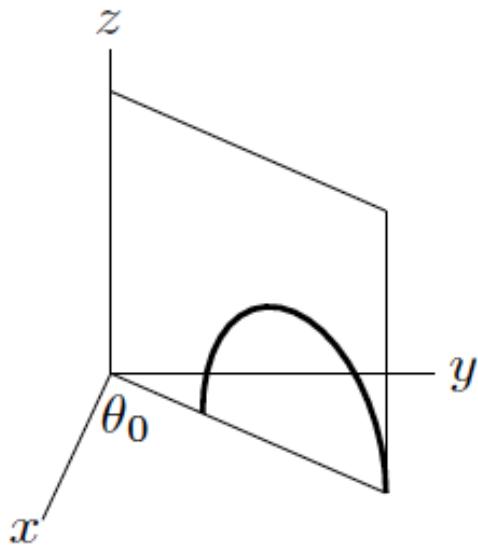
If the \mathbf{y} above is denoted $\mathbf{f}(\mathbf{x})$, then $\mathbf{f} : W \rightarrow \mathbb{R}^m$ is C^∞ , $\mathbf{f}(\mathbf{a}) = \mathbf{b}$ and $\mathbf{g}(\mathbf{x}, \mathbf{f}(\mathbf{x})) = 0$ for all $\mathbf{x} \in W$.

The n -sphere S^n is the n -dimensional surface in \mathbb{R}^{n+1} given implicitly by the equation $g(\mathbf{x}_1, \dots, \mathbf{x}_{n+1}) = \mathbf{x}_1^2 + \dots + \mathbf{x}_{n+1}^2 - 1 = 0$. In a neighbourhood of the north pole (for example, the northern hemisphere), S^n is given explicitly by the equation $\mathbf{x}_{n+1} = \sqrt{\mathbf{x}_1^2 + \dots + \mathbf{x}_n^2}$.

Example 5 (The Torus) The torus T^2 is the two dimensional surface

$$T^2 = \{ (x, y, z) \in \mathbb{R}^3 \mid (\sqrt{x^2 + y^2} - 1)^2 + z^2 = \frac{1}{4} \}$$

in \mathbb{R}^3 . In cylindrical coordinates $x = r \cos \theta$, $y = r \sin \theta$, $z = z$, the equation of the torus is $(r - 1)^2 + z^2 = \frac{1}{4}$. Fix any θ , say θ_0 . Recall that the set of all points in \mathbb{R}^3 that have $\theta = \theta_0$ is like one page in an open book. It is a half-plane that starts at the z axis. The intersection of the torus with that half plane is a circle of radius $\frac{1}{2}$ centred on $r = 1$, $z = 0$. As φ runs from 0 to 2π , the point $r = 1 + \frac{1}{2} \cos \varphi$, $z = \frac{1}{2} \sin \varphi$,



$\theta = \theta_0$ runs over that circle. If we now run θ from 0 to 2π , the circle on the page sweeps out the whole torus. So, as φ runs from 0 to 2π and θ runs from 0 to 2π , the point $(x, y, z) = ((1 + \frac{1}{2} \cos \varphi) \cos \theta, (1 + \frac{1}{2} \cos \varphi) \sin \theta, \frac{1}{2} \sin \varphi)$ runs over the whole torus. So we may build coordinate patches for T^2 using θ and φ (with ranges $(0, 2\pi)$ or $(-\pi, \pi)$) as coordinates.

The material in this aside is not examinable.]

In other words, a Lie Group is a continuous group with the added property that the elements are not just continuous, but the group mapping law is also analytic. Now, $f(a, b)$ is a function of two variables, so analyticity means that the function must be *differentiable in both sets of variables*.

(Definition) Dimensions: The number of independent parameters of the Lie Group is called its Dimensions.

Instead of the approach to Lie groups through manifolds, we will take a mathematically simpler approach. We shall study matrix Lie groups. A matrix Lie group is a set of $n \times n$ matrices (for some fixed n) that is closed under products, inverses and non-singular limits. The last condition means that if A_1, A_2, A_3, \dots is a sequence of matrices which converge to A in G , then A^{-1} exists in G . We are just considering here the limit concept in n^2 dimensional Euclidean space. We will view all matrix groups as groups of real matrices. Real entries will suffice since complex numbers and quaternions can be represented as real matrices as we will see below.

Matrix representation of complex numbers

A good way to see why the matrices $R_\theta = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$ behave the same as the complex numbers $z_\theta = \cos \theta + i \sin \theta$ is to write R_θ as the linear combination

$$R_\theta = \cos \theta \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \sin \theta \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

of the *basis matrices*

$$\mathbf{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{i} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

It is easily checked that

$$\mathbf{1}^2 = \mathbf{1}, \quad \mathbf{1}\mathbf{i} = \mathbf{i}\mathbf{1} = \mathbf{i}, \quad \mathbf{i}^2 = -\mathbf{1},$$

so the matrices $\mathbf{1}$ and \mathbf{i} behave exactly the same as the complex numbers 1 and i .

In fact, the matrices

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix} = a\mathbf{1} + b\mathbf{i}, \quad \text{where } a, b \in \mathbb{R},$$

behave exactly the same as the complex numbers $a + bi$ under addition and multiplication, so we can represent *all* complex numbers by 2×2 real matrices, not just the complex numbers z_θ that represent rotations. This representation offers a “linear algebra explanation” of certain properties of complex numbers, for example:

- The *squared absolute value*, $|a + bi|^2 = a^2 + b^2$ of the complex number $a + bi$ is the *determinant* of the corresponding matrix $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$.
- Therefore, the *multiplicative property* of absolute value, $|z_1 z_2| = |z_1| |z_2|$, follows from the multiplicative property of determinants,

$$\det(A_1 A_2) = \det(A_1) \det(A_2).$$

(Take A_1 as the matrix representing z_1 , and A_2 as the matrix representing z_2 .)

- The *inverse* $z^{-1} = \frac{a-bi}{a^2+b^2}$ of $z = a + bi \neq 0$ corresponds to the *inverse matrix*

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}^{-1} = \frac{1}{a^2 + b^2} \begin{pmatrix} a & b \\ -b & a \end{pmatrix}.$$

The two-square identity

If we set $z_1 = a_1 + ib_1$ and $z_2 = a_2 + ib_2$, then the multiplicative property of (squared) absolute value states that

$$(a_1^2 + b_1^2)(a_2^2 + b_2^2) = (a_1 a_2 - b_1 b_2)^2 + (a_1 b_2 + a_2 b_1)^2,$$

as can be checked by working out the product $z_1 z_2$ and its squared absolute value. This identity is particularly interesting in the case of *integers* a_1, b_1, a_2, b_2 , because it says that

$$(\text{a sum of two squares}) \times (\text{a sum of two squares}) = (\text{a sum of two squares}).$$

This fact was noticed nearly 2000 years ago by Diophantus, who mentioned an instance of it in Book III, Problem 19, of his *Arithmetica*. However, Diophantus said nothing about sums of three squares—with good reason, because there is no such three-square identity. For example

$$(1^2 + 1^2 + 1^2)(0^2 + 1^2 + 2^2) = 3 \times 5 = 15,$$

and 15 is *not* a sum of three integer squares.

Quaternions

By associating the ordered pair (a, b) with the complex number $a + ib$ or the matrix $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ we can speak of the “sum,” “product,” and “absolute value” of ordered pairs. In the same way, we can speak of the “sum,” “product,” and “absolute value” of ordered quadruples by associating each ordered quadruple (a, b, c, d) of real numbers with the matrix

$$q = \begin{pmatrix} a + id & -b - ic \\ b - ic & a - id \end{pmatrix}. \quad (*)$$

We call any matrix of the form (*) a *quaternion*. (This is not the only way to associate a matrix with a quadruple. I have chosen these complex

matrices because they extend the real matrices used in the previous section to represent complex numbers. Thus complex numbers are the special quaternions with $c = d = 0$.)

It is clear that the sum of any two matrices of the form (*) is another matrix of the same form, and it can be checked (Exercise 1.3.2) that the product of two matrices of the form (*) is of the form (*). Thus we can define the *sum* and *product* of quaternions to be just the matrix sum and product. Also, if the *squared absolute value* $|q|^2$ of a quaternion q is defined to be the determinant of q , then we have

$$\det q = \det \begin{pmatrix} a+id & -b-ic \\ b-ic & a-id \end{pmatrix} = a^2 + b^2 + c^2 + d^2.$$

So $|q|^2$ is the squared distance of the point (a, b, c, d) from O in \mathbb{R}^4 .

The quaternion sum operation has the same basic properties as addition for numbers, namely

$$\begin{aligned} q_1 + q_2 &= q_2 + q_1, && \text{(commutative law)} \\ q_1 + (q_2 + q_3) &= (q_1 + q_2) + q_3, && \text{(associative law)} \\ q + (-q) &= \mathbf{0} \quad \text{where } \mathbf{0} \text{ is the zero matrix,} && \text{(inverse law)} \\ q + \mathbf{0} &= q. && \text{(identity law)} \end{aligned}$$

The quaternion product operation does *not* have all the properties of multiplication of numbers—in general, the commutative property $q_1 q_2 = q_2 q_1$ fails—but well-known properties of the matrix product imply the following properties of the quaternion product:

$$\begin{aligned} q_1(q_2 q_3) &= (q_1 q_2) q_3, && \text{(associative law)} \\ q q^{-1} &= \mathbf{1} \quad \text{for } q \neq \mathbf{0}, && \text{(inverse law)} \\ q \mathbf{1} &= q, && \text{(identity law)} \\ q_1(q_2 + q_3) &= q_1 q_2 + q_1 q_3. && \text{(left distributive law)} \end{aligned}$$

Here $\mathbf{0}$ and $\mathbf{1}$ denote the 2×2 zero and identity matrices, which are also quaternions. The right distributive law $(q_2 + q_3) q_1 = q_2 q_1 + q_3 q_1$ of course holds too, and is distinct from the left distributive law because of the non-commutative product.

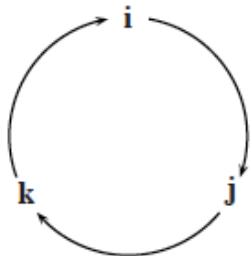
The noncommutative nature of the quaternion product is exposed more clearly when we write

$$\begin{pmatrix} a+di & -b-ci \\ b-ci & a-di \end{pmatrix} = a\mathbf{1} + b\mathbf{i} + c\mathbf{j} + d\mathbf{k},$$

where

$$\mathbf{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{i} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \mathbf{j} = \begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix}, \quad \mathbf{k} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}.$$

Thus $\mathbf{1}$ behaves like the number 1, $\mathbf{i}^2 = -\mathbf{1}$ as before, and also $\mathbf{j}^2 = \mathbf{k}^2 = -\mathbf{1}$. The noncommutativity is concentrated in the products of $\mathbf{i}, \mathbf{j}, \mathbf{k}$, which are summarized in Figure 1.2. The product of any two distinct elements is



the third element in the circle, with a + sign if an arrow points from the first element to the second, and a - sign otherwise. For example, $\mathbf{ij} = \mathbf{k}$, but $\mathbf{ji} = -\mathbf{k}$, so $\mathbf{ij} \neq \mathbf{ji}$.

The failure of the commutative law is actually a good thing, because it enables quaternions to represent other things that do not commute, such as rotations in three and four dimensions.

As with complex numbers, there is a linear algebra explanation of some less obvious properties of quaternion multiplication.

- The absolute value has the *multiplicative property* $|q_1 q_2| = |q_1||q_2|$, by the multiplicative property of \det : $\det(q_1 q_2) = \det(q_1) \det(q_2)$.
- Each nonzero quaternion q has an inverse q^{-1} , namely the matrix inverse of q .
- From the matrix (*) for q we get an explicit formula for q^{-1} . If $q = a\mathbf{l} + b\mathbf{i} + c\mathbf{j} + d\mathbf{k} \neq 0$ then

$$q^{-1} = \frac{1}{a^2 + b^2 + c^2 + d^2}(a\mathbf{l} - b\mathbf{i} - c\mathbf{j} - d\mathbf{k}).$$

- The quaternion $a\mathbf{l} - b\mathbf{i} - c\mathbf{j} - d\mathbf{k}$ is called the *quaternion conjugate* \bar{q} of $q = a\mathbf{l} + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$, and we have $q\bar{q} = a^2 + b^2 + c^2 + d^2 = |q|^2$.
- The quaternion conjugate is *not* the result of taking the complex conjugate of each entry in the matrix q . In fact, \bar{q} is the result of taking the complex conjugate of each entry in the *transposed* matrix q^T . Then it follows from $(q_1 q_2)^T = q_2^T q_1^T$ that $\overline{(q_1 q_2)} = \bar{q}_2 \bar{q}_1$.

The algebra of quaternions was discovered by Hamilton in 1843, and it is denoted by \mathbb{H} in his honor. He started with just \mathbf{i} and \mathbf{j} (hoping to find an algebra of triples analogous to the complex algebra of pairs), but later introduced $\mathbf{k} = \mathbf{ij}$ to escape from apparently intractable problems with triples (he did not know, at first, that there is no three-square identity). The matrix representation was discovered in 1858, by Cayley.

Pure imaginary quaternions

The *pure imaginary quaternions* are those of the form

$$p = b\mathbf{i} + c\mathbf{j} + d\mathbf{k}.$$

They form a three-dimensional space that we will denote by $\mathbb{R}\mathbf{i} + \mathbb{R}\mathbf{j} + \mathbb{R}\mathbf{k}$, or sometimes \mathbb{R}^3 for short. The space $\mathbb{R}\mathbf{i} + \mathbb{R}\mathbf{j} + \mathbb{R}\mathbf{k}$ is the *orthogonal complement* to the line $\mathbb{R}\mathbf{1}$ of quaternions of the form $a\mathbf{1}$, which we will call *real quaternions*. From now on we write the real quaternion $a\mathbf{1}$ simply as a , and denote the line of real quaternions simply by \mathbb{R} .

It is clear that the sum of any two members of $\mathbb{R}\mathbf{i} + \mathbb{R}\mathbf{j} + \mathbb{R}\mathbf{k}$ is itself a member of $\mathbb{R}\mathbf{i} + \mathbb{R}\mathbf{j} + \mathbb{R}\mathbf{k}$, but this is not generally true of products. In fact, if $u = u_1\mathbf{i} + u_2\mathbf{j} + u_3\mathbf{k}$ and $v = v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}$ then the multiplication diagram for \mathbf{i}, \mathbf{j} , and \mathbf{k} (Figure 1.2) gives

$$\begin{aligned} uv &= -(u_1v_1 + u_2v_2 + u_3v_3) \\ &\quad + (u_2v_3 - u_3v_2)\mathbf{i} - (u_1v_3 - u_3v_1)\mathbf{j} + (u_1v_2 - u_2v_1)\mathbf{k}. \end{aligned}$$

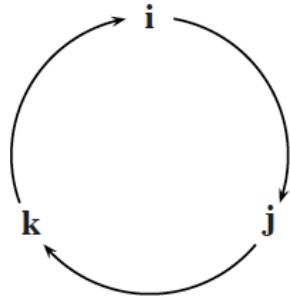


Figure 1.2: Products of the imaginary quaternion units.

This relates the quaternion product uv to two other products on \mathbb{R}^3 that are well known in linear algebra: the inner (or “scalar” or “dot”) product,

$$u \cdot v = u_1v_1 + u_2v_2 + u_3v_3,$$

and the vector (or “cross”) product

$$u \times v = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} = (u_2v_3 - u_3v_2)\mathbf{i} - (u_1v_3 - u_3v_1)\mathbf{j} + (u_1v_2 - u_2v_1)\mathbf{k}.$$

In terms of the scalar and vector products, the quaternion product is

$$uv = -u \cdot v + u \times v.$$

Since $u \cdot v$ is a real number, this formula shows that uv is in $\mathbb{R}\mathbf{i} + \mathbb{R}\mathbf{j} + \mathbb{R}\mathbf{k}$ only if $u \cdot v = 0$, that is, only if u is *orthogonal to* v .

The formula $uv = -u \cdot v + u \times v$ also shows that uv is real if and only if $u \times v = 0$, that is, if u and v have the same (or opposite) direction. In particular, if $u \in \mathbb{R}\mathbf{i} + \mathbb{R}\mathbf{j} + \mathbb{R}\mathbf{k}$ and $|u| = 1$ then

$$u^2 = -u \cdot u = -|u|^2 = -1.$$

Thus every unit vector in $\mathbb{R}\mathbf{i} + \mathbb{R}\mathbf{j} + \mathbb{R}\mathbf{k}$ is a “square root of -1 . ” (This, by the way, is another sign that \mathbb{H} does not satisfy all the usual laws of algebra. If it did, the equation $u^2 = -1$ would have at most two solutions.)

The 3-sphere of unit quaternions

The quaternions $a\mathbf{l} + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$ of absolute value 1, or *unit quaternions*, satisfy the equation

$$a^2 + b^2 + c^2 + d^2 = 1.$$

Hence they form the analogue of the sphere, called the *3-sphere* S^3 , in the space \mathbb{R}^4 of all 4-tuples (a, b, c, d) . It follows from the multiplicative property and the formula for inverses above that the product of unit quaternions is again a unit quaternion, and hence S^3 is a group under quaternion multiplication. Like the 1-sphere S^1 of unit complex numbers, the 3-sphere of unit quaternions encapsulates a group of rotations, though not quite so directly. In the next two sections we show how unit quaternions may be used to represent rotations of ordinary space \mathbb{R}^3 .

Lemma:

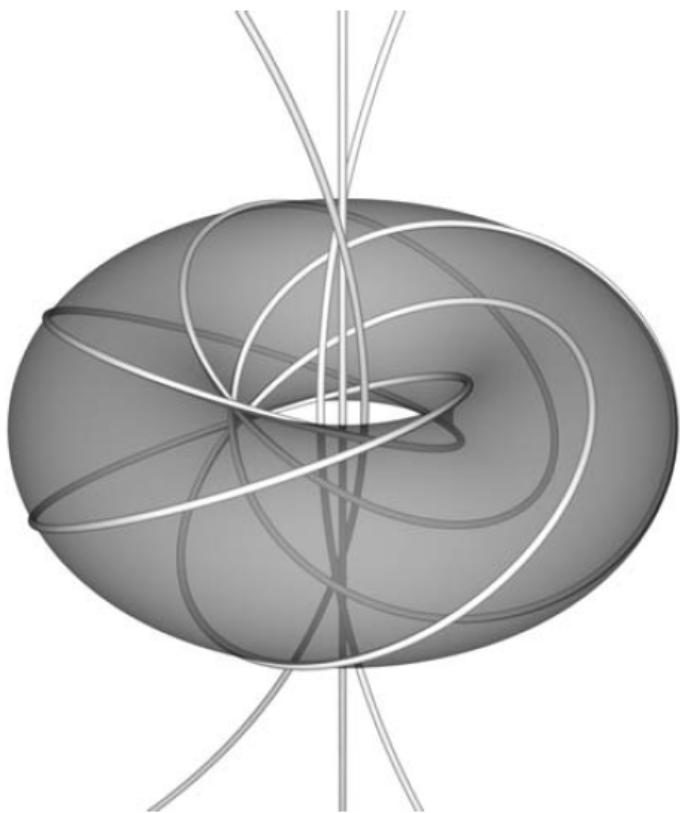
S^3 can be decomposed into disjoint congruent circles.

Proof. As we saw in Section 1.3, the quaternions $a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$ of unit length satisfy

$$a^2 + b^2 + c^2 + d^2 = 1,$$

and hence they form a 3-sphere S^3 . The unit quaternions also form a group G , because the product and inverse of unit quaternions are also unit quaternions, by the multiplicative property of absolute value.

One subgroup H of G consists of the unit quaternions of the form $\cos \theta + \mathbf{i} \sin \theta$, and these form a unit circle in the plane spanned by \mathbf{l} and \mathbf{i} . It follows that any coset qH is also a unit circle, because multiplication by a quaternion q of unit length is an isometry, as we saw in Section 1.4. Since the cosets qH fill the whole group and are disjoint, we have a decomposition of the 3-sphere into unit circles. \square



This construction is known as a Hopf fibration.

The \det homomorphism

An important homomorphism for real and complex matrix groups G is the *determinant* map

$$\det : G \rightarrow \mathbb{C}^\times,$$

where \mathbb{C}^\times denotes the multiplicative group of nonzero complex numbers. The determinant map is a homomorphism because \det is multiplicative— $\det(AB) = \det(A)\det(B)$ —a fact well known from linear algebra.

The kernel of \det , consisting of the matrices with determinant 1, is therefore a normal subgroup of G . Many important Lie groups arise in precisely this way, as we will see in Chapter 3.

Simple groups

A many-to-1 homomorphism of a group G maps it onto a group G' that is “simpler” than G (or, at any rate, not more complicated than G). For this reason, groups that admit no such homomorphism, other than the homomorphism sending all elements to 1, are called *simple*. Equivalently, a nontrivial group is simple if it contains no normal subgroups other than itself and the trivial group.

Linear/Matrix Groups

Let's now talk about more interesting continuous groups. We begin with the group that string theorists love¹⁹.

(Definition) Special Linear Group $SL(2, \mathbb{C})$: Let $SL(2, \mathbb{C})$ denotes a set of 2×2 matrices which has determinant 1 and complex entries, with the group composition law being given by usual matrix multiplication. Each element $A \in SL(2, \mathbb{C})$ has the following form

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad a, b, c, d \in \mathbb{C}$$

Since all the matrices must have $\det(A) = 1$, this means that

$$ad - bc = 1.$$

We can then prove that $SL(2, \mathbb{C})$ is a group:

- Identity: There exists an identity e such that $Ae = eA = A \forall A$. This is just the usual identity matrix

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

- Associativity: Since matrix multiplication is associative, the group *inherits* this property $A(BC) = (AB)C$ for all $A, B, C \in SL(2, \mathbb{C})$.
- Inverse: Since the determinant of all matrices $\det(A) = 1$ and hence it is never zero, this means that an inverse A^{-1} exists for all A such that $AA^{-1} = e$.
- Closure: Using the rule of determinant multiplication $\det(AB) = \det(A)\det(B)$, and now since $\det(A) = \det(B) = 1$, $\det(AB) = 1$ and hence the group is closed.

Since a, b, c, d are all complex numbers²⁰, and it inherits the continuity from being a pair of real numbers. So the elements of $SL(2, \mathbb{C})$ are then similarly continuous and thus is a **continuous group**.

This group is called the **Special Linear Group** in 2 dimensions over the Complex Field

$SL(2, \mathbb{C})$ is a big group, and as we will see, counts the group of 2-dimensional rotations and the dihedral groups D_n as some of its subgroups. But itself is a subgroup of an even larger group called $GL(2, \mathbb{C})$, or the **General Linear Group** in 2 dimensions, which itself can be generalized to N dimensions as follows.

(Definition) General Linear Group $GL(N, \mathbb{C})$: Let $GL(N, \mathbb{C})$ be a set of $N \times N$ square matrices with non-zero determinant, and complex entries. Then together with usual matrix multiplication and the algebra of the complex field, forms a group. (It should be easy for you to prove this.)

The groups $SU(2)$ and $SO(3)$

The group $\text{SO}(2)$ of rotations of \mathbb{R}^2 about O can be viewed as a geometric object, namely the *unit circle* in the plane, as we observed in Section 1.1.

The unit circle, S^1 , is the first in the series of *unit n-spheres* S^n , the n th of which consists of the points at distance 1 from the origin in \mathbb{R}^{n+1} . Thus S^2 is the ordinary sphere, consisting of the points at distance 1 from the origin in \mathbb{R}^3 . Unfortunately (for those who would like an example of an easily visualized but nontrivial Lie group) there is no rule for multiplying points that makes S^2 a Lie group. In fact, the only other Lie group among the n -spheres is S^3 . As we saw in Section 1.3, it becomes a group when its points are viewed as unit quaternions, under the operation of quaternion multiplication.

The group S^3 of unit quaternions can also be viewed as the group of 2×2 complex matrices of the form

$$Q = \begin{pmatrix} a+di & -b-ci \\ b-ci & a-di \end{pmatrix}, \quad \text{where} \quad \det(Q) = 1,$$

because these are precisely the quaternions of absolute value 1. Such matrices are called *unitary*, and the group S^3 is also known as the *special unitary group* $\text{SU}(2)$. Unitary matrices are the complex counterpart of orthogonal matrices, and we study the analogy between the two in Chapters 3 and 4.

The group $\text{SU}(2)$ is closely related to the group $\text{SO}(3)$ of rotations of \mathbb{R}^3 . As we saw in Section 1.5, rotations of \mathbb{R}^3 correspond 1-to-1 to

the pairs $\pm t$ of antipodal unit quaternions, the rotation being induced on $\mathbb{R}\mathbf{i} + \mathbb{R}\mathbf{j} + \mathbb{R}\mathbf{k}$ by the conjugation map $q \mapsto t^{-1}qt$. Also, the group operation of $\text{SO}(3)$ corresponds to quaternion multiplication, because if one rotation is induced by conjugation by t_1 , and another by conjugation by t_2 , then conjugation by t_1t_2 induces the product rotation (first rotation followed by the second). Of course, we multiply *pairs* $\pm t$ of quaternions by the rule

$$(\pm t_1)(\pm t_2) = \pm t_1t_2.$$

We therefore identify $\text{SO}(3)$ with the group \mathbb{RP}^3 of unit quaternion pairs $\pm t$ under this product operation. The map $\varphi : \text{SU}(2) \rightarrow \text{SO}(3)$ defined by $\varphi(t) = \{\pm t\}$ is a 2-to-1 homomorphism, because the two elements t and $-t$ of $\text{SU}(2)$ go to the single pair $\pm t$ in $\text{SO}(3)$. Thus $\text{SO}(3)$ looks “simpler” than $\text{SU}(2)$ because $\text{SO}(3)$ has only one element where $\text{SU}(2)$ has two. Indeed, $\text{SO}(3)$ is “simpler” because $\text{SU}(2)$ is not simple—it has the normal subgroup $\{\pm 1\}$ —and $\text{SO}(3)$ is. We now prove this famous property of $\text{SO}(3)$ by showing that $\text{SO}(3)$ has no nontrivial normal subgroup.

(\mathbb{RP}^n stands for the real projective space of lines passing through the origin 0 in \mathbb{R}^{n+1})

The above has relied extensively on:

Quaternion representation of space rotations

A quaternion t of absolute value 1, like a complex number of absolute value 1, has a “real part” $\cos \theta$ and an “imaginary part” of absolute value $\sin \theta$, orthogonal to the real part and hence in $\mathbb{R}\mathbf{i} + \mathbb{R}\mathbf{j} + \mathbb{R}\mathbf{k}$. This means that

$$t = \cos \theta + u \sin \theta,$$

where u is a unit vector in $\mathbb{R}\mathbf{i} + \mathbb{R}\mathbf{j} + \mathbb{R}\mathbf{k}$, and hence $u^2 = -1$ by the remark at the end of the previous section.

Such a unit quaternion t induces a rotation of $\mathbb{R}\mathbf{i} + \mathbb{R}\mathbf{j} + \mathbb{R}\mathbf{k}$, though not simply by multiplication, since the product of t and a member q of $\mathbb{R}\mathbf{i} + \mathbb{R}\mathbf{j} + \mathbb{R}\mathbf{k}$ may not belong to $\mathbb{R}\mathbf{i} + \mathbb{R}\mathbf{j} + \mathbb{R}\mathbf{k}$. Instead, we send each $q \in \mathbb{R}\mathbf{i} + \mathbb{R}\mathbf{j} + \mathbb{R}\mathbf{k}$ to $t^{-1}qt$, which turns out to be a member of $\mathbb{R}\mathbf{i} + \mathbb{R}\mathbf{j} + \mathbb{R}\mathbf{k}$.

To see why, first note that

$$t^{-1} = \bar{t}/|t|^2 = \cos \theta - u \sin \theta,$$

by the formulas for q^{-1} and \bar{q}

Since t^{-1} exists, multiplication of \mathbb{H} on either side by t or t^{-1} is an invertible map and hence a bijection of \mathbb{H} onto itself. It follows that the map $q \mapsto t^{-1}qt$, called *conjugation by t* , is a bijection of \mathbb{H} . Conjugation by t also maps the real line \mathbb{R} onto itself, because $t^{-1}rt = r$ for a real number r ; hence it also maps the orthogonal complement $\mathbb{R}\mathbf{i} + \mathbb{R}\mathbf{j} + \mathbb{R}\mathbf{k}$ onto itself. This is because conjugation by t is an isometry, since multiplication on either side by a unit quaternion is an isometry.

It looks as though we are onto something with conjugation by $t = \cos \theta + u \sin \theta$, and indeed we have the following theorem.

Rotation by conjugation. *If $t = \cos \theta + u \sin \theta$, where $u \in \mathbb{R}\mathbf{i} + \mathbb{R}\mathbf{j} + \mathbb{R}\mathbf{k}$ is a unit vector, then conjugation by t rotates $\mathbb{R}\mathbf{i} + \mathbb{R}\mathbf{j} + \mathbb{R}\mathbf{k}$ through angle 2θ about axis u .*

First, observe that the line $\mathbb{R}u$ of real multiples of u is fixed by the conjugation map, because

$$\begin{aligned} t^{-1}ut &= (\cos \theta - u \sin \theta)u(\cos \theta + u \sin \theta) \\ &= (u \cos \theta - u^2 \sin \theta)(\cos \theta + u \sin \theta) \\ &= (u \cos \theta + \sin \theta)(\cos \theta + u \sin \theta) \quad \text{because } u^2 = -1 \\ &= u(\cos^2 \theta + \sin^2 \theta) + \sin \theta \cos \theta + u^2 \sin \theta \cos \theta \\ &= u \quad \text{also because } u^2 = -1. \end{aligned}$$

The general proof is left to the problem sheet.

Simplicity of $\text{SO}(3)$. *The only nontrivial subgroup of $\text{SO}(3)$ closed under conjugation is $\text{SO}(3)$ itself.*

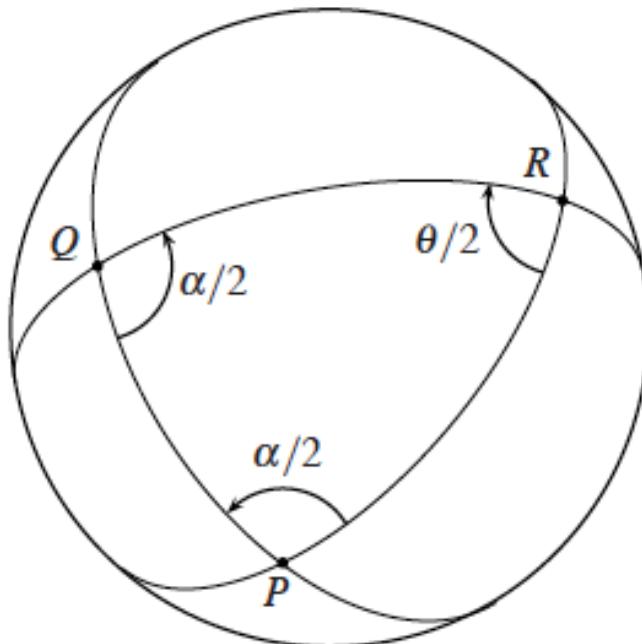
Proof. Suppose that H is a nontrivial subgroup of $\text{SO}(3)$, so H includes a nontrivial rotation, say the rotation h about axis l through angle α .

Now suppose that H is normal, so H also includes all elements $g^{-1}hg$ for $g \in \text{SO}(3)$. If g moves axis l to axis m , then $g^{-1}hg$ is the rotation about axis m through angle α . (In detail, g^{-1} moves m to l , h rotates through angle α about l , then g moves l back to m .) Thus the normal subgroup H includes the rotations through angle α about *all possible axes*.

Now a rotation through α about P , followed by rotation through α about Q , equals rotation through angle θ about R , where R and θ are as shown in Figure 2.3. As in Exercise 1.5.6, we obtain the rotation about P by successive reflections in the great circles PR and PQ , and then the rotation about Q by successive reflections in the great circles PQ and QR . In this sequence of four reflections, the reflections in PQ cancel out, leaving the reflections in PR and QR that define the rotation about R .

As P varies continuously over some interval of the great circle through P and Q , θ varies continuously over some interval. (R may also vary, but this does not matter.) It follows that θ takes some value of the form

$$\frac{m\pi}{n}, \quad \text{where } m \text{ is odd,}$$



Angle of the product rotation.

because such numbers are dense in \mathbb{R} . The n -fold product of this rotation also belongs to H , and it is a rotation about R through $m\pi$, where m is odd. The latter rotation is simply rotation through π , so H includes rotations through π about any point on the sphere (by conjugation with a suitable g again).

Finally, taking the product of rotations with $\alpha/2 = \pi/2$ in Figure 2.3, it is clear that we can get a rotation about R through any angle θ between 0 and 2π . Hence H includes all the rotations in $\text{SO}(3)$. \square

Special Orthogonal Group $\text{SO}(N)$ and Rotations

A subgroup of $SL(N, \mathbb{C})$ is the group of rotations in N dimensions $\text{SO}(N)$:

(Definition) Special Orthogonal Group $\text{SO}(N)$: Let $\text{SO}(N)$ be a set of $N \times N$ matrices R such that $\det R = 1$, and the condition $R^T R = e$, where R^T is the **Transpose**²¹ of the matrix R . Then $\text{SO}(N)$ is a group.

Most of the proof is very similar to the case for $SL(2, \mathbb{C})$. The additional condition $R^\dagger R = 1$ requires just a bit of work to show closure, i.e. suppose $R_1, R_2 \in \text{SO}(N)$, then

$$\begin{aligned} (R_1 R_2)^T (R_1 R_2) &= R_2^T R_1^T R_1 R_2 \\ &= R_2^T (R_1^T R_1) R_2 \\ &= R_2^T R_2 \\ &= e \end{aligned}$$

so the group is closed under matrix multiplication as required. This condition $R^T R = e$ is known as the **Orthogonality condition** (just in case you are wondering). This group is a subgroup of the **Orthogonal Group $O(N)$** , which relaxes the condition on the determinant to be ± 1 (instead of just $+1$).

Rotations of \mathbb{R}^4 and pairs of quaternions

A linear map is called *orientation-preserving* if its determinant is positive, and *orientation-reversing* otherwise. Reflections are linear and orientation-reversing, so a product of reflections is orientation-preserving if and only if it contains an even number of terms. We define a *rotation of \mathbb{R}^n about O* to be an orientation-preserving isometry that fixes O .

Thus it follows from the Cartan–Dieudonné theorem that any rotation of \mathbb{R}^4 is the product of 0, 2, or 4 reflections. The exact number is not important here—what we really want is a way to represent reflections by quaternions, as a stepping-stone to the representation of rotations by quaternions. Not surprisingly, each reflection is specified by the quaternion orthogonal to the hyperplane of reflection. More surprisingly, a rotation is specified by just *two* quaternions, regardless of the number of reflections needed to compose it.

Quaternion representation of reflections. *Reflection of $\mathbb{H} = \mathbb{R}^4$ in the hyperplane through O orthogonal to the unit quaternion u is the map that sends each $q \in \mathbb{H}$ to $-u\bar{q}u$.*

Proof. First observe that the map $q \mapsto -u\bar{q}u$ is an isometry. This is because

- $q \mapsto -\bar{q}$ reverses the real part of q and keeps the imaginary part fixed, hence it is reflection in the hyperplane spanned by \mathbf{i} , \mathbf{j} , and \mathbf{k} .
- Multiplication on the left by the unit quaternion u is an isometry by the argument in Section 1.4, and there is a similar argument for multiplication on the right.

Next notice that the map $q \mapsto -u\bar{q}u$ sends

$$\begin{aligned} vu &\text{ to } -u(\overline{vu})u = -u\bar{u}\bar{v}u \quad \text{because } (\overline{vu}) = \bar{u}\bar{v}, \\ &= -\bar{v}u \quad \text{because } u\bar{u} = |u|^2 = 1. \end{aligned}$$

In particular, the map sends u to $-u$, so vectors parallel to u are reversed. And it sends iu to iu , because $\bar{\mathbf{i}} = -\mathbf{i}$, and similarly ju to ju and ku to ku . Thus the vectors iu , ju , and ku , which span the hyperplane orthogonal to u , are fixed. Hence the map $q \mapsto -u\bar{q}u$ is reflection in this hyperplane. \square

Quaternion representation of rotations. Any rotation of $\mathbb{H} = \mathbb{R}^4$ about O is a map of the form $q \mapsto vqw$, where v and w are unit quaternions.

Proof. It follows from the quaternion representation of reflections that the result of successive reflections in the hyperplanes orthogonal to the unit quaternions u_1, u_2, \dots, u_{2n} is the map

$$q \mapsto u_{2n} \cdots \overline{u_3} u_2 \overline{u_1} q \overline{u_1} u_2 \overline{u_3} \cdots u_{2n},$$

because an even number of sign changes and conjugations makes no change. The pre- and postmultipliers are in general two different unit quaternions, $u_{2n} \cdots \overline{u_3} u_2 \overline{u_1} = v$ and $\overline{u_1} u_2 \overline{u_3} \cdots u_{2n} = w$, say, so the general rotation of \mathbb{R}^4 is a map of the form

$$q \mapsto vqw, \quad \text{where } v \text{ and } w \text{ are unit quaternions.}$$

Conversely, any map of this form is a rotation, because *multiplication of $\mathbb{H} = \mathbb{R}^4$ on either side by a unit quaternion is an orientation-preserving isometry*. We already know that multiplication by a unit quaternion is an isometry, by Section 1.4. And it preserves orientation by the following argument.

Multiplication of $\mathbb{H} = \mathbb{R}^4$ by a unit quaternion

$$v = \begin{pmatrix} a+id & -b-ic \\ b-ic & a-id \end{pmatrix}, \quad \text{where } a^2 + b^2 + c^2 + d^2 = 1,$$

is a linear transformation of \mathbb{R}^4 with matrix

$$R_v = \left(\begin{array}{cc|cc} a & -d & -b & c \\ d & a & -c & -b \\ \hline b & c & a & d \\ -c & b & -d & a \end{array} \right),$$

where the 2×2 submatrices represent the complex-number entries in v . It can be checked that $\det(R_v) = 1$. So multiplication by v , on either side, preserves orientation. \square

An $n \times n$ real matrix A is said to be **orthogonal** if the column vectors that make up A are orthonormal, that is, if

$$\sum_{l=1}^n A_{lj} A_{lk} = \delta_{jk}, \quad 1 \leq j, k \leq n.$$

(Here δ_{jk} is the Kronecker delta, equal to 1 if $j = k$ and equal to zero if $j \neq k$.) Equivalently, A is orthogonal if it preserves the inner product, namely if $\langle x, y \rangle = \langle Ax, Ay \rangle$ for all vectors x, y in \mathbb{R}^n . (Angled brackets denote the usual inner product on \mathbb{R}^n , $\langle x, y \rangle = \sum_k x_k y_k$.) Still another equivalent definition is that A is orthogonal if $A^{tr}A = I$, i.e., if $A^{tr} = A^{-1}$. (Here, A^{tr} is the transpose of A , $(A^{tr})_{kl} = A_{lk}$.) See Exercise 2.

Since $\det A^{tr} = \det A$, we see that if A is orthogonal, then $\det(A^{tr}A) = (\det A)^2 = \det I = 1$. Hence, $\det A = \pm 1$, for all orthogonal matrices A .

This formula tells us in particular that every orthogonal matrix must be invertible. However, if A is an orthogonal matrix, then

$$\langle A^{-1}x, A^{-1}y \rangle = \langle A(A^{-1}x), A(A^{-1}y) \rangle = \langle x, y \rangle.$$

Thus, the inverse of an orthogonal matrix is orthogonal. Furthermore, the product of two orthogonal matrices is orthogonal, since if A and B both preserve inner products, then so does AB . Thus, the set of orthogonal matrices forms a group.

The set of all $n \times n$ real orthogonal matrices is the **orthogonal group** $O(n)$, and it is a subgroup of $GL(n; \mathbb{C})$. The limit of a sequence of orthogonal matrices is orthogonal, because the relation $A^{tr}A = I$ is preserved under taking limits. Thus, $O(n)$ is a matrix Lie group.

If $\mathbf{u} = (u_1, u_2, \dots, u_n)$ and $\mathbf{v} = (v_1, v_2, \dots, v_n)$ are two vectors in \mathbb{R}^n , their *inner product* $\mathbf{u} \cdot \mathbf{v}$ is defined by

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.$$

It follows immediately that

$$\mathbf{u} \cdot \mathbf{u} = u_1^2 + u_2^2 + \cdots + u_n^2 = |\mathbf{u}|^2,$$

so the *length* $|\mathbf{u}|$ of \mathbf{u} (that is, the distance of \mathbf{u} from the origin $\mathbf{0}$) is definable in terms of the inner product. It also follows (as one learns in linear algebra courses) that $\mathbf{u} \cdot \mathbf{v} = 0$ if and only if \mathbf{u} and \mathbf{v} are orthogonal, and more generally that

$$\mathbf{u} \cdot \mathbf{v} = |\mathbf{u}| |\mathbf{v}| \cos \theta,$$

where θ is the angle between the lines from $\mathbf{0}$ to \mathbf{u} and $\mathbf{0}$ to \mathbf{v} . Thus angle is also definable in terms of inner product. Conversely, inner product is definable in terms of length and angle. Moreover, an angle θ is determined by $\cos \theta$ and $\sin \theta$, which are the ratios of lengths in a certain triangle, so inner product is in fact definable in terms of length alone.

This means that a transformation T preserves length if and only if T preserves the inner product, that is,

$$T(\mathbf{u}) \cdot T(\mathbf{v}) = \mathbf{u} \cdot \mathbf{v} \quad \text{for all } \mathbf{u}, \mathbf{v} \in \mathbb{R}^n.$$

The inner product is a more convenient concept than length when one is working with linear transformations, because linear transformations are represented by matrices and the inner product occurs naturally within matrix multiplication: if A and B are matrices for which AB exists then

$$(i, j)\text{-element of } AB = (\text{row } i \text{ of } A) \cdot (\text{column } j \text{ of } B).$$

This observation is the key to the following concise and practical criterion for recognizing rotations, involving the matrix A and its transpose A^T . To state it we introduce the notation $\mathbf{1}$ for the identity matrix, of any size, extending the notation used in Chapter 1 for the 2×2 identity matrix.

Rotation criterion. *An $n \times n$ real matrix A represents a rotation of \mathbb{R}^n if and only if*

$$AA^T = \mathbf{1} \quad \text{and} \quad \det(A) = 1.$$

Proof. First we show that the condition $AA^T = \mathbf{1}$ is equivalent to preservation of the inner product by A .

$$\begin{aligned} AA^T = \mathbf{1} &\Leftrightarrow (\text{row } i \text{ of } A) \cdot (\text{col } j \text{ of } A^T) = \delta_{ij} \\ &\quad \text{where } \delta_{ij} = 1 \text{ if } i = j \text{ and } \delta_{ij} = 0 \text{ if } i \neq j \\ &\Leftrightarrow (\text{row } i \text{ of } A) \cdot (\text{row } j \text{ of } A) = \delta_{ij} \\ &\Leftrightarrow \text{rows of } A \text{ form an orthonormal basis} \\ &\Leftrightarrow \text{columns of } A \text{ form an orthonormal basis} \\ &\quad \text{because } AA^T = \mathbf{1} \text{ means } A^T = A^{-1}, \text{ so } \mathbf{1} = A^T A = A^T (A^T)^T, \\ &\quad \text{and hence } A^T \text{ has the same property as } A \\ &\Leftrightarrow A\text{-images of the standard basis form an orthonormal basis} \\ &\Leftrightarrow A \text{ preserves the inner product} \end{aligned}$$

because $A\mathbf{e}_i \cdot A\mathbf{e}_j = \mathbf{e}_i \cdot \mathbf{e}_j$, where $\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$, ..., $\mathbf{e}_n = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$ are the standard basis vectors of \mathbb{R}^n .

Second, the condition $\det(A) = 1$ says that A preserves orientation, as mentioned at the beginning of Section 2.5. Standard properties of determinants give

$$\det(AA^T) = \det(A)\det(A^T) \quad \text{and} \quad \det(A^T) = \det(A),$$

so we already have

$$1 = \det(\mathbf{1}) = \det(AA^T) = \det(A)\det(A^T) = \det(A)^2.$$

And the two solutions $\det(A) = 1$ and $\det(A) = -1$ occur according as A preserves orientation or not. \square

A rotation matrix is called a *special orthogonal* matrix, presumably because its rows (or columns) form an orthonormal basis. The matrices

that preserve length, but not necessarily orientation, are called *orthogonal*. (However, orthogonal matrices are not the only matrices that preserve orthogonality. Orthogonality is also preserved by the *dilation* matrices $k\mathbf{1}$ for any nonzero constant k .)

The general linear groups are indeed groups under the operation of matrix multiplication: The product of two invertible matrices is invertible, the identity matrix is an identity for the group, an invertible matrix has (by definition) an inverse, and matrix multiplication is associative.

An example of a subgroup of $\mathrm{GL}(n; \mathbb{C})$ which is not closed (and hence is not a matrix Lie group) is the set of all $n \times n$ invertible matrices all of whose entries are real and rational. This is in fact a subgroup of $\mathrm{GL}(n; \mathbb{C})$, but not a closed subgroup. That is, one can (easily) have a sequence of invertible matrices with rational entries converging to an invertible matrix with some irrational entries. (In fact, *every* real invertible matrix is the limit of some sequence of invertible matrices with rational entries.)

Another example of a group of matrices which is not a matrix Lie group is the following subgroup of $\mathrm{GL}(2; \mathbb{C})$. Let a be an irrational real number and let

$$G = \left\{ \begin{pmatrix} e^{it} & 0 \\ 0 & e^{ita} \end{pmatrix} \mid t \in \mathbb{R} \right\}.$$

Clearly, G is a subgroup of $\mathrm{GL}(2; \mathbb{C})$. Because a is irrational, the matrix $-I$ is not in G , since to make e^{it} equal to -1 , we must take t to be an odd integer multiple of π , in which case ta cannot be an odd integer multiple of π . On the other hand (Exercise 1), by taking $t = (2n+1)\pi$ for a suitably chosen integer n , we can make ta arbitrarily close to an odd integer multiple of π . Hence, we can find a sequence of matrices in G which converges to $-I$, and so G is not a matrix Lie group.

The unitary and special unitary groups, $\mathbf{U}(n)$ and $\mathbf{SU}(n)$

An $n \times n$ complex matrix A is said to be **unitary** if the column vectors of A are orthonormal, that is,

$$\sum_{l=1}^n \overline{A_{lj}} A_{lk} = \delta_{jk}.$$

Equivalently, A is unitary if it preserves the inner product, namely if $\langle x, y \rangle = \langle Ax, Ay \rangle$ for all vectors x, y in \mathbb{C}^n . (Angled brackets here denote the inner product on \mathbb{C}^n , $\langle x, y \rangle = \sum_k \bar{x}_k y_k$. We will adopt the convention of putting the complex conjugate on the left.) Still another equivalent definition is that A is unitary if $A^* A = I$, i.e., if $A^* = A^{-1}$. (Here, A^* is the **adjoint** of A , $(A^*)_{jk} = \overline{A_{kj}}$.)

Since $\det A^* = \overline{\det A}$, we see that if A is unitary, then $\det(A^* A) = |\det A|^2 = \det I = 1$. Hence, $|\det A| = 1$, for all unitary matrices A .

This, in particular, shows that every unitary matrix is invertible. The same argument as for the orthogonal group shows that the set of unitary matrices forms a group.

The set of all $n \times n$ unitary matrices is the **unitary group** $U(n)$, and it is a subgroup of $GL(n; \mathbb{C})$. The limit of unitary matrices is unitary, so $U(n)$ is a matrix Lie group. The set of unitary matrices with determinant one is the **special unitary group** $SU(n)$. It is easy to check that $SU(n)$ is a matrix Lie group. Note that a unitary matrix can have determinant $e^{i\theta}$ for any θ , and so $SU(n)$ is a smaller subset of $U(n)$ than $SO(n)$ is of $O(n)$. (Specifically, $SO(n)$ has the same dimension as $O(n)$, whereas $SU(n)$ has dimension one less than that of $U(n)$.)

The generalized orthogonal and Lorentz groups

Let n and k be positive integers, and consider \mathbb{R}^{n+k} . Define a symmetric bilinear form $[\cdot, \cdot]_{n,k}$ on \mathbb{R}^{n+k} by the formula

$$[x, y]_{n,k} = x_1 y_1 + \cdots + x_n y_n - x_{n+1} y_{n+1} - \cdots - x_{n+k} y_{n+k}$$

The set of $(n+k) \times (n+k)$ real matrices A which preserve this form (i.e., such that $[Ax, Ay]_{n,k} = [x, y]_{n,k}$ for all $x, y \in \mathbb{R}^{n+k}$) is the **generalized orthogonal group** $O(n; k)$. It is a subgroup of $GL(n+k; \mathbb{R})$ and a matrix Lie group

If A is an $(n+k) \times (n+k)$ real matrix, let $A^{(i)}$ denote the i^{th} column vector of A , that is,

$$A^{(i)} = \begin{pmatrix} A_{1,i} \\ \vdots \\ A_{n+k,i} \end{pmatrix}.$$

Then, A is in $O(n; k)$ if and only if the following conditions are satisfied:

$$\begin{aligned} [A^{(l)}, A^{(j)}]_{n,k} &= 0 & l \neq j, \\ [A^{(l)}, A^{(l)}]_{n,k} &= 1 & 1 \leq l \leq n, \\ [A^{(l)}, A^{(l)}]_{n,k} &= -1 & n+1 \leq l \leq n+k. \end{aligned}$$

Let g denote the $(n+k) \times (n+k)$ diagonal matrix with ones in the first n diagonal entries and minus ones in the last k diagonal entries. Then, A is in $O(n; k)$ if and only if $A^{tr} g A = g$ (Exercise 4). Taking the determinant of this equation gives $(\det A)^2 \det g = \det g$, or $(\det A)^2 = 1$. Thus, for any A in $O(n; k)$, $\det A = \pm 1$.

Of particular interest in physics is the **Lorentz group** $O(3; 1)$.

The groups \mathbb{R}^* , \mathbb{C}^* , S^1 , \mathbb{R} , and \mathbb{R}^n

Several important groups which are not naturally groups of matrices can (and will in these notes) be thought of as such.

The group \mathbb{R}^* of non-zero real numbers under multiplication is isomorphic to $GL(1; \mathbb{R})$. Thus, we will regard \mathbb{R}^* as a matrix Lie group. Similarly, the group \mathbb{C}^* of nonzero complex numbers under multiplication is isomorphic to $GL(1; \mathbb{C})$, and the group S^1 of complex numbers with absolute value one is isomorphic to $U(1)$.

The group \mathbb{R} under addition is isomorphic to $GL(1; \mathbb{R})^+$ (1×1 real matrices with positive determinant) via the map $x \rightarrow [e^x]$. The group \mathbb{R}^n (with vector addition) is isomorphic to the group of diagonal real matrices with positive diagonal entries, via the map

$$(x_1, \dots, x_n) \rightarrow \begin{pmatrix} e^{x_1} & & 0 \\ & \ddots & \\ 0 & & e^{x_n} \end{pmatrix}.$$

The Euclidean and Poincaré groups $E(n)$ and $P(n; 1)$

The **Euclidean group** $E(n)$ is, by definition, the group of all one-to-one, onto, distance-preserving maps of \mathbb{R}^n to itself, that is, maps $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ such that $d(f(x), f(y)) = d(x, y)$ for all $x, y \in \mathbb{R}^n$. Here, d is the usual distance on \mathbb{R}^n : $d(x, y) = |x - y|$. Note that we do not assume *anything* about the structure of f besides the above properties. In particular, f need not be linear. The orthogonal group $O(n)$ is a subgroup of $E(n)$ and is the group of all *linear* distance-preserving maps of \mathbb{R}^n to itself. For $x \in \mathbb{R}^n$, define the **translation by x** , denoted T_x , by

$$T_x(y) = x + y.$$

The set of translations is also a subgroup of $E(n)$.

Every element T of $E(n)$ can be written uniquely as an orthogonal linear transformation followed by a translation, that is, in the form

$$T = T_x R$$

with $x \in \mathbb{R}^n$ and $R \in O(n)$.

We will not prove this. The key step is to prove that every one-to-one, onto, distance-preserving map of \mathbb{R}^n to itself which fixes the origin must be linear. We will write an element $T = T_x R$ of $E(n)$ as a pair $\{x, R\}$. Note that for $y \in \mathbb{R}^n$,

$$\{x, R\} y = Ry + x$$

and that

$$\{x_1, R_1\} \{x_2, R_2\} y = R_1(R_2y + x_2) + x_1 = R_1R_2y + (x_1 + R_1x_2).$$

Thus, the product operation for $E(n)$ is the following:

$$\{x_1, R_1\} \{x_2, R_2\} = \{x_1 + R_1x_2, R_1R_2\}.$$

The inverse of an element of $E(n)$ is given by

$$\{x, R\}^{-1} = \{-R^{-1}x, R^{-1}\}.$$

As already noted, $E(n)$ is not a subgroup of $GL(n; \mathbb{R})$, since translations are not linear maps. However, $E(n)$ is isomorphic to a subgroup of $GL(n+1; \mathbb{R})$, via the map which associates to $\{x, R\} \in E(n)$ the following matrix:

$$\begin{pmatrix} & & x_1 \\ & R & \vdots \\ & & x_n \\ 0 & \cdots & 0 & 1 \end{pmatrix}.$$

This map is clearly one-to-one, and direct computation shows that multiplication of elements of the form (1.6) follows the multiplication rule in (1.5), so that this map is a homomorphism. Thus, $E(n)$ is isomorphic to the group of all matrices of the form (1.6) (with $R \in O(n)$). The limit of things of the form (1.6) is again of that form, and so we have expressed the Euclidean group $E(n)$ as a matrix Lie group.

We similarly define the Poincaré group $P(n; 1)$ to be the group of all transformations of \mathbb{R}^{n+1} of the form

$$T = T_x A$$

with $x \in \mathbb{R}^{n+1}$ and $A \in O(n; 1)$. This is the group of affine transformations of \mathbb{R}^{n+1} which preserve the Lorentz “distance” $d_L(x, y) = (x_1 - y_1)^2 + \cdots + (x_n - y_n)^2 - (x_{n+1} - y_{n+1})^2$. (An affine transformation is one of the form $x \rightarrow Ax + b$, where A is a linear transformation and b is constant.) The group product is the obvious analog of the product (1.5) for the Euclidean group.

The Poincaré group $P(n; 1)$ is isomorphic to the group of $(n+2) \times (n+2)$ matrices of the form

$$\begin{pmatrix} & x_1 \\ A & \vdots \\ & x_{n+1} \\ 0 \cdots 0 & 1 \end{pmatrix}$$

The map from $SU(2) \times SU(2)$ to $SO(4)$

we showed that the rotations of \mathbb{R}^4 are precisely the maps $q \mapsto vqw$, where v and w run through all the unit quaternions. Since v^{-1} is a unit quaternion if and only if v is, it is equally valid to represent each rotation of \mathbb{R}^4 by a map of the form $q \mapsto v^{-1}qw$, where v and w are unit quaternions. The latter representation is more convenient for what comes next.

The pairs of unit quaternions (v, w) form a group under the operation defined by

$$(v_1, w_1) \cdot (v_2, w_2) = (v_1 v_2, w_1 w_2),$$

where the products $v_1 v_2$ and $w_1 w_2$ on the right side are ordinary quaternion products. Since the v come from the group $SU(2)$ of unit quaternions, and the w likewise, the group of pairs (v, w) is the direct product $SU(2) \times SU(2)$ of $SU(2)$ with itself.

The map that sends each pair $(v, w) \in SU(2) \times SU(2)$ to the rotation $q \mapsto v^{-1}qw$ in $SO(4)$ is a *homomorphism* $\varphi : SU(2) \times SU(2) \rightarrow SO(4)$. This is because

- the product of the map $q \mapsto v_1^{-1}qw_1$ corresponding to (v_1, w_1)
- with the map $q \mapsto v_2^{-1}qw_2$ corresponding to (v_2, w_2)
- is the map $q \mapsto v_2^{-1}v_1^{-1}qw_1w_2$,
- which is the map $q \mapsto (v_1v_2)^{-1}q(w_1w_2)$ corresponding to the product (v_1v_2, w_1w_2) of (v_1, w_1) and (v_2, w_2) .

This homomorphism is onto $\mathrm{SO}(4)$, because each rotation of \mathbb{R}^4 can be expressed in the form $q \mapsto v^{-1}qw$, but one might expect it to be very many-to-one, since many pairs (v, w) of unit quaternions conceivably give the same rotation. Surprisingly, this is not so. The representation of rotations by pairs is “unique up to sign” in the following sense: *if (v, w) gives a certain rotation, the only other pair that gives the same rotation is $(-v, -w)$.*

To prove this, it suffices to prove that the kernel of the homomorphism $\varphi : \mathrm{SU}(2) \times \mathrm{SU}(2) \rightarrow \mathrm{SO}(4)$ has two elements.

Size of the kernel. *The homomorphism $\varphi : \mathrm{SU}(2) \times \mathrm{SU}(2) \rightarrow \mathrm{SO}(4)$ is 2-to-1, because its kernel has two elements.*

SO(4) is not simple. *There is a nontrivial normal subgroup of SO(4), not equal to SO(4).*

Proof. The subgroup of pairs $(v, 1) \in \mathrm{SU}(2) \times \mathrm{SU}(2)$ is normal; in fact, it is the kernel of the map $(v, w) \mapsto (1, w)$, which is clearly a homomorphism.

The corresponding subgroup of SO(4) consists of maps of the form $q \mapsto v^{-1}qv$, which likewise form a normal subgroup of SO(4). But this subgroup is not the whole of SO(4). For example, it does not include the map $q \mapsto qw$ for any $w \neq \pm 1$, by the “unique up to sign” representation of rotations by pairs (v, w) . \square

Proof. Suppose that (v, w) is in the kernel, so $q \mapsto v^{-1}qw$ is the identity rotation. In particular, this rotation fixes 1, so

$$v^{-1}1w = 1; \quad \text{hence} \quad v = w.$$

Thus the map is in fact $q \mapsto v^{-1}qv$, which we know (from Section 1.5) fixes the real axis and rotates the space of pure imaginary quaternions. Only if $v = 1$ or $v = -1$ does the map fix everything; hence the kernel of φ has only two elements, $(1, 1)$ and $(-1, -1)$.

The left cosets of the kernel are therefore the 2-element sets

$$(v, w)(\pm 1, \pm 1) = (\pm v, \pm w),$$

and each coset corresponds to a distinct rotation of \mathbb{R}^4 , by the fundamental homomorphism theorem of Section 2.2. \square

This theorem shows that SO(4) is “almost” the same as $\mathrm{SU}(2) \times \mathrm{SU}(2)$, and the latter is far from being a simple group. For example, the subgroup of pairs $(v, 1)$ is a nontrivial normal subgroup, but clearly not the whole of $\mathrm{SU}(2) \times \mathrm{SU}(2)$. This gives us a way to show that SO(4) is not simple.

We shall next give an introduction to the representation theory of finite groups. We shall omit the proofs of some of the important theorems since our main emphasis will be applications. Within solid state physics the representation theory is used widely.

Representations of finite groups

We will concentrate on matrix representations of *finite* groups, particularly rotation and reflection groups (the so-called crystal point groups). The general ideas carry over to infinite groups, such as the continuous rotation groups, but in a book such as this, which aims to cover many areas of applicable mathematics, some topics can only be mentioned and not explored. We now give the formal definition of a representation.

Definition. A representation $D = \{D(X)\}$ of a group \mathcal{G} is an assignment of a non-singular square $n \times n$ matrix $D(X)$ to each element X belonging to \mathcal{G} , such that

- (i) $D(I) = I_n$, the unit $n \times n$ matrix,
- (ii) $D(X)D(Y) = D(XY)$ for any two elements X and Y belonging to \mathcal{G} , i.e. the matrices multiply in the same way as the group elements they represent.

As mentioned previously, a representation by $n \times n$ matrices is said to be an *n-dimensional representation* of \mathcal{G} . The dimension n is not to be confused with g , the order of the group, which gives the number of matrices needed in the representation, though they might not all be different.

A consequence of the two defining conditions for a representation is that the matrix associated with the inverse of X is the inverse of the matrix associated with X . This follows immediately from setting $Y = X^{-1}$ in (ii):

$$D(X)D(X^{-1}) = D(XX^{-1}) = D(I) = I_n;$$

hence

$$D(X^{-1}) = [D(X)]^{-1}.$$

Equivalent representations

If D is an *n*-dimensional representation of a group \mathcal{G} , and Q is any fixed invertible $n \times n$ matrix ($|Q| \neq 0$), then the set of matrices defined by the similarity transformation

$$D_Q(X) = Q^{-1}D(X)Q$$

also forms a representation D_Q of \mathcal{G} , said to be *equivalent* to D . We can see from a comparison with the definition in section 29.2 that they do form a representation:

- (i) $D_Q(I) = Q^{-1}D(I)Q = Q^{-1}I_nQ = I_n$,
- (ii) $D_Q(X)D_Q(Y) = Q^{-1}D(X)QQ^{-1}D(Y)Q = Q^{-1}D(X)D(Y)Q$
 $= Q^{-1}D(XY)Q = D_Q(XY)$.

Since we can always transform between equivalent representations using a non-singular matrix Q , we will consider such representations to be one and the same.

The general invertible matrix Q that appears in the definition

describes changes arising from a change in the coordinate system

(i.e. in the set of basis functions). As before, suppose that the effect of an operation X on the basis functions is expressed by the action of $M(X)$ (which is equal to $D^T(X)$) on the corresponding basis vector:

$$u' = M(X)u = D^T(X)u.$$

A change of basis would be given by $u_Q = Qu$ and $u'_Q = Qu'$, and we may write

$$u'_Q = Qu' = QM(X)u = QD^T(X)Q^{-1}u_Q.$$

This is of the same form as (29.6), i.e.

$$u'_Q = D_{Q^T}(X)u_Q,$$

where $D_{Q^T}(X) = (Q^T)^{-1}D(X)Q^T$ is related to $D(X)$ by a similarity transformation. Thus $D_{Q^T}(X)$ represents the same linear transformation as $D(X)$, but with respect to a new basis vector u_Q ; this supports our contention that representations connected by similarity transformations should be considered as the *same* representation.

Reducibility of a representation

We have seen already that it is possible to have more than one representation of any particular group. For example, the group $\{1, i, -1, -i\}$ under ordinary multiplication has been shown to have a set of 2×2 matrices, and a set of four unit $n \times n$ matrices I_n , as two of its possible representations.

Consider two or more representations, $D^{(1)}, D^{(2)}, \dots, D^{(N)}$, which may be of different dimensions, of a group G . Now combine the matrices $D^{(1)}(X), D^{(2)}(X), \dots, D^{(N)}(X)$ that correspond to element X of G into a larger *block-diagonal* matrix:

$$D(X) = \begin{pmatrix} D^{(1)}(X) & & & & 0 \\ & D^{(2)}(X) & & & \\ & & \ddots & & \\ & & & D^{(N)}(X) & \\ 0 & & & & \end{pmatrix}$$

Then $D = \{D(X)\}$ is the matrix representation of the group obtained by combining the basis vectors of $D^{(1)}, D^{(2)}, \dots, D^{(N)}$ into one larger basis vector. If, knowingly or unknowingly, we had started with this larger basis vector and found the matrices of the representation D to have the form shown in (29.9), or to have a form that can be transformed into this by a similarity transformation (29.5) (using, of course, the *same* matrix Q for each of the matrices $D(X)$) then we would say that D is *reducible* and that each matrix $D(X)$ can be written as the *direct sum* of smaller representations:

$$D(X) = D^{(1)}(X) \oplus D^{(2)}(X) \oplus \cdots \oplus D^{(N)}(X).$$

It may be that some or all of the matrices $D^{(1)}(X), D^{(2)}(X), \dots, D^{(N)}$ themselves can be further reduced – i.e. written in block diagonal form. For example, suppose that the representation $D^{(1)}$, say, has a basis vector $(x \ y \ z)^T$; then, for the symmetry group of an equilateral triangle, whilst x and y are mixed together for at least one of the operations X , z is never changed. In this case the 3×3 representative matrix $D^{(1)}(X)$ can itself be written in block diagonal form as a 2×2 matrix and a 1×1 matrix. The direct-sum matrix $D(X)$ can now be written

$$D(X) = \begin{pmatrix} \boxed{\begin{matrix} a & b \\ c & d \end{matrix}} & & & 0 \\ & \boxed{1} & & \\ & & \boxed{D^{(2)}(X)} & \\ & & & \ddots \\ 0 & & & \boxed{D^{(N)}(X)} \end{pmatrix}$$

but the first two blocks can be reduced no further.

When all the other representations $D^{(2)}(X), \dots$ have been similarly treated, what remains is said to be *irreducible* and has the characteristic of being block diagonal, with blocks that individually cannot be reduced further. The blocks are known as the *irreducible representations of \mathcal{G}* , often abbreviated to the *irreps of \mathcal{G}* , and we denote them by $\hat{D}^{(i)}$. They form the building blocks of representation theory, and it is their properties that are used to analyse any given physical situation which is invariant under the operations that form the elements of \mathcal{G} . Any representation can be written as a linear combination of irreps.

If, however, the initial choice u of basis vector for the representation D is arbitrary, as it is in general, then it is unlikely that the matrices $D(X)$ will assume obviously block diagonal forms (it should be noted, though, that since the matrices are square, even a matrix with non-zero entries only in the extreme top right and bottom left positions is technically block diagonal). In general, it will be possible to reduce them to block diagonal matrices with more than one block; this reduction corresponds to a transformation Q to a new basis vector u_0 , as described in section 29.3.

In any particular representation D , each constituent irrep $\hat{D}^{(i)}$ may appear any number of times, or not at all, subject to the obvious restriction that the sum of all the irrep dimensions must add up to the dimension of D itself. Let us say that $\hat{D}^{(i)}$ appears m_i times. The general expansion of D is then written

$$D = m_1 \hat{D}^{(1)} \oplus m_2 \hat{D}^{(2)} \oplus \dots \oplus m_N \hat{D}^{(N)},$$

where if \mathcal{G} is finite so is N .

- (i) Those components of the basis vector that correspond to rows in the representation matrices with a single-entry block, i.e. a 1×1 block, are unchanged by the operations of the group. Such a coordinate or function is said to transform according to a one-dimensional irrep of \mathcal{G} . In the example given in (29.10), that the entry on the third row forms a 1×1 block implies that the third entry in the basis vector $(x \ y \ z \ \dots)^T$, namely z , is invariant under the two-dimensional symmetry operations on an equilateral triangle in the xy -plane.
- (ii) If, in any of the g matrices of the representation, the largest-sized block located on the row or column corresponding to a particular coordinate (or function) in the basis vector is $n \times n$, then that coordinate (or function) is mixed by the symmetry operations with $n - 1$ others and is said to transform according to an n -dimensional irrep of \mathcal{G} . Thus in the matrix (29.10), x is the first entry in the complete basis vector; the first row of the matrix contains two non-zero entries, as does the first column, and so x is part of a two-component basis vector whose components are mixed by the symmetry operations of \mathcal{G} . The other component is y .

The result

may also be formulated in terms of the more abstract notion

of vector spaces. The set of g matrices that forms an n -dimensional representation D of the group \mathcal{G} can be thought of as acting on column matrices corresponding to vectors in an n -dimensional vector space V spanned by the basis functions of the representation. If there exists a *proper subspace* W of V , such that if a vector whose column matrix is w belongs to W then the vector whose column matrix is $D(X)w$ also belongs to W , for all X belonging to \mathcal{G} , then it follows that D is reducible. We say that the subspace W is invariant under the actions of the elements of \mathcal{G} . With D unitary, the orthogonal complement W_{\perp} of W , i.e. the vector space V remaining when the subspace W has been removed, is also invariant, and all the matrices $D(X)$ split into two blocks acting separately on W and W_{\perp} . Both W and W_{\perp} may contain further invariant subspaces, in which case the matrices will be split still further.

The orthogonality theorem for irreducible representations

We come now to the central theorem of representation theory, a theorem that justifies the relatively routine application of certain procedures to determine the restrictions that are inherent in physical systems that have some degree of rotational or reflection symmetry. The development of the theorem is long and quite complex when presented in its entirety, and the reader will have to refer elsewhere for the proof.[§]

The theorem states that, in a certain sense, the irreps of a group \mathcal{G} are as orthogonal as possible, as follows. If, for each irrep, the elements in any one position in each of the g matrices are used to make up g -component column matrices then

- (i) any two such column matrices coming from different irreps are orthogonal;
- (ii) any two such column matrices coming from different positions in the matrices of the same irrep are orthogonal.

This orthogonality is in addition to the irreps' being in the form of orthogonal (unitary) matrices and thus each comprising mutually orthogonal rows and columns.

More mathematically, if we denote the entry in the i th row and j th column of a matrix $D(X)$ by $[D(X)]_{ij}$, and $\hat{D}^{(\lambda)}$ and $\hat{D}^{(\mu)}$ are two irreps of \mathcal{G} having dimensions n_λ and n_μ respectively, then

$$\sum_X \left[\hat{D}^{(\lambda)}(X) \right]_{ij}^* \left[\hat{D}^{(\mu)}(X) \right]_{kl} = \frac{g}{n_\lambda} \delta_{ik} \delta_{jl} \delta_{\lambda\mu}.$$

Firstly, the asterisk indicates that the complex conjugate should be taken if necessary, though all our representations so far have involved only real matrix elements. Each Kronecker delta function on the right-hand side has the value 1 if its two subscripts are equal and has the value 0 otherwise. Thus the right-hand side is only non-zero if $i = k$, $j = l$ and $\lambda = \mu$, all at the same time.

Secondly, the summation over the group elements X means that g contributions have to be added together, each contribution being a product of entries drawn from the representative matrices in the two irreps $\hat{D}^{(\lambda)} = \{\hat{D}^{(\lambda)}(X)\}$ and $\hat{D}^{(\mu)} = \{\hat{D}^{(\mu)}(X)\}$. The g contributions arise as X runs over the g elements of \mathcal{G} .

Thus, putting these remarks together, the summation will produce zero if either

- (i) the matrix elements are not taken from exactly the same position in every matrix, including cases in which it is not possible to do so because the irreps $\hat{D}^{(\lambda)}$ and $\hat{D}^{(\mu)}$ have different dimensions, or
- (ii) even if $\hat{D}^{(\lambda)}$ and $\hat{D}^{(\mu)}$ do have the same dimensions and the matrix elements are from the same positions in every matrix, they are different irreps, i.e. $\lambda \neq \mu$.

These theorems have many applications in areas such as crystal field splitting, quantum mechanical selection rules, electronic states of molecules, molecular vibrations and atomic spectra and band structure.

Characters

The actual matrices of general representations and irreps are cumbersome to work with, and they are not unique since there is always the freedom to change the coordinate system, i.e. the components of the basis vector (see section 29.3), and hence the entries in the matrices. However, one thing that does not change for a matrix under such an equivalence (similarity) transformation – i.e. under a change of basis – is the trace of the matrix. This was shown in chapter 8, but is repeated here. The trace of a matrix A is the sum of its diagonal elements,

$$\text{Tr } A = \sum_{i=1}^n A_{ii}$$

or, using the summation convention (section 26.1), simply A_{ii} . Under a similarity transformation, again using the summation convention,

$$\begin{aligned} [D_Q(X)]_{ii} &= [Q^{-1}]_{ij} [D(X)]_{jk} [Q]_{ki} \\ &= [D(X)]_{jk} [Q]_{ki} [Q^{-1}]_{ij} \\ &= [D(X)]_{jk} [I]_{kj} \\ &= [D(X)]_{jj}, \end{aligned}$$

showing that the traces of equivalent matrices are equal.

This fact can be used to greatly simplify work with representations, though with some partial loss of the information content of the full matrices. For example, using trace values alone it is not possible to distinguish between the two groups known as $4mm$ and $\bar{4}2m$, or as C_{4v} and D_{2d} respectively, even though the two groups are not isomorphic. To make use of these simplifications we now define the characters of a representation.

Definition. The characters $\chi(D)$ of a representation D of a group G are defined as the traces of the matrices $D(X)$, one for each element X of G .

At this stage there will be g characters, but, as we noted in subsection 28.7.3, elements A, B of G in the same conjugacy class are connected by equations of the form $B = X^{-1}AX$. It follows that their matrix representations are connected by corresponding equations of the form $D(B) = D(X^{-1})D(A)D(X)$, and so by the argument just given their representations will have equal traces and hence equal characters. Thus *elements in the same conjugacy class have the same characters*,

though, in general, these will vary from one representation to another. However, it might also happen that two or more conjugacy classes have the same characters in a representation

Orthogonality property of characters

Some of the most important properties of characters can be deduced from the orthogonality theorem |

$$\sum_X \left[\hat{D}^{(\lambda)}(X) \right]_{ij}^* \left[\hat{D}^{(\mu)}(X) \right]_{kl} = \frac{g}{n_\lambda} \delta_{ik} \delta_{jl} \delta_{\lambda\mu}.$$

If we set $j = i$ and $l = k$, so that both factors in any particular term in the summation refer to diagonal elements of the representative matrices, and then sum both sides over i and k , we obtain

$$\sum_X \sum_{i=1}^{n_\lambda} \sum_{k=1}^{n_\mu} \left[\hat{D}^{(\lambda)}(X) \right]_{ii}^* \left[\hat{D}^{(\mu)}(X) \right]_{kk} = \frac{g}{n_\lambda} \sum_{i=1}^{n_\lambda} \sum_{k=1}^{n_\mu} \delta_{ik} \delta_{ik} \delta_{\lambda\mu}.$$

Expressed in term of characters, this reads

$$\sum_X [\chi^{(\lambda)}(X)]^* \chi^{(\mu)}(X) = \frac{g}{n_\lambda} \sum_{i=1}^{n_\lambda} \delta_{ii}^2 \delta_{\lambda\mu} = \frac{g}{n_\lambda} \sum_{i=1}^{n_\lambda} 1 \times \delta_{\lambda\mu} = g \delta_{\lambda\mu}.$$

In words, the (g -component) ‘vectors’ formed from the characters of the various irreps of a group are mutually orthogonal, but each one has a squared magnitude (the sum of the squares of its components) equal to the order of the group.

Since, as noted in the previous subsection, group elements in the same class have the same characters, (29.14) can be written as a sum over classes rather than elements. If c_i denotes the number of elements in class \mathcal{C}_i and X_i any element of \mathcal{C}_i , then

$$\sum_i c_i [\chi^{(\lambda)}(X_i)]^* \chi^{(\mu)}(X_i) = g \delta_{\lambda\mu}.$$

Although we do not prove it here, there also exists a ‘completeness’ relation for characters. It makes a statement about the products of characters for a fixed pair of group elements, X_1 and X_2 , when the products are summed over all possible irreps of the group. This is the converse of the summation process defined by (29.14). The completeness relation states that

$$\sum_\lambda [\chi^{(\lambda)}(X_1)]^* \chi^{(\lambda)}(X_2) = \frac{g}{c_1} \delta_{\mathcal{C}_1 \mathcal{C}_2},$$

where element X_1 belongs to conjugacy class \mathcal{C}_1 and X_2 belongs to \mathcal{C}_2 . Thus the sum is zero unless X_1 and X_2 belong to the same class. For table 29.1 we can verify that these results are valid.

Counting irreps using characters

The expression of a general representation $D = \{D(X)\}$ in terms of irreps, as given in (29.11), can be simplified by going from the full matrix form to that of characters. Thus

$$D(X) = m_1 \hat{D}^{(1)}(X) \oplus m_2 \hat{D}^{(2)}(X) \oplus \cdots \oplus m_N \hat{D}^{(N)}(X)$$

becomes, on taking the trace of both sides,

$$\chi(X) = \sum_{\lambda=1}^N m_\lambda \chi^{(\lambda)}(X).$$

Given the characters of the irreps of the group \mathcal{G} to which the elements X and the characters of the representation $D = \{D(X)\}$, the g equations can be solved as simultaneous equations in the m_λ , either by inspection or by multiplying both sides by $[\chi^{(\mu)}(X)]^*$ and summing over X , making use of (29.14) and (29.15), to obtain

$$m_\mu = \frac{1}{g} \sum_X [\chi^{(\mu)}(X)]^* \chi(X) = \frac{1}{g} \sum_i c_i [\chi^{(\mu)}(X_i)]^* \chi(X_i).$$

That an unambiguous formula can be given for each m_λ , once the *character set* (the set of characters of each of the group elements or, equivalently, of each of the conjugacy classes) of D is known, shows that, for any particular group, two representations with the same characters are equivalent. This strongly suggests something that can be shown, namely, *the number of irreps = the number of conjugacy classes*. The argument is as follows.

is a set of

simultaneous equations for N unknowns, the m_λ , some of which may be zero. The value of N is equal to the number of irreps of \mathcal{G} . There are g different values of X , but the number of *different* equations is only equal to the number of distinct

conjugacy classes, since any two elements of \mathcal{G} in the same class have the same character set and therefore generate the same equation. For a unique solution to simultaneous equations in N unknowns, exactly N independent equations are needed. Thus N is also the number of classes, establishing the stated result.

The first summation rule for irreps is a simple restatement

with μ set

equal to λ ; it then reads

$$\sum_v [\chi^{(\lambda)}(X)]^* \chi^{(\lambda)}(X) = g.$$

Theorem. If n_μ is the dimension of the μ th irrep of a group \mathcal{G} then

$$\sum_{\mu} n_{\mu}^2 = g,$$

where g is the order of the group.

Proof. Define a representation of the group in the following way. Rearrange the rows of the multiplication table of the group so that whilst the elements in a particular order head the columns, their inverses in the same order head the rows. In this arrangement of the $g \times g$ table, the leading diagonal is entirely occupied by the identity element. Then, for each element X of the group, take as representative matrix the multiplication-table array obtained by replacing X by 1 and all other element symbols by 0. The matrices $D^{\text{reg}}(X)$ so obtained form the *regular representation* of \mathcal{G} ; they are each $g \times g$, have a single non-zero entry ‘1’ in each row and column and (as will be verified by a little experimentation) have

the same multiplication structure as the group \mathcal{G} itself, i.e. they form a faithful representation of \mathcal{G} .

The above is sometimes known as Burnside’s theorem which we will formally restate as:

(Theorem) Burnside:

- (Part 1): Suppose n_m is the dimensions of the matrix representations of the m -th inequivalent irrep of the group G and G has M number of such irreps, then the following equality holds

$$\boxed{\sum_{m=1}^M n_m^2 = |G|}$$

- (Part 2): The number of inequivalent irreducible representations is equal to the number of conjugacy classes in the group.

Construction of a character table

In order to decompose representations into irreps on a routine basis using characters, it is necessary to have available a character table for the group in question. Such a table gives, for each irrep μ of the group, the character $\chi^{(\mu)}(X)$ of the class to which group element X belongs. To construct such a table the following properties of a group, established earlier in this chapter, may be used:

- (i) the number of classes equals the number of irreps;
- (ii) the ‘vector’ formed by the characters from a given irrep is orthogonal to the ‘vector’ formed by the characters from a different irrep;
- (iii) $\sum_{\mu} n_{\mu}^2 = g$, where n_{μ} is the dimension of the μ th irrep and g is the order of the group;
- (iv) the identity irrep (one-dimensional with all characters equal to 1) is present for every group;
- (v) $\sum_X |\chi^{(\mu)}(X)|^2 = g$.
- (vi) $\chi^{(\mu)}(X)$ is the sum of n_{μ} m th roots of unity, where m is the order of X .

Some Applications:

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:

$$\sum_c n_c (\chi^{(r)}(c))^k \chi^{(s)}(c) = N(G) \delta^{rs} \quad (2)$$

Row orthogonality:

$$\sum_c \chi^{(r)}(c)^k \chi^{(s)}(c) = \frac{N(G)}{n_r} \delta^{rs} \quad (3)$$

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:

$$\sum_g D^{(r)\dagger}(g)_j^l D^{(s)}(g)_l^k = \frac{N(G)}{d_r} \delta^{rs} \delta_j^l \delta_k^l \quad (5)$$

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

$$\sum_c n_c \chi^{*(r)}(c) \chi(c) = N(G) n_r \quad (7)$$

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.

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-by-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	n_c	1	$1'$	$1''$
	1	I	1	1
Z_3	1	$c = (123)$	1	ω
Z_3	1	$a = (132)$	1	ω^*

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

abelian finite group with N elements.

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-by-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 in chapter II.2, 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

$$\sum_{j=0}^{N-1} e^{i2\pi(k-k')j/N} = N\delta_{kk'}$$

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 $1^2 + 1^2 + 2^2 = 7 > 6$, 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
Z_2	3	(12), (23), (31)	1	-1	0

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 $1 \cdot 2^2 + 2 \cdot x^2 + 3 \cdot 0^2 = 6 = 4 + 2x^2 \implies x^2 = 1$, 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: $1 \cdot 1 \cdot 2 + 2 \cdot 1 \cdot (-1) + 3 \cdot 1 \cdot 0 = 0$, 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-by-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* $(123) \sim \begin{pmatrix} \omega & 0 \\ 0 & \omega^*$ 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 $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \omega & 0 \\ 0 & \omega^4 \end{pmatrix} = \begin{pmatrix} 0 & \omega^4 \\ \omega & 0 \end{pmatrix}$. So, Confusio guessed wrong. The three transpositions (12), (23), and (31) are represented by $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\begin{pmatrix} 0 & \omega^4 \\ \omega & 0 \end{pmatrix}$, and $\begin{pmatrix} 0 & \omega \\ \omega^4 & 0 \end{pmatrix}$.

Some simple checks: they are unitary, traceless, and generate Z_2 , since, for example, $\begin{pmatrix} 0 & \omega^4 \\ \omega & 0 \end{pmatrix} \begin{pmatrix} 0 & \omega^4 \\ \omega & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. We could check the multiplication table some more, for example, $(23)(31) = (231) = (123)$: $\begin{pmatrix} 0 & \omega^4 \\ \omega & 0 \end{pmatrix} \begin{pmatrix} 0 & \omega \\ \omega^4 & 0 \end{pmatrix} = \begin{pmatrix} \omega & 0 \\ 0 & \omega^4 \end{pmatrix}$, 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

$$\sum_g D^{(r)}(g)^l_j = 0$$

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 from chapter I.1 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.

In chapter II.1, I remarked that S_n has an n -dimensional representation, its defining¹ 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

characters for the 3: $\begin{array}{|c|} \hline 3 \\ \hline 0 \\ \hline 1 \\ \hline \end{array}$

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.²

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

$$\text{add up correctly: } \begin{array}{|c|} \hline 3 \\ \hline 0 \\ \hline 1 \\ \hline \end{array} = \begin{array}{|c|} \hline 2 \\ \hline -1 \\ \hline 0 \\ \hline \end{array} + \begin{array}{|c|} \hline 1 \\ \hline 1 \\ \hline 1 \\ \hline \end{array}.$$

The reader might have recognized that this is the same problem as decomposing a vector that we started the review of linear algebra with: 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 square boxes.) 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.

We can perform some additional checks, for example, orthogonality between 3 and $\bar{1}$ gives $1 \cdot 1 \cdot 3 + 2 \cdot 1 \cdot 0 + 3 \cdot (-1) \cdot 1 = 0$, as expected (3 does not contain $\bar{1}$). Also, from (6), $1 \cdot 3^2 + 2 \cdot 0^2 + 3 \cdot 1^2 = 12 = 2 \cdot 6$ tells us that 3 contains two irreducible representations.

As you can see, it is difficult to make a careless arithmetical error without catching ourselves while playing this game; the orthogonality theorems provide a web of interlocking checks.

$$\sum_g D^{(r)}(g)^I_j = 0$$

Harmonic systems of springs and masses

case[†] of N particles of equal mass tied together by ideal springs and moving in

D -dimensional space. The deviation of the particles from their equilibrium positions provides the relevant coordinates. Denote the coordinates of the a th particle ($a = 1, \dots, N$) by x_a^i ($i = 1, 2, \dots, D$). Absorb the common mass into the spring constant and write Newton's law in the form*

$$\frac{d^2x_a^i}{dt^2} = - \sum_{b=1}^N \sum_{j=1}^D H^{ia,jb} x_b^j$$

(We are again temporarily suspending the repeated summation convention.) This system of equations is entirely linear, and thus linear combinations of solutions are also solutions.

Indeed, assemble the coordinates x_a^i into a DN -dimensional vector x^A , $A = 1, 2, \dots, DN$. Set \dagger $x^A(t) = x^A \sin(\omega t + \phi)$, thus obtaining the eigenvalue equation (reverting to the repeated index summation convention)

$$H^{AB} x^B = \omega^2 x^A$$

The real symmetric matrix H will have in general DN eigenvalues ω_α^2 and eigenvectors x_α^A with $\alpha = 1, 2, \dots, DN$. The vector x_α^A for a given α describes the α th eigenmode or harmonic mode with eigenfrequency squared ω_α^2 .

All this will become clear with a couple of examples.

The power of group theory

Even for relatively simple cases, the DN -by- DN matrix H^{AB} can be quite a mess. But if the system

of springs and masses, or the “molecule,” exhibits a symmetry, then H will be invariant under a group of transformations. The awesome power of group theory then manifests itself. Using the character table and a touch of physical intuition, we can often figure out what the harmonic modes are without even writing down H . In favorable cases, we can even learn a lot about the eigenfrequencies.

First let us discuss Schur's lemma:

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" (see chapter II.1). 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

I Schur

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.*

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 take the ij -component of the statement $HD(g) = D(g)H$ (using the upper and lower indices explained in the review of linear algebra but suspending the repeated index summation convention for the moment): $(HD(g))_j^i = H_i^j D_j^i(g) = (D(g)H)_j^i = D_j^i(g) H_j^i$, which implies that $(H_i^j - H_j^i) D_j^i(g) = 0$. 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 $D_j^i(g) = 0$ for all g (note the emphasis on “all” here), we can conclude $H_i^j = H_j^i$. 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:

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

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.

Schur's first lemma gives us a way to check whether the representation D is irreducible or not – if one can find a matrix M which is not of the form ce which commutes with all D_α then D_α is not an irrep. Generalizing Schur's first lemma to *different irreducible representations* gives Schur's Second Lemma.

Schur's Second Lemma: Suppose D_α and D'_α are two irreducible representations of G with dimensions n and n' respectively. Then if there exist a matrix $n \times n' M$ such that

$$MD_\alpha = D'_\alpha M$$

then if $n = n'$, $M = 0$ or D_α and D'_α are equivalent (i.e. related by a similarity transform), or if $n \neq n'$ then $M = 0$.

Let us return to the harmonic crystal problem:

Let us sketch the approach to be followed here, which amounts to an application of Schur's lemma and its consequences. The action of the symmetry transformation on the masses gives a DN -dimensional representation $D(g)$ of the symmetry group G . Since we know how the coordinates x_a^i change under the transformation g , in principle $D(g)$ is easily written down, although in practice some tedious work would be involved. Fortunately, as we learned, the characters of the representation often suffice. Once we know the characters,

we know how $D(g)$ falls apart into the irreducible representations of G , as pictured in

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

Recall also that the number of times n_r the irreducible representation r appears is determined by

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

and

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

Schur's lemma then tells us that

H has the form

$$H = \begin{pmatrix} \ddots & 0 & 0 & 0 & 0 \\ 0 & \omega_{(r)}^2 I_d & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & \omega_{(r)}^2 I_d & 0 \\ 0 & 0 & 0 & 0 & \ddots \end{pmatrix}$$

with I_d the d -by- d identity matrix.

Group theory cannot tell us what the eigenfrequencies $\omega_{(r)}$ are—that clearly has to do with the details of the system and what is sometimes called dynamics in physics—but it does tell us how many modes have the eigenfrequency $\omega_{(r)}$. That is given by the dimension d_r of the irreducible representation r . This is completely analogous to the situation in quantum mechanics: group theory can determine the pattern of degenerate levels in the energy spectrum, but it cannot tell us what the energies of the levels are.

Zero mode

After all this formalism, some examples, as was already promised, would be most welcome. One issue in pedagogy concerns the simplicity of the examples to be given (as I already said in chapter II.3). If the example is too simple, then it could provoke yawns, and worse, obscure or negate the power of the formalism just developed. But if the example is too complicated, it could hide the trees as well as the forest. In any case, we will start with an almost ridiculously simple example, that of two (equal) masses connected by a spring and moving in 1-dimensional space. See figure 1.



Figure 1

We can write down Newton's equations of motion by inspection: $\frac{d^2x_1}{dt^2} = -(x_1 - x_2)$ and $\frac{d^2x_2}{dt^2} = -(x_2 - x_1)$, which we can of course solve immediately. But that's not the point; instead, we would like to see how the group theoretic formalism works.

From the equations of motion we can read off¹

$$H = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

(Throughout this chapter, I will absorb all irrelevant constants.)

The symmetry group here is S_2 , consisting of the identity I and the element (12) exchanging particles 1 and 2. The character table is trivially constructed; we barely have to use any of the powerful theorems* in chapters II.2 and II.3. But do construct the table for practice before reading on.

Here it is.

S_2	n_c	c	1	$\bar{1}$
	1	I	1	1
Z_2	1	(12)	1	-1

The group has two elements, separating into two classes each with one element, and two irreducible representations, called the 1 and the $\bar{1}$, also known as even and odd in this context.

The $DN = 2 \cdot 1 = 2$ representation furnished by the two masses is simply given by $D(I) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $D((12)) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ (that is, in the same basis as used in (7)).

In the basis in which the $D(g)$ are diagonal, consisting of two 1-by-1 blocks, that is, the basis in which $D((12)) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. Schur's lemma forces H to have the diagonal form in (6), namely,[†] $H = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}$.

We see that there exists a mode with $\omega = 0$, known as a zero mode.² To understand what it means, go back to when we solve the time dependence by setting $x^A(t) = x^A \sin(\omega t + \phi)$. Zero frequency $\omega = 0$ actually does not mean that $x^A(t)$ does not depend on time. Since Newton's equation, unlike the Schrödinger equation, involves the second derivative in time, $x^A(t)$ can actually be a linear function of time for $\omega = 0$. Indeed, we now recognize this zero mode as the center of mass motion of the whole system, moving at some constant velocity. The eigenfrequency is zero precisely because the two masses are moving with the same velocity: the spring is not stretched at all.

Section 2

We shall move away from finite groups to study more continuous matrix groups.

Tensors and Representations of the Rotation Groups $SO(N)$

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?

But to help those readers seeing this for the first time focus, I will often specialize to

$N = 3$.

We have

$$R^T R = I$$

and

$$\det R = 1$$

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 -by- 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:

$$V^i \rightarrow V'^i = R^{ij} V^j$$

with $i, j = 1, 2, \dots, N$.

Several questions naturally suggest themselves. How many irreducible representations does the rotation group have? How are they to be characterized? What are their dimensions?

Our understanding of the representation theory of finite groups offers some suggestive answers to the questions posed above.

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: $D(e^{i2\pi j_1/N})D(e^{i2\pi j_2/N}) = e^{i2\pi k j_1/N} e^{i2\pi k j_2/N} = e^{i2\pi k(j_1+j_2)/N} = D(e^{i2\pi (j_1+j_2)/N})$. (I am being extra pedantic here.)

As $N \rightarrow \infty$?

Indeed, as proud physicists, let's take a flying guess at the representations of $SO(2)$.

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 chapter.) 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 $SO(N)$

would like to construct larger or higher-dimensional irreducible representations.

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

$$T^{ij} \rightarrow T'^{ij} = R^{ik}R^{jl}T^{kl}$$

then we say that T transforms like a tensor, and hence is a tensor.¹ (Here we are using the Einstein summation convention introduced in the review of linear algebra: 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.

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 $W^{ijn} \rightarrow W'^{ijn} = R^{ik}R^{jl}R^{nm}W^{klm}$. 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 $\begin{pmatrix} T^{11} \\ T^{12} \\ T^{13} \\ T^{21} \\ T^{22} \\ T^{23} \\ T^{31} \\ T^{32} \\ T^{33} \end{pmatrix}$. The linear transformation on the nine objects can then be represented by a 9-by-9 matrix $D(R)$ acting on this column.[†]

For every rotation, specified by a 3-by-3 matrix R , we can thus associate a 9-by-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

$$T^{ij} \rightarrow T'^{ij} = R_1^{ik}R_1^{jl}T^{kl}$$

$$\rightarrow T''^{ij} = R_2^{ik}R_2^{jl}T'^{kl} = R_2^{ik}R_1^{km}R_2^{jl}R_1^{ln}T^{mn} = (R_2R_1)^{lm}(R_2R_1)^{jn}T^{mn}$$

Thus, indeed, $D(R_2)D(R_1) = D(R_2R_1)$. 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.² (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,

$$\begin{aligned} A^{ij} &\rightarrow A^{ij} = T^{ij} - T^{ji} = R^{ik}R^{jl}T^{kl} - R^{jk}R^{il}T^{kl} \\ &= R^{ik}R^{jl}T^{kl} - R^{jl}R^{ik}T^{kl} = R^{ik}R^{jl}(T^{kl} - T^{lk}) = R^{ik}R^{jl}A^{kl} \end{aligned}$$

A^{ij} transforms like a tensor and is thus a tensor.

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-by-9 matrix referred to above actually breaks up into a 3-by-3 block and a 6-by-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

$$S^{il} \rightarrow S^{il} = R^{ik}R^{il}S^{kl} = (R^T)^{kl}R^{il}S^{kl} = (R^{-1})^{kl}R^{il}S^{kl} = \delta^{kl}S^{kl} = S^{kk}$$

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-by-6 matrix describing the linear transformation of the six objects S^{ij} breaks up into a 1-by-1 block and a 5-by-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

$$\tilde{S}^{ij} = S^{ij} - \delta^{ij}(S^{kk}/N)$$

(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-by-9 matrix $D(R)$ breaks up into a 3-by-3 matrix, a 1-by-1 matrix, and a 5-by-5 matrix “stacked on top of one another.” This is represented schematically as

$$S^{-1}D(R)S = \begin{pmatrix} \text{(3-by-3 block)} & 0 & 0 \\ 0 & \text{(1-by-1 block)} & 0 \\ 0 & 0 & \text{(5-by-5 block)} \end{pmatrix}$$

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 $\frac{1}{2}N(N-1)$ -dimensional representation, a $(\frac{1}{2}N(N+1)-1)$ -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.³

Invarlant 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.

$$S^{ii} = \delta^{ij}S^{ij} = (\delta^{ij}R^{ik}R^{jl})S^{kl} = \delta^{kl}S^{kl} = S^{kk},$$

it furnishes the 1-dimensional trivial representation]

$$\delta^{lj}R^{ik}R^{jl} = \delta^{ki}$$

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:

$$\epsilon^{\dots i \dots m \dots} = -\epsilon^{\dots m \dots i \dots} \quad \text{and} \quad \epsilon^{12\dots N} = 1$$

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 $N = 2$, $\epsilon^{12} = -\epsilon^{21} = 1$, with all other components vanishing. For $N = 3$, $\epsilon^{123} = \epsilon^{231} = \epsilon^{312} = -\epsilon^{213} = -\epsilon^{132} = -\epsilon^{321} = 1$, with all other components vanishing.

The determinant is defined, for any matrix R , by

$$\epsilon^{ijk\dots n} R^{ip} R^{jq} R^{kr} \dots R^{ns} = \epsilon^{pqr\dots s} \det R$$

See the review of linear algebra. (Verify this for $N = 2$ and 3.) For R a rotation, $\det R = 1$, and hence

$$\epsilon^{ijk\dots n} R^{ip} R^{jq} R^{kr} \dots R^{ns} = \epsilon^{pqr\dots s}$$

Dual tensors

In light of this discussion, given an antisymmetric tensor A^{ij} , we can define another antisymmetric tensor $B^{k\dots n} = \epsilon^{ijk\dots n} A^{ij}$ 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, $B^{k\dots n} \rightarrow \epsilon^{ijk\dots n} R^{ip} R^{jq} A^{pq}$. But multiplying (13) by a bunch of R^T 's carrying appropriate indices, we have $\epsilon^{ijk\dots n} R^{ip} R^{jq} = \epsilon^{pqr\dots s} R^{kr} \dots R^{ns}$. (Derive this!) Thus,

$$B^{k\dots n} \rightarrow \epsilon^{ijk\dots n} R^{ip} R^{jq} A^{pq} = \epsilon^{pqr\dots s} R^{kr} \dots R^{ns} A^{pq} = R^{kr} \dots R^{ns} B^{r\dots s}$$

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 $N = 3$, $B^k = \epsilon^{ijk} A^{ij}$ 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.

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 $T^{ij\cdots np} \rightarrow T^{ij\cdots np} = R^{ik}R^{jl}\cdots R^{nm}R^{pq}T^{kl\cdots mq}$. Take any two indices, say j and n , and contract them. Then

$$T^{ij\cdots jp} \rightarrow T^{ij\cdots jp} = R^{ik}R^{jl}\cdots R^{jm}R^{pq}T^{kl\cdots mq} = R^{ik}\cdots R^{pq}T^{kl\cdots iq}$$

since $R^{jl}R^{jm} = \delta^{lm}$. In other words, $T^{ij\cdots jp}$ transforms like a tensor $T^{i\cdots p}$ 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 $T^{ij\cdots jp}$, this actually stands for $T^{i1\cdots 1p} + T^{i2\cdots 2p} + \cdots + T^{iN\cdots Np}$."

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 $S^{i_1 i_2 \cdots i_j}$ 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 $T^{[ij]k} = (T^{ijk} + T^{jik})/2$ and the antisymmetric combination $T^{[ij]k} = (T^{ijk} - T^{jik})/2$. (The standard notation $[ij]$ 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 it is just a 2-indexed tensor $B^{lk} = \epsilon^{ijl}T^{[ij]k}$, 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 three indices by brute force. (Go ahead, do it before reading on.) Explicitly, write $3T^{[ij]k}$ as $(T^{[ij]k} + T^{[ki]j} + T^{[kj]i}) + (T^{[ij]k} - T^{[jk]i}) + (T^{[ij]k} - T^{[ki]j})$. Verify that the expression in the first parenthesis is completely symmetric in all three indices; indeed, it is just a sum over the $6 - 3!$ permutations of the three indices carried by $T^{[ij]k}$. The expressions in the other two round parentheses are antisymmetric in ki and kj , respectively; we can multiply them by ϵ^{kil} and ϵ^{kil} , respectively, turning them into 2-indexed tensors, and 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;

Dimension of the irreducible representations of $SO(3)$

Let us count the number of independent components contained in $S^{i_1 i_2 \dots i_j}$, which gives the dimension of the irreducible representation furnished by $S^{i_1 i_2 \dots i_j}$ and labeled by the integer j .

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 3s, 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:

$$\sum_{k=0}^j (k+1) = \frac{1}{2}j(j+1) + (j+1) = \frac{1}{2}(j+1)(j+2)$$

(As an interim check, we have $\frac{1}{2} \cdot 2 \cdot 3 = 3$ for $j = 1$, and $\frac{1}{2} \cdot 3 \cdot 4 = 6$ for $j = 2$.)

But we are not done yet: we have to impose the traceless condition: $\delta^{i_1 i_2} S^{i_1 i_2 \dots i_j} = 0$. The left hand side here is a totally symmetric tensor carrying $(j - 2)$ indices, which according to (16), has $\frac{1}{2}(j - 2 + 1)(j - 2 + 2) = \frac{1}{2}(j - 1)j$ components. Therefore, setting these to zero amounts to imposing $\frac{1}{2}j(j - 1)$ conditions. So, finally, the dimension of the irreducible representation j is

$$d = \frac{1}{2}(j+1)(j+2) - \frac{1}{2}j(j-1) = \frac{1}{2}(j^2 + 3j + 2 - j^2 + j) = 2j + 1$$

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 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 $\epsilon^{ij} T^{\dots i \dots j \dots}$ 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 $\delta^{i_1 i_2} S^{i_1 i_2 \dots i_j} = 0$. 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:

$$D^{(j)}(\theta) = \begin{pmatrix} \cos j\theta & \sin j\theta \\ -\sin j\theta & \cos j\theta \end{pmatrix}$$

In particular, $j = 1$ corresponds to the defining or fundamental representation.

This representation is reducible. Consider
the unitary transformation

$$\begin{aligned} U^\dagger D^{(j)}(\theta) U &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix}^\dagger \begin{pmatrix} \cos j\theta & \sin j\theta \\ -\sin j\theta & \cos j\theta \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix} \\ &= \frac{1}{2} \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix} \begin{pmatrix} e^{ij\theta} & e^{-ij\theta} \\ ie^{ij\theta} & -ie^{-ij\theta} \end{pmatrix} \\ &= \begin{pmatrix} e^{ij\theta} & 0 \\ 0 & e^{-ij\theta} \end{pmatrix} \end{aligned}$$

We have shown explicitly that the 2-dimensional representation $D^{(j)}(\theta)$ reduces to two 1-dimensional representations $e^{ij\theta}$ and $e^{-ij\theta}$.

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 $\epsilon^{klmn} T^{hijkl}$, 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.

The complicated symmetry properties are dealt with using Young tableaux but we will not discuss this.

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 $A^{i_1 i_2 \dots i_n}$ with $2n(2n-1) \dots (n+1)/n! = (2n)/(n!)^2$ components. From our discussion, you would think that it furnishes an irreducible representation. What could possibly reduce it?

Construct the tensor $B^{i_1 i_2 \dots i_n} \equiv \frac{1}{n!} \epsilon^{i_1 i_2 \dots i_n i_{n+1} i_{n+2} \dots i_{2n}} A^{i_{n+1} i_{n+2} \dots i_{2n}}$ dual to A . Then A is dual to B , that is, $A^{i_{n+1} i_{n+2} \dots i_{2n}} = \frac{1}{n!} \epsilon^{i_1 i_2 \dots i_n i_{n+1} i_{n+2} \dots i_{2n}} B^{i_1 i_2 \dots i_n}$.

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, $\epsilon T_{\pm} \sim \epsilon(A \pm B) \sim \epsilon A \pm \epsilon B \sim B \pm A \sim \pm(A \pm B) \sim \pm T_{\pm}$. Thus, $T_+ \sim \epsilon T_+$ is dual to itself, while $T_- \sim -\epsilon T_-$ 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 $G = SO(4)$, with the defining or vector representation consisting of the components of the 4-vector V^i , $i = 1, 2, 3, 4$. Consider the subgroup $H = SO(3)$ consisting of those elements of $SO(4)$ that leave V^4 alone. In other words, the subgroup $SO(3)$ 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 $SO(4)$ breaks into a 3-dimensional representation and a 1-dimensional representation of $SO(3)$.

In chapter VII.2 we will discuss in detail the Lorentz group, which at this point I simply say is, roughly speaking, just $SO(4)$ suitably modified. The subgroup $SO(3)$ 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 $SO(4)$ 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 in chapter VII.2, this corresponds to the electromagnetic field breaking into the electric and the magnetic fields.

How about the $\frac{1}{2} \cdot 4 \cdot 5 - 1 = 9$ -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 $\tilde{S}^{ab} = S^{ab} - \frac{1}{3}\delta^{ab}S^{cc}$

Rotations and the Notion of Lie Algebra

With a solid background on group theory in the past few chapters, we now embark the study of the capstone of our course – Lie Groups and Lie Algebras. Lie Groups are Continuous Groups which also possess the *additional property that functions of the elements of this Group is smooth and differentiable*, and furthermore analytic. We will elucidate on the bold words soon enough. Lie Groups are, without doubt, the most important class of group in physics. It pops its head out in every field from particle physics, quantum mechanics, condensed matter physics, to General Relativity, String Theory, and basically anything that you can think of in physics.

Despite its critical nature, the study of Lie Groups has its humble beginnings in Sophus Lie's (1842-1899) hope to solve some differential equations. He was inspired by Galois, who invented Galois Theory to show that one can use the properties of the symmetric group to solve algebraic equations. You probably have learned that the quadratic equation $ax^2 + bx + c = 0$ has the "solution by radicals" $x = (1/2a)(-b \pm \sqrt{b^2 - 4ac})$ – i.e. the roots are some "square root" of some clever combination of the coefficients. Galois proved that one can solve equations up to quartic order using such "solutions by radicals", but not quintic and above³⁹. To do this, he exploited the fact that some combinations of the roots are symmetric under the permutation group, the details of which we will not pursue here. Lie was wondering if he could use similar ideas to help him solve differential equations "by radicals", and found that he could. However, to do that, he developed the entire machinery of Lie Groups and Lie Algebras, which ended up revolutionizing both mathematics and physics.

Continuity, Smoothness and Analyticity

The study of continuous groups needs additional mathematical structure, namely what do we mean for two elements in a set to be “infinitesimally” close together. In addition to being continuous, the set of elements may or may not possess the additional property of being **smooth**. For example, consider the kink :

$$f(x) = \begin{cases} x + 1, & x < 0 \\ -x + 1, & x \geq 0 \end{cases}$$

which is continuous. However, $x = 0$ the kink does not “look smooth”. We make the definition of “smoothness” precise by considering the first derivative of this function

$$\frac{df}{dx} = \begin{cases} 1, & x < 0 \\ -1, & x \geq 0 \end{cases}$$

At $x = 0$, the derivative $d^2 f / dx^2$ is “undefined”. In other words, while $f(x)$ is differentiable, its derivative $df(x)/dx$ is *not differentiable*. More generally, some functions are differentiable *everywhere* for up to n times, which we categorized them as C_n functions. Hence a smooth function is a continuous C_∞ function.

(Definition) Smoothness: A C_n function $f(x)$ is a function which can be differentiated n times. A **smooth function** is a C_∞ function. If a function has more than one variable, i.e $f(x, y, z, \dots)$, each variable may have different levels of differentiability⁴⁰.

A function which is smooth within some (open) domain M can hence be differentiated an infinite number of times within this domain. This allow us to expand this function as a **Taylor Series** at any point $x_0 \in M$, i.e.

$$f(x) = f(x_0) + \left. \frac{df}{dx} \right|_{x_0} (x - x_0) + \frac{1}{2!} \left. \frac{d^2 f}{dx^2} \right|_{x_0} (x - x_0)^2 + \dots$$

This series may or may not converge⁴¹, depending on the exact form of f . If this series converges to its functional value *everywhere* in the (open) domain considered, then we say that the function f is **analytic** in M . All this is familiar to you when you do Calculus 101. Since we will be exclusively be dealing with C_∞ functions, we will now drop the awkward term “infinitely” from now on.

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.

Recall our working definition for a continuous group is that the elements can be parameterized by \mathbb{R}^n or a compact subset of \mathbb{R}^n . Consider a simple $n = 1$ continuous group, $SO(2)$.

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

$$R(\theta) \simeq I + A$$

Here **A denotes some infinitesimal matrix of order θ** .

The neglected terms are of higher order in θ

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

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

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

But there is basically only one 2-by-2 antisymmetric matrix:

$$\mathcal{J} \equiv \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

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[†]

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

[†]An equivalent way of saying this is that for infinitesimal θ , the transformation $x' \simeq x + \theta y$ and $y' \simeq y - \theta x$ satisfies the Pythagorean condition $x'^2 + y'^2 = x^2 + y^2$ to first order in θ . (You could verify that (1) indeed reduces to this transformation to leading order in θ .) Or, write $x' = x + \delta x$, $y' = y + \delta y$, and solve the condition $x \delta x + y \delta y = 0$ that $\delta \vec{r}$ is perpendicular to \vec{r} .

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 $e^x = \lim_{N \rightarrow \infty} (1 + \frac{x}{N})^N$ (which you can easily prove by differentiating both sides). Then, for a finite (that is, not infinitesimal) angle θ , we have

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

which is precisely $R(\theta)$

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

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.

But it is very easy to write down all possible antisymmetric N -by- N matrices! For $N = 2$, there is only one, namely, the \mathcal{J} introduced earlier. For $N = 3$, there are basically three of them:

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

Any 3-by-3 antisymmetric matrix can be written as $A = \theta_x \mathcal{J}_x + \theta_y \mathcal{J}_y + \theta_z \mathcal{J}_z$, 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-by-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

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

(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 imaginary³ unit i . Thus, define⁴ $J_x \equiv -i \mathcal{J}_x$, $J_y \equiv -i \mathcal{J}_y$, $J_z \equiv -i \mathcal{J}_z$, and write a general rotation as

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

treating the three real numbers θ_j and the three matrices J_j as two 3-dimensional vectors.

In general, consider an n -parameter Lie Group labeled by (a_1, a_2, \dots, a_n) acting on an N -th dimensional target space labeled by coordinates $\{x_i\}$ for $i = 1, 2, \dots, N$. A general transformation is given by the set of equations

$$\begin{aligned} x'_1 &= f_1(x_1, x_2, \dots, a_1, a_2, \dots) \\ x'_2 &= f_2(x_1, x_2, \dots, a_1, a_2, \dots) \\ &\vdots \\ x'_i &= f_i(x_1, x_2, \dots, a_1, a_2, \dots) \end{aligned}$$

and as long as all the f_i 's are analytic in the a_j 's Lie parameters the transformation is analytic.

Before we close this section, we want to make a subtle point clear. In this section, we have chosen to represent Lie Group transformations as functions of some choice of parameterization of their group elements, and argue that analyticity of the parameters determines whether they are Lie Groups or not. However, sometimes we may choose to parameterize Lie Groups with some choices of parameters which *may not be analytic throughout the entire group* – perhaps in some physical situations the operations which do not overlap with the region of the group which is not “covered by the parameterization”. An example is the 3-parameter Lie Group $SL(2, \mathbb{R})$, which one can parameterize by the following 2D representation

$$T(a_1, a_2, a_3) = \begin{pmatrix} a_1 & a_2 \\ a_3 & \frac{1+a_2a_3}{a_1} \end{pmatrix} \in SL(2, \mathbb{R}), \quad a_i \in \mathbb{R}$$

where it is clear that the parameterization is not analytic at $a_1 \rightarrow 0$. It turns out that the reason this is so is that $SL(2, \mathbb{R})$ is actually not parameterized by \mathbb{R}^3 , but by $\mathbb{R}^2 \times S^1$. In other words, one of the parameters is periodic, and hence compact so we have to restrict the domain of the parameterization to the region where the transform is analytic. You will be guided through how to show this in a homework problem. The point here is that it is often a non-trivial task to show analyticity of a Lie Group – one has to first choose a parameterization which “covers” the whole group in an isomorphic way. We will not delve too deeply into this discussion – which requires knowledge of **Topology** – in our introductory symmetry class though.

Lie algebra

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

$$[A, B] \equiv AB - BA,$$

known as the commutator* between A and B .

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:

$$[J_i, J_j] = i c_{ijk} J_k$$

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 (convince yourself of this) numbers. Evidently, $c_{ijk} = -c_{jik}$.

By explicit computation using (14), we find

$$[J_x, J_y] = i J_z$$

$$[J_y, J_z] = i J_x$$

$$[J_z, J_x] = i J_y$$

The three commutation relations

may be summarized by $[J_i, J_j] = i\epsilon_{ijk}J_k$ (sum over k implied)

From these commutation rules it follows that there exists an operator

$$J^2 = J_x^2 + J_y^2 + J_z^2$$

which has the property of commuting with all the angular momentum operators:

$$[J^2, J_x] = [J^2, J_y] = [J^2, J_z] = 0.$$

J^2 is known as a Casimir operator.

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 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

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

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.

Rotations in higher-dimensional space

Start with an N -by- 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 $J_{(mn)}$. 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 $J_{(mn)} = -i J_{(mn)}$. Explicitly,

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

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 $SO(N)$.

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 ,

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

The Quantum Harmonic Oscillator

The Classical Hamiltonian can be **Quantized** by replacing the variables x and p with the operators \hat{x} and \hat{p} .

$$H = \frac{\hat{p}^2}{2m} + \frac{1}{2}m\omega^2\hat{x}^2$$

We now define the **Creation Operator** a^\dagger and the **Annihilation Operator** a in terms of \hat{x} and \hat{p} :

$$\hat{x} = \sqrt{\frac{\hbar}{2m\omega}}(a^\dagger + a), \quad \hat{p} = i\sqrt{\frac{\hbar m\omega}{2}}(a^\dagger - a)$$

where \hbar is a constant (the normalised Planck Constant).

$$[a, a^\dagger] = I, \quad [a, I] = [a^\dagger, I] = 0$$

This is known as the Heisenberg algebra and encodes the canonical commutation relation $[\hat{x}, \hat{p}] = i$.

Harmonic oscillator approach to SU(2) Lie algebra

Let a_p and a_q denote two independent harmonic oscillators where p will denote proton and q will denote neutron (with the application to isospin in mind).

Let us now construct the Lie algebra of all possible bilinear products of these operators which do not change the number of particles. These are products of one creation operator and one annihilation operator. Since there are two possible operators of each type, there are in all four possible bilinear products which do not change the number of particles:

$$a_p^\dagger a_n, \quad a_n^\dagger a_p, \quad a_p^\dagger a_p \quad \text{and} \quad a_n^\dagger a_n.$$

The first of these operators annihilates a neutron and creates a proton in the same quantum state; i.e. it changes a neutron into a proton. The second operator does the reverse, changing a proton into a neutron. These two operators are thus the ordinary isospin operators τ_+ and τ_- . The other two operators annihilate either a proton or a neutron and create the same particle back again. These are just number operators which count the number of protons and neutrons. The sum of the last two operators is just the total number operator which counts the number of nucleons. Since all of the other operators do not change the number of nucleons, this total number operator commutes with all the others. It is therefore convenient to divide the set of four operators into a set of three plus the total number, or baryon number, operator which commutes with all of the others.

$$B = a_p^\dagger a_p + a_n^\dagger a_n,$$

$$\tau_+ = a_p^\dagger a_n,$$

$$\tau_- = a_n^\dagger a_p,$$

$$\tau_0 = \frac{1}{2}(a_p^\dagger a_p - a_n^\dagger a_n) = Q - \frac{1}{2}B.$$

The operator τ_0 defined as half the difference between the number of protons and the number of neutrons is just equal to the total charge Q minus half the baryon number, since the protons carry one unit of charge and the neutrons carry no charge. The operators τ_+ , τ_- and τ_0 satisfy commutation rules exactly like angular momenta

$$[\tau_0, \tau_+] = \tau_+,$$

$$[\tau_0, \tau_-] = -\tau_-,$$

$$[\tau_+, \tau_-] = 2\tau_0.$$

Harmonic oscillator approach to SU(3) Lie algebra

Consider now the case where there are three quantum states. A convenient example of this case is the Sakata model of elementary particles in which the transformations of isospin are extended to include the lambda hyperon as well as the proton and the neutron. Let a_A^\dagger and a_A be operators for the creation and annihilation of a lambda particle. We now construct the Lie algebra of all possible bilinear products of the nucleon and lambda operators which do not change the number of particles. With three creation operators and three annihilation operators, there are nine possible bilinear products. These are conveniently written as follows:

$$B = a_p^\dagger a_p + a_n^\dagger a_n + a_\Lambda^\dagger a_\Lambda,$$

$$\tau_+ = a_p^\dagger a_n, \quad \tau_- = a_n^\dagger a_p,$$

$$\tau_0 = \frac{1}{2}(a_p^\dagger a_p - a_n^\dagger a_n),$$

$$B_+ = a_p^\dagger a_\Lambda, \quad B_- = a_n^\dagger a_\Lambda,$$

$$C_+ = a_\Lambda^\dagger a_n, \quad C_- = a_\Lambda^\dagger a_p,$$

$$N = \frac{1}{3}(a_p^\dagger a_p + a_n^\dagger a_n - 2a_\Lambda^\dagger a_\Lambda) = \frac{1}{3}B + S.$$

Let us now consider which Lie group is associated with these operators. By an extension of isospin we see that these operators generate infinitesimal transformations in a three-dimensional proton–neutron–lambda Hilbert space. These transformations are again unitary; thus the Lie group associated with these operators is the group of unitary transformations in a three-dimensional space.

As in the case of isospin, some of the bilinear products are operators which change one kind of particle into another, while others are number operators which simply count the number of particles of a particular kind. Again the sum of all the number operators is just the baryon number and commutes with all of the other operators which do not change the baryon number. We therefore divide the set of nine operators into a set of eight plus the baryon number which commutes with all of the rest. In the set of eight operators, there are still two number operators and it is convenient to choose the linear combinations given above.

The two operators τ_0 and N

commute with one another, and it is impossible to find a third operator which commutes with both of these.

Such a Lie algebra is said to have rank 2. The set of 8 operators excluding B generate $SU(3)$.

The algebra is

$$\begin{aligned} [\tau_0, \tau_{\pm}] &= \pm \tau_{\pm}, & [N, \tau_{\pm}] &= 0, \\ [\tau_0, B_{\pm}] &= \pm \frac{1}{2} B_{\pm}, & [N, B_{\pm}] &= B_{\pm}, \\ [\tau_0, C_{\pm}] &= \pm \frac{1}{2} C_{\pm}, & [N, C_{\pm}] &= -C_{\pm}. \end{aligned}$$

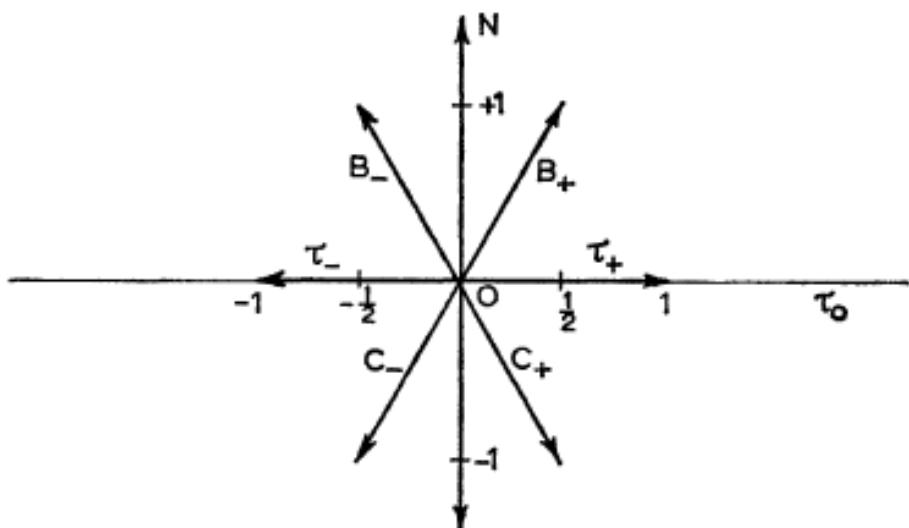
The remaining commutators are easily obtained by simple algebra.

$$\begin{aligned}
[\tau_{\pm}, B_{\pm}] &= [\tau_{\pm}, C_{\pm}] = 0 = [C_+, C_-] = [B_+, B_-], \\
[\tau_{\pm}, B_{\mp}] &= B_{\pm}, \quad [B_{\pm}, C_{\pm}] = \tau_{\pm}, \\
[\tau_{\pm}, C_{\mp}] &= -C_{\mp}, \quad [\tau_+, \tau_-] = 2\tau_0, \\
[B_+, C_-] &= \frac{1}{2}(3N + 2\tau_0), \\
[B_-, C_+] &= \frac{1}{2}(3N - 2\tau_0).
\end{aligned}$$

Following the analogy with isospin we might attempt to find operators C_{μ} which are functions of these operators and commute with all of them. The eigenvalues of these operators would then be used to label the multiplets as the eigenvalues of the operator T^2 label the isospin multiplets. However, the operators C_{μ} for the SU_3 are rather complicated.

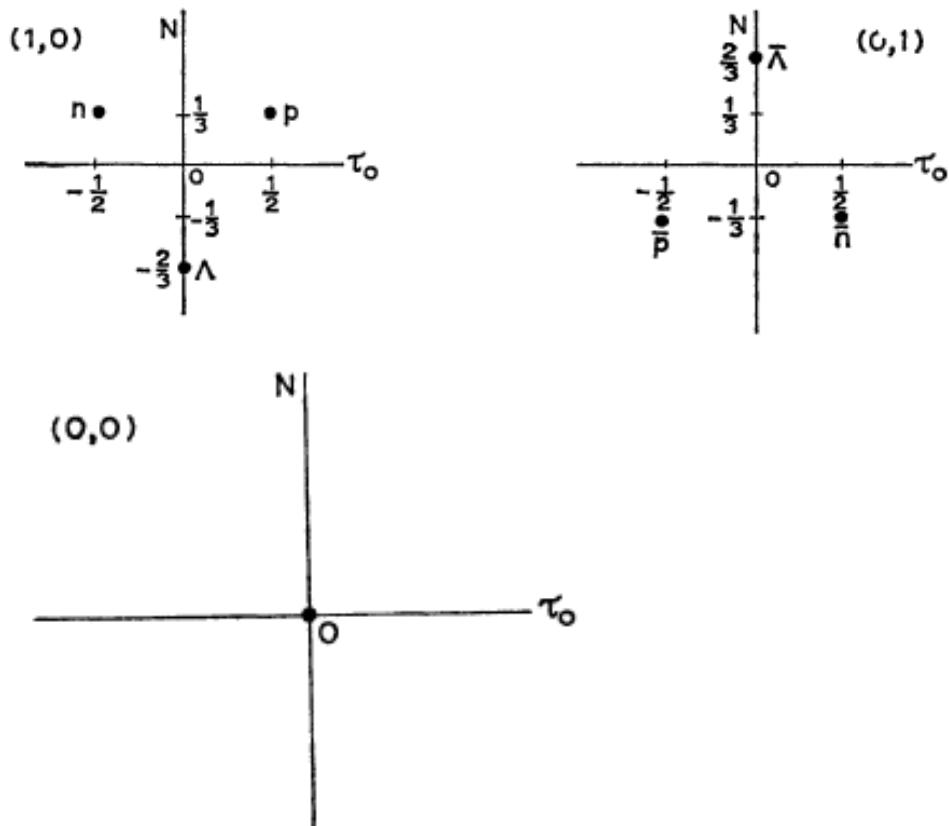
THE STRUCTURE OF THE MULTIPLETS

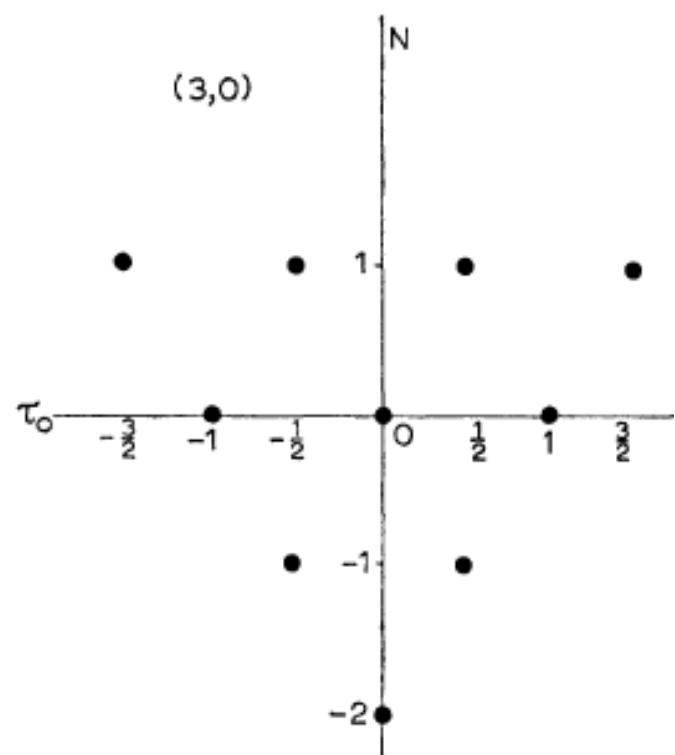
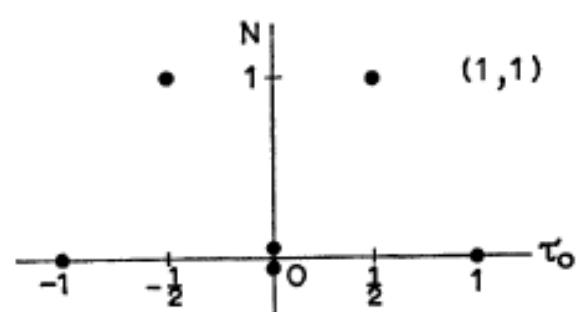
The SU_3 multiplets are generated by successive operation on any state within the multiplet with the eight operators of the Lie algebra. The states of each multiplet are represented as points on a two-dimensional plot of the eigenvalues of τ_0 and N . The points representing the states of a given multiplet should appear in such a plot as a two-dimensional lattice in which the lattice vectors are just the vectors of

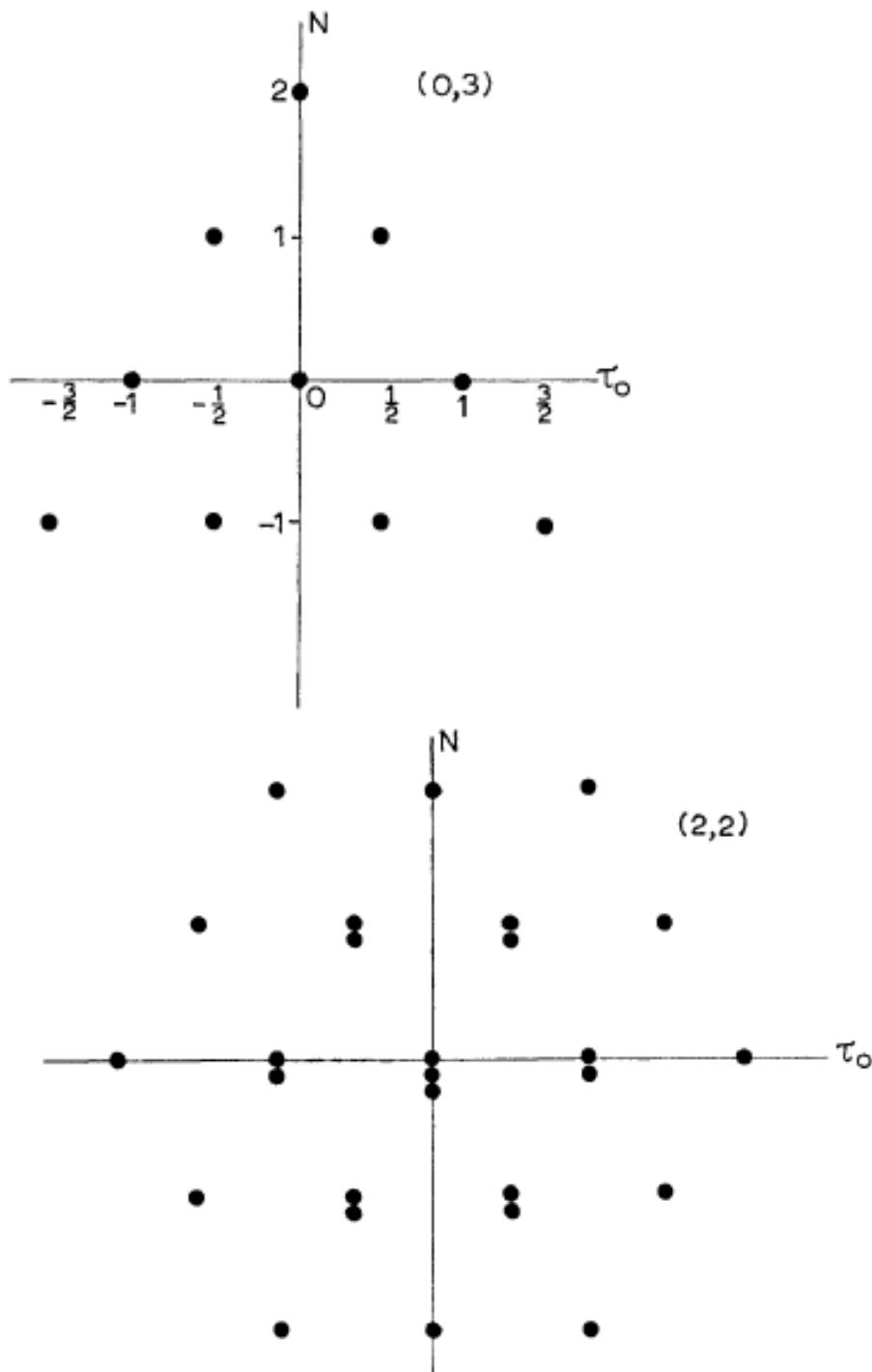


The lattice therefore has the hexagonal

character in which a change of N by ± 1 is always accompanied by a change in T_0 of $\pm \frac{1}{2}$. Since the SU_3 algebra is larger than the isospin algebra and includes it as a subset, we expect the multiplets for SU_3 to be larger than isospin multiplets and to contain several isospin multiplets at different values of N . Since the B - and C -operators change T_0 by $\pm \frac{1}{2}$, both integral and half-integral isospins occur in the same multiplet in contrast to the case of isospin multiplets. From the hexagonal character of the lattice we see that integral and half-integral isospin multiplets appear alternately with increasing values of the quantum number N .







General properties of Lie algebras:

(Theorem) **Jacobi Identity:** Given any 3 operator A, B, C and a binary (commutator) operation $[A, B] = AB - BA$, then the operators obey the Jacobi Identity

$$[[[A, B], C] + [[B, C], A] + [[C, A], B] = 0]$$

- **Linear Vector Space:** Lie Algebras \mathfrak{g} form a linear vector space under vector addition and scalar multiplication, i.e. if $X, Y \in \mathfrak{g}$ and $a, b \in \mathbb{F}$ is some field (such as \mathbb{R} or \mathbb{C}), then

$$aX + bY \in \mathfrak{g}.$$

- *Closure under commutation:* If $X, Y \in \mathfrak{g}$, then

$$[X, Y] \in \mathfrak{g}.$$

- *Jacobi Identity:* It obeys the Jacobi Identity.

It should be clear to you that there are many different representations for Lie Algebras. However, there is a particularly special representation called the **Adjoint Representation** which we will now discuss.

(Definition) Adjoint Representation: The set of matrices defined by the structure constants $(C_i)_j^k \equiv -c_{ij}^k$ of some Lie Algebra forms a representation of the parent Lie Group⁴⁹.

Let's parse the above definition as follows. We substitute Eq. (414) into the Jacobi Identity Eq. (415), using $A = l_i, B = l_j$ and $C = l_j$, some algebra leads us to the relation

$$c_{ij}^m c_{mk}^n + c_{jk}^m c_{mi}^n + c_{ki}^m c_{mj}^n = 0.$$

Now if we consider c_{ij}^k as a set of $n \times n$ matrices C_i , defined as

$$(C_i)_j^k \equiv -c_{ij}^k$$

where the upper index labels the columns and the lower index labels the rows of the matrix, we can rewrite as

$$(c_{ij}^m C_m)_k^n + (C_j C_i)_k^n - (C_i C_j)_k^n = 0.$$

Dropping the subscripts k and superscript n ,

and replacing the index $m \rightarrow k$, we get

$$C_i C_j - C_j C_i = [C_i, C_j] = c_{ij}^k C_k$$

which defines the algebra

Consider $\mathfrak{so}(3)$, the structure constants $c_{ij}^k = i\epsilon_{ijk}$, and we can see that the set of 3 matrices $(C_i)_j^k = -c_{ij}^k$ are

$$\begin{aligned} C_1 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} = L_1 \\ C_2 &= \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix} = L_2 \\ C_3 &= \begin{pmatrix} 0 & i & 0 \\ -i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = L_3. \end{aligned}$$

Application : Rotation Symmetry in Quantum Mechanics

In this chapter, we will show an application of group theory to a physics problem – the problem of finding the energy eigenstates of a quantum mechanical particle in a spherical symmetric potential. You will have seen this problem done in your usual Quantum Mechanics class, so I hope you will be familiar with the results. We will attack this problem from a Group Theory perspective.

Quantum Mechanics of Hydrogen Atom Review

We begin by recalling the time-independent Schrödinger's Equation in 3 dimensions with potential $V(\mathbf{x})$ using Cartesian coordinates

$$\left[-\frac{\hbar^2}{2m} \left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2} \right) + V(\mathbf{x}) \right] \psi(\mathbf{x}) = E\psi(\mathbf{x})$$

where $\psi(\mathbf{x})$ is the wavefunction. We can express this in terms of the Hamiltonian operator

$$\hat{H} \equiv -\frac{\hbar^2}{2m} \left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2} \right) + V(\mathbf{x})$$

So

$$\hat{H}\psi(\mathbf{x}) = E\psi(\mathbf{x}).$$

Recall that the derivative operators $D \equiv \partial^2/\partial x_i^2$ are linear i.e.,

$$D(a\psi(\mathbf{x}) + b\phi(\mathbf{x})) = aD\psi(\mathbf{x}) + bD\phi(\mathbf{x})$$

If, furthermore, the potential $V(\mathbf{x})$ is also spherically symmetric, i.e. $V(\mathbf{x}) = V(r)$ where $r = \sqrt{x_1^2 + x_2^2 + x_3^2}$

$$\begin{aligned} x_1 &= r \cos \phi \sin \theta \\ x_2 &= r \sin \phi \sin \theta \\ x_3 &= r \cos \theta \\ r &= \sqrt{x^2 + y^2 + z^2}. \end{aligned}$$

so

$$-\frac{\hbar^2}{2m} \left[\frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial \psi}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial \psi}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 \psi}{\partial \phi^2} \right] + V(r)\psi = E\psi.$$

Use the ansatz:

$$\psi(r, \theta, \phi) = \chi_{nlm}(r)Y_{l,m}(\phi, \theta)$$

where $Y_{l,m}$ are the Spherical Harmonics which form a complete basis for the compact 2-sphere S_2 labeled by (ϕ, θ) – you can think of them as “fourier transforms” of a compact 2-sphere. The radial function χ_{nl} is dependent on the Angular Momentum Quantum number l , and the Principal Quantum Number n but not the Magnetic Angular Momentum Quantum Number m . Plugging in Eq. (437) into Eq. (435), we get a set of equations labeled by l

$$-\frac{\hbar^2}{2m} \left[\frac{1}{r} \frac{\partial^2}{\partial r^2} (r\chi_{nlm}) \right] + \frac{l(l+1)\hbar^2}{2mr^2} \chi_{nlm} - U(r)\chi_{nlm} = E\chi_{nlm}.$$

You then solve Eq. (438) using a power series, and found that χ_{nlm} is a Laguerre Polynomial

$$\chi_{E_n,l}(r) = e^{-\zeta/n} \sum_{k=l+1}^n a_k \zeta^{k-1}.$$

with

$$\zeta = \frac{m_e e^2}{4\pi e_0 \hbar^2 r}, \quad a_{k+1} = \frac{2(ak-1)}{k(k+1)-l(l+1)} a_k, \quad a^2 = -\frac{32\pi^2 e_0^2 \hbar^2}{m_e e^4} E.$$

The solutions are labeled by quantum number (n, l, m) . For each Principal number $n > 0$, there exists an n -fold degeneracy in the Angular Momentum number $l = 0, 1, 2, \dots, n-1$, and for each l there exists a $(2l+1)$ -fold degeneracy in the Magnetic Angular Momentum number $m = -l, -l+1, \dots, -1, 0, 1, \dots, l-1, l$. So far so familiar. One of the things that you did *not* derive from this method is the fact that we may have *half-integer* spins $l = 1/2, 3/2, \dots$ states. Indeed, the electron spin is $1/2$, with two possible states $(-1/2, 1/2)$. Where did that come from?

In this lecture however, we will take a different tack and look at spherical symmetry in Quantum Mechanics from the perspective of Group Theory. We will show that for a quantum mechanical system with spherical symmetry $SO(3)$, we can construct all the possible states – including the spin $1/2$ states – using only our knowledge of the symmetry of the system.

Spherical Symmetry of the Hamiltonian

Let's begin by recalling what a rotation in 3D does. From our study in chapter 5, we recognize that $SO(3)$ is the group we want to consider. The rotation operators are then elements of the rotation group in 3D, or $R \in SO(3)$. If we arrange the 3D Cartesian space as a vector $\mathbf{x} = (x_1, x_2, x_3)$, then a rotation simply rotates it to another point $\mathbf{x}' = x'_i = (x'_1, x'_2, x'_3)$, i.e.

$$R\mathbf{x} = \mathbf{x}'.$$

Now consider the 3D Schrödinger's equation with spherically symmetric potential $V(r)$,

$$\left[-\frac{\hbar^2}{2m} \left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2} \right) + V(r) \right] \psi(\mathbf{x}) = E\psi(\mathbf{x}).$$

Let $\psi(\mathbf{x})$ be a solution to Eq. (442). Now, if we can define a new function $\psi'(\mathbf{x})$,

$$\psi'(\mathbf{x}) = \psi(R\mathbf{x}) = \psi(\mathbf{x}').$$

It is now straightforward to show that $\psi'(\mathbf{x})$ must also be a solution to Eq. (442). Taking derivative of $\psi(\mathbf{x}')$ and using the chain rule, we get

$$\frac{\partial}{\partial x_i} \psi(\mathbf{x}') = \frac{\partial x'_j}{\partial x_i} \frac{\partial}{\partial x'_j} \psi(\mathbf{x}').$$

Now we can define

$$R_{ij} \equiv \frac{\partial x'_j}{\partial x_i}$$

which we have suggestively called " R_{ij} " since it is just the rotation operator. To see this, let's write Eq.

$$R_{ij}x_j = x'_i,$$

so since R_{ij} is a linear transformation, we can take derivatives on both side to get

$$R_{ij} = \frac{\partial x'_i}{\partial x_j}.$$

R is an element of $SO(3)$

it must obey the orthogonality and positive unit determinant conditions

$$R^T R = I \rightarrow R_{ik} R_{kj} = \delta_{kj}, \quad \det R_{ij} = 1.$$

Taking the second derivatives of $\psi(x_i)$, we then get

$$\begin{aligned} \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_i} \psi(\mathbf{x}') &= \frac{\partial x'_j}{\partial x_i} \frac{\partial x'_k}{\partial x_i} \frac{\partial}{\partial x'_j} \frac{\partial}{\partial x'_k} \psi(\mathbf{x}') \\ &= R_{ji} R_{ki} \frac{\partial}{\partial x'_j} \frac{\partial}{\partial x'_k} \psi(\mathbf{x}') \\ &= \delta_{jk} \frac{\partial}{\partial x'_j} \frac{\partial}{\partial x'_k} \psi(\mathbf{x}') \\ &= \frac{\partial}{\partial x'_j} \frac{\partial}{\partial x'_j} \psi(\mathbf{x}'), \end{aligned}$$

We can investigate the potential $V(\mathbf{x})$ under the same transformation, but since it is spherical symmetric:

$$V(\mathbf{x}') = V(R\mathbf{x})$$

it remains invariant.

Hence, we have proven that $\psi'(\mathbf{x}) = \psi(R\mathbf{x})$ is also a solution

$$\psi'_i(\mathbf{x}) = D_{ij}(R)\psi_j(\mathbf{x}) \quad \text{where } D_{ij} \text{ are determined by } R |$$

v' = matrix representation of R :

$$\psi(R_1\mathbf{x}) = D(R_1)\psi(\mathbf{x}), \quad \psi(R_2\mathbf{x}) = D(R_2)\psi(\mathbf{x})$$

Also

$$\psi(R_1 R_2 \mathbf{x}) = D(R_1 R_2)\psi(\mathbf{x}) = D(R_1)\psi(R_2 \mathbf{x}) = D(R_1)D(R_2)\psi(\mathbf{x})$$

So

$$D(R_1 R_2) = D(R_1) D(R_2), \quad \text{which is exactly the definition for matrix representations.}$$

From our earlier discussion of the representations of $SO(3)$ we choose the common eigenvectors of J^2 and J_z .

$$J_z |J, M\rangle = M |J, M\rangle$$

and

$$J^2 |J, M\rangle = J(J+1) |J, M\rangle.$$

$$(J_x \pm iJ_y) |J, M\rangle = \sqrt{J(J+1) - M(M \pm 1)} |J, M \pm 1\rangle$$

Beginning with any particular state, $|J, M\rangle$, a set of states can be generated by operating successively with the operators $(J_x + iJ_y)$ and $(J_x - iJ_y)$. This process cannot be continued indefinitely because M can never be greater than J . Thus one finds restrictions on the possible eigenvalues of J and M , and obtains the well-known result that these may be either integral or half-integral and that for any eigenvalue J there corresponds a set or multiplet of $2J+1$ states all having the same eigenvalue of J and having values of M equal to $-J, -J+1, \dots, +J$. The full set of states in a multiplet can be generated from any one of the states by successive operation with the operators $(J_x \pm iJ_y)$.

More generally, operator multiplets can be defined in a manner analogous to the wave function multiplets. These operator multiplets are called irreducible tensors and their components transform into linear combinations of one another under rotations. The commutators of such tensor operators with the angular momentum

$$[(J_x \pm iJ_y), T_{kq}] = \sqrt{k(k+1) - q(q \pm 1)} T_{k(q \pm 1)}$$

$$[J_z, T_{kq}] = q T_{kq}$$

where T_{kq} is the q -component of an irreducible tensor of degree k , and the indices k and q are analogous to the angular momentum eigenvalues J and M for the corresponding wave function multiplet. Such an irreducible tensor has $2k+1$ components and the index q takes on $2k+1$ values from $-k$ to $+k$.

Irreducible tensors combine in the same way as angular momentum multiplets.

N.B. An irreducible tensor of degree k is a set of $(2k+1)$ operators which transform under rotations in the representation of $SO(3)$ labeled by k .

We could change the problem to a non-rotationally invariant one where the Hamiltonian H is

$$H = H_0 + K J_z$$

where H_0 commutes with all the angular momentum operators and the K is a constant. For this case, the eigenstates of the Hamiltonian can still be chosen to be simultaneous eigenfunctions of J^2 and J_z , but the $2J+1$ states of a multiplet are no longer degenerate. The splitting of the energy levels in a multiplet is determined by the terms in the Hamiltonian that do not commute with J_x and J_y . The Hamiltonian does not commute with all the angular momentum operators, but the commutators have a simple form.

$$[H, (J_x \pm i J_y)] = \pm K (J_x \pm i J_y) .$$

The commutator of H with J_z vanishes. The splitting of the energies in the multiplet is proportional to K .