Symmetry in Physics

Lecturer: Eugene A. Lim
King's College London
Department of Physics
April 17, 2022

Recommended Books

- H. F. Jones, *Groups, Representations and Physics*, Taylor and Francis 1998.
- C. Isham, *Lectures on Groups and Vector Spaces: For Physicists*, World Scientific 1989.
- B. Schutz, *Geometrical Methods of Mathematical Physics*, Cambridge University Press 1989.
- S. Sternberg, *Group Theory and Physics*, Cambridge University Press 1994.
- R. Gilmore, *Lie Groups, Physics, and Geometry: An Introduction for Physicists, Engineers and Chemists*, Cambridge University Press 2008.
- J. Schwichtenberg, *Physics from Symmetry*, Springer 2015.

Some online resources

- G. t'Hoof, `http://www.staff.science.uu.nl/~hooft101/lectures/lieg07.pdf` Very good and clear sets of lecture notes, for advanced undergraduates and beginning PhD. Chapter 6 of these lecture notes is based on it.
- P. Woit, `http://www.math.columbia.edu/~woit/QM/fall2012.html` For those who has finished this class, and wants to know how to look at QM from a group theory perspective.
- R. Earl, `http://www.maths.ox.ac.uk/courses/course/22895/material` The Oxford University class on group theory has lots of concise notes and problems which are suited for this class.

Contents

Chapter 1

Introduction to Symmetries

> *...the shapes of symmetries, which though broken, are exact principles governing all phenomena, expressions of the beauty of the world outside.*
> Steven Weinberg

Symmetries are our friends.

When you were learning dynamics of a particle of mass m moving under the influence of a central mass M located at the origin O at $\mathbf{x} = 0$, you wrote down a formula using Newton's Law of Gravity which looks like the following

$$\ddot{\mathbf{x}} = -\frac{GM}{r^2}\frac{\mathbf{x}}{|\mathbf{x}|} \tag{1.1}$$

where the radial coordinate r

$$r = \sqrt{x^2 + y^2 + z^2} = \sqrt{|\mathbf{x}^2|}. \tag{1.2}$$

We say that the Gravitational force, and hence its **potential** $U(\mathbf{x})$, is **spherically symmetric** around $\mathbf{x} = r = 0$, and this allows us to simplify the equation that we want to solve as you might recall. In this simple example, the **symmetry** in "spherically symmetric" refers to the fact that it doesn't matter which angle we are looking at the central mass. One way to think about it is that we can rotate the point mass in any of the elevation and azimuthal angles, and the dynamical problem remains exactly the same. In the viewpoint of the particle which is orbiting around the central mass, it means that the gravitational force that it feels only depends on the radial distance r and nothing else.

While we like the fact that this symmetry makes equations easier to solve, more importantly it actually buys us something crucial. Consider the Angular Momentum of a particle around O

$$\mathbf{L} = \mathbf{x} \times \mathbf{p},\ \mathbf{p} = m\dot{\mathbf{x}} = m\mathbf{v}. \tag{1.3}$$

Taking the time derivative of $\mathbf{L}$, we get

$$\begin{aligned} \frac{d\mathbf{L}}{dt} &= m\frac{d\mathbf{x}}{dt} \times \mathbf{v} + m\mathbf{x} \times \frac{d\mathbf{v}}{dt} \\ &= 0 + m\mathbf{x} \times \frac{\mathbf{F}}{m}. \end{aligned} \tag{1.4}$$

But $\mathbf{F} \parallel \mathbf{x}$ for spherically symmetric potential about O since the force must be pointing the same direction as the position vector, so the last term vanishes

$$\frac{d\mathbf{L}}{dt} = 0. \tag{1.5}$$

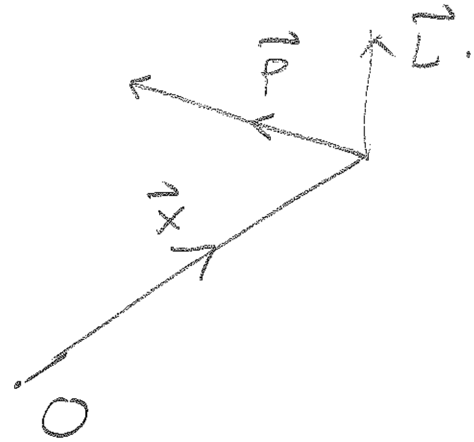


Figure 1.1: Classical Angular Momentum.

We can immediately integrate Eq. (1.5) to obtain $\mathbf{L} = \mathbf{C}$ where $\mathbf{C}$ is some constant vector. We call this constant a **constant of motion** and we say that angular momentum is a **conserved quantity** in the presence of a spherically symmetric potential. In other words, given a spherically symmetric potential like the Gravitational potential, the angular momentum of any particle cannot change without additional (non-spherically symmetric) external forces.

This notion that *symmetries lead to the existence of conserved quantities* is the most important thing[1] you will learn in these lectures (although not the only thing). The laws of conservations that you have studied – conservation of momentum, conservation of energy, conservation of electric charges, indeed conservation of X, are *consequences* of one or more existing symmetries that underlie the dynamical equations which describe them[2]. Indeed, you will find that the *quantum mechanical* notion of a particle and its charge – be it the electric charge or something more complicated like the color charges of quarks, arises from the fact that the equations that describe them[3] can be cast in the language of symmetries; Nature organizes herself using the language of symmetries. Hence if you want to study Nature, you'd better study the language of symmetries.

The mathematical structure that underlies the study of symmetries is known as **Group Theory**, and in this class we will be spending a lot of time studying it. As an introduction, in this chapter we will give you a broad picture of the concepts without going into too much mathematical details. Instead , we will rely on your innate intuition (I hope) to motivate the ideas, so don't worry if there seems to be a lot of holes in our study or when we make assertions without proofs. After this, we will come back and cover the same material in much greater detail, and perhaps reorganize your intuition a bit.

Finally, for many of you, this will be the first class that you will encounter what some people will term "formal mathematics" or "pure mathematics", at least in a diluted semi-serious manner. Indeed, the presentation is deliberately designed to be more formal than your usual physics class. Personally, I find the most difficult aspect of learning anything is to understand the *notation*, and formal math have a lot of those. In these lectures, you will sometimes find a bewildering amount of new mathematical notation that can look very intimidating. There is indeed a learning curve, perhaps even a steep one. Hopefully, once you get to grips with the notation, you will find that learning physics that uses fancy

[1]To be precise, this applies to *continuous* symmetries.

[2]It is important to remark that conservation laws do *not* follow the invariance of the potential of some hypothetical surface. In other words, a potential which possess an ellipsoidal equipotential surface do not necessary possess a conserved quantity associated with it.

[3]For the electric charge, the equations of electrodynamics, and for the color charges of quarks, Quantum Chromodynamics.

mathematics less daunting. This is particular true for those of you who are interested in pursuing a career in theoretical physics.

1.1 Symmetries of the Square

What is a **symmetry**? Roughly speaking, a symmetry is a property of some **object** which remains **invariant** under some **operations**. Note that the notion of symmetry requires both an object, and the operation which **acts** or **operates** on the object. **Invariance** means that the property of the object "remains the same" before and after the operation has been acted upon it. In the example of the spherically symmetric potential, the object is the "amplitude of the force or potential U" and the operation is "rotation" around the elevation and azimuthal angles. The proper term for the action is **operator** and for the object is **target**. All these are very abstract – and indeed I encourage you to start thinking abstractly immediately.

As a concrete example, let's consider the square. We will label the four corners of the square A, B, C, D. Intuitively, if we rotate the square by $90°$ clockwise, the square remains the same with the original orientation. What do we mean by "the same" – it means that *if we drop the labels* A, B, C, D*, then you won't be able to tell whether or not the square have been rotated by* $90°$ *or not*. We say that *the square is* **invariant** *under the* **operation** R, and we say that R is an **operator**, see Fig. 1.2. Another operation

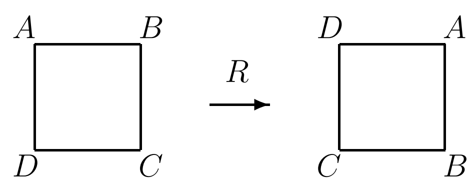


Figure 1.2: Rotation operation R: rotating the square by $90°$ clockwise.

which leaves the square invariant is the **reflection** about a vertical axis that goes through the geometric center of the square, let's call this operation m_1

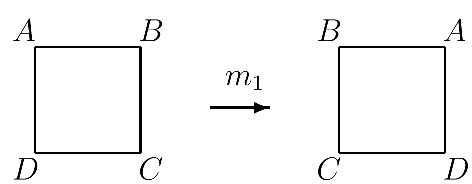


Figure 1.3: Reflection operation m_1: reflecting the square about a vertical axis which goes through the geometric center.

Now rotation about $180°$ also leaves the square invariant, let's call this operation T. But now if we allow ourselves to **compose** operations, by introducing the notion of "multiplication", i.e. for any two operations X,Y we can create the new operator Z, i.e.

$$X \cdot Y = Z \text{ or simply } XY = Z \tag{1.6}$$

where the "XY" means "operate Y and then X". We are very careful in defining the *order* of the operation as it is important as we will see below. Then it is intuitively clear that

$$T = RR = R^2. \tag{1.7}$$

And hence it follows that clockwise rotation by 270° would then be R^3. You can also check that the composition rule is **associative**, i.e.

$$X(YZ) = (XY)Z, \tag{1.8}$$

which you should convince yourself is true.

Of course, we are not restricted to rotating clockwise, but it's clear that rotating counterclockwise by 90° would be the same operation as R^3. If we now denote R^{-1} as rotation by 90° counterclockwise, then

$$R^{-1} = R^3 \tag{1.9}$$

and hence $R^{-1}R = e$ where e is the "do nothing" operator, which we also call the **identity**. R^{-1} is the **inverse** of R – let's formalize this notation by saying that X^{-1} is the inverse of X.

For reflections, in addition to m_1, there exist 3 other **axes of symmetries** about the geometric center which we will call m_2, m_3 and m_4, see Fig. 1.4 below. If we act on the square successively with the same

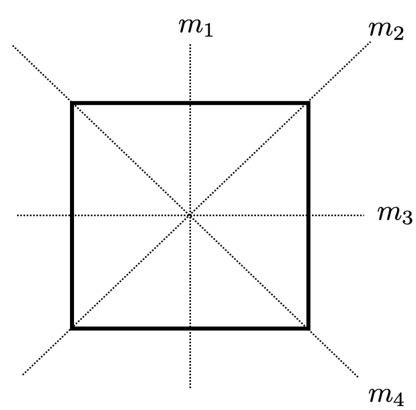


Figure 1.4: The axes of reflection symmetry of the square.

reflection operator, we have effectively done nothing. Mathematically

$$m_i m_i = m_i^2 = e \tag{1.10}$$

or we can also write, using

$$m_i m_i^{-1} = e \tag{1.11}$$

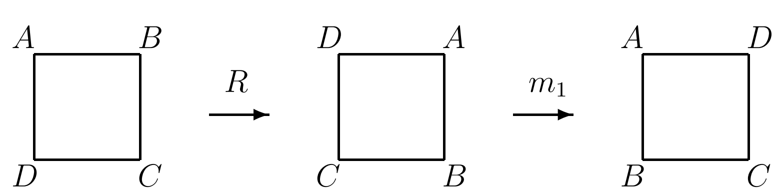
where $m_i = m_i^{-1}$. In other words, *reflections operators m_i are their own inverses.*

It turns out (which you can and should check) that the set of operations

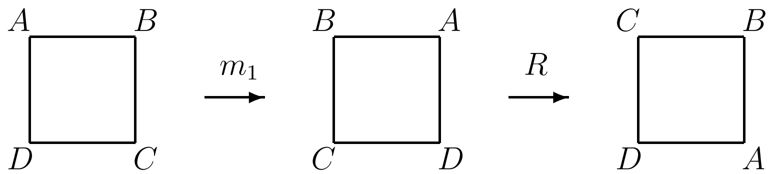
$$D_4 = \{e, R, R^2, R^3, m_1, m_2, m_3, m_4\}$$

is all the operations which will leave the square invariant; let's give this set a name D_4. Notice that the inverses are already part of the set, e.g. $R^{-1} = R^3$, $(R^2)^{-1} = R^2$ and of course the reflections are their own inverses.

When we discuss composition of operators Eq. (1.6), we note the importance of the ordering. Let's consider two operations m_1 and R. First we act on the square with R then m_1 to get



which we can compare to acting with m_1 first and then R which results in



Notice that the labels of the corners for the two compositions are different (even though both are squares). This imply that

$$m_1 R \neq R m_1. \tag{1.12}$$

We say that "m_1 and R **do not commute**". On the other hand, rotation operators R clearly **commute** with each other. Exercise: do reflection operators commute with one another?

Also, you can check that composition of *any* two operators from the set $\{e, R, R^2, R^3, m_1, m_2, m_3, m_4\}$ will result in an operator which also belongs to the set. You already have seen $RR = R^2$, $m_1^2 = e$, etc. You can also check that $m_1m_4 = m_2m_1 = m_3m_2 = m_4m_3 = R$, $m_1m_3 = m_2m_4 = m_3m_1 = m_4m_2 = R^2$, $Rm_1 = m_3$ etc. You cannot compose an operator which is not part of the set – it's like an exclusive club! We say that the "algebra is **closed** under the composition rule". Roughly speaking, an **algebra** is a set of objects (here this is the set of operators) equipped with a composition rule. We will formally define algebras and **closure** in the next Chapter 2.

Let's take stock. We have showed that for a square, there exists 8 operations

$$D_4 = \{e, R, R^2, R^3, m_1, m_2, m_3, m_4\}$$

which leaves the square invariant. We also note that we can define a composition rule $XY = Z$, which give us the following results

- The result of composition of two operators is also an operator in D_4, i.e. it is **closed**.
- Each element in D_4 has an **inverse** which is also belongs to D_4. The inverse for each element is unique.
- The composition rule is **associative** $(XY)Z = X(YZ)$.
- There is an identity e, where $eX = Xe = X$ for all elements X which belong to D_4. Furthermore, the identity is unique.

Any set of objects with a composition rule which obeys all the above **axioms** is called a **Group**. D_4 is called the **Dihedral-4 Group** – the 4 means that it is the symmetry group of the 4-gon. i.e. the square. It won't be a surprise to you that the symmetries of the triangle also forms a group called D_3. The symmetry group of the N-gon is D_N.

Remark: We have asserted above that the identity e and the inverses are unique, but has not proven it. We will come back to proofs when we study Groups in a more systematic way.

Since the number of elements of the group D_4 is 8, it is a **finite group**, and the number of elements is called its **order** and denoted $|D_4|$. Don't confuse the term order here with the "order" in "ordering" (i.e. which one goes first) – unfortunately sometimes mathematicians overused the same word and in fact this won't be the last thing we call order (sadly). On the other hand, the symmetry group of a circle is of infinite order, and forms a **continuous group** – can you guess what are the operations which leave the circle invariant? We will study both discrete and continuous groups in this class.

For a finite group whose order is small, sometimes it can be illuminating to construct a **multiplication table** or **Cayley Table** which tells us what are the results of the multiplication of two elements in the group. For D_4, it is the following:

$$\begin{array}{c|ccccccc} e & R & R^2 & R^3 & m_1 & m_2 & m_3 & m_4 \\ \hline R & R^2 & R^3 & e & m_2 & m_3 & m_4 & m_1 \\ R^2 & R^3 & e & R & m_3 & m_4 & m_1 & m_2 \\ R^3 & e & R & R^2 & m_4 & m_1 & m_2 & m_3 \\ m_1 & m_4 & m_3 & m_2 & e & R^3 & R^2 & R \\ m_2 & m_1 & m_4 & m_3 & R & e & R^3 & R^2 \\ m_3 & m_2 & m_1 & m_4 & R^2 & R & e & R^3 \\ m_4 & m_3 & m_2 & m_1 & R^3 & R^2 & R & e \end{array} \tag{1.13}$$

As a convention, the *column* goes first (i.e. to the right), and then the *row*. So $Rm_1 = m_2$ etc.

The study of the algebra of groups, as written above, is known as **Group Theory**.

Finally, before we end this section on the symmetries of the square, let's introduce a new concept. You might have noticed by now that while the set of the 8 operators are distinct, some of them are compositions of other operators. We can then ask the question: *what is the minimal number of operators we need to generate the entire set of D_4?*

For example, we have already being using $R^2 = RR$ to denote rotation about 180° – in words we **generate** R^2 from composing RR. We can also generate m_4 by first rotating by R and then reflecting with m_1, i.e. $m_4 = Rm_1$. Indeed we can generate the entire set of operators in D_4 using only two operators $\{R, m_1\}$ – which you can easily convince yourself. This minimal subset of operators are called **generators**, and we say that "the group D_4 is generated by the subset of generators $\{R, m_1\}$". Generators will play a crucial role in understanding the symmetries of physics.

1.2 Representations of Finite Groups

At this stage, we want to make an important remark. While we have motivated the study of groups using the symmetries of the square, *this is not necessary to study groups*. In other words, we can drop the square, and then study the properties of a set of 8 objects D_4 which possesses the algebraic properties as described above. This **abstraction** is a crucial part of our study of group theory, and *it is the reason why group theory so powerful in describing symmetries* – since once we know all the properties of a group, then *any physical system that possess the same symmetries will obey exactly the same underlying group-theoretic properties*. This paragraph is a bit mysterious for now, but we will crystalize it in due time. Let's plow on for now.

In the lectures so far, we have used the words "rotate" and "reflect", and intuitively we understand that we are describing the **action** of some **operators** on an object – here we this object is the square and we have drawn pictures (or visualize it in our heads) to describe the actions. However, drawing pictures while illustrative, is not very efficient nor very precise. What we want is a way to describe the square and the operations on it in a much more mathematical way. One possible way is the following

$$\text{result} = \text{operator} \times \text{object} \tag{1.14}$$

where the "action" **acts** or **operates** on the object *from the left*, i.e. left multiplication. As you have seen before, one can also define right multiplication as the action

$$\text{result} = \text{object} \times \text{operator}\ ,\ \ \text{(not used)} \tag{1.15}$$

but we will follow convention and use left multiplication.

Now, how do we describe the "object" mathematically? To be specific, how do we describe the square? It turns out that there are many ways, and not all of them will be useful to us. In these lectures, we will instead of interested in the case when the "object lives" in a **vector space**. For this reason, sometimes we call the objects being acted upon the **Target Space**. We will get to vector spaces in great detail in Chapter 4, but for now let's motivate it.

Consider the square again, and imagine we lay down a coordinate system with the usual x-axis and y-axis on it, with the origin $(0,0)$ in the geometric center of the square. Any point (x,y) can then be described by a vector $\mathbf{A}$ as you have studied in high school

$$\mathbf{A} = x\mathbf{i} + y\mathbf{j} \tag{1.16}$$

where $\mathbf{i}$ and $\mathbf{j}$ are the unit vectors pointing in the $+x$ and $+y$ directions respectively. This notation, while well loved, is actually quite cumbersome for our purposes. It will be more convenient to introduce the **matrix notation**, so the unit vectors can be written as

$$\mathbf{i} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \ , \ \mathbf{j} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{1.17}$$

so the vector $\mathbf{A}$ is

$$\mathbf{A} = x\begin{pmatrix} 1 \\ 0 \end{pmatrix} + y\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}. \tag{1.18}$$

We have used your familiarity with **vector addition** rules, i.e. any pair of vectors can be added together to yield a third vector, and the fact that a vector $\mathbf{A}$ can be multiplied by a number r to get a new vector $r\mathbf{A}$ **linearly** in the following sense $r(\mathbf{A}+\mathbf{B}) = r\mathbf{A} + r\mathbf{B}$. These rules (and a few others) define what we mean by the object "living", or belonging to a **vector space**. We will be a lot more formal in Chapter 4.

From Eq. (1.14), then it is clear that the symbol $\times$ denotes **matrix multiplication**. Furthermore, we *insist* that the "result" *must* also be a vector of the same form as $\mathbf{A}$, i.e.

$$\mathbf{A}' = x'\begin{pmatrix} 1 \\ 0 \end{pmatrix} + y'\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \text{operator} \times \mathbf{A}, \tag{1.19}$$

then for a valid matrix multiplication requires that the "operator" (let's call it M) be a 2×2 matrix of the following form

$$\textbf{operator } M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \tag{1.20}$$

where a,b,c,d are real numbers.

If we want the answer to the question "what happens when we act on the vector $\mathbf{A}$ with one of the elements of the set D_4", we then have to **represent** the elements of D_4 by 2×2 matrices. The identity e element leaves the square invariant, so if

$$\mathbf{A} = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \tag{1.21}$$

is the top left corner of the square, this means that e in matrix representation is simply

$$e = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{1.22}$$

which is of course the identity matrix. Of course, there are more than one possible 2×2 matrix which when acted on $\mathbf{A}$ returns $\mathbf{A}$. You can easily convince yourself that a square matrix of the form

$$\begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}$$

will also keep **A** invariant. But that's not surprising – there are 3 other points of the square

$$\mathbf{B}=\begin{pmatrix}1\\1\end{pmatrix}\ ,\ \mathbf{C}=\begin{pmatrix}1\\-1\end{pmatrix}\ ,\ \mathbf{D}=\begin{pmatrix}-1\\-1\end{pmatrix}\ ,$$

so that any matrix that represents the operators must also return to the right position, so we need to use this additional information to figure out what the matrix operator is. Let's consider the rotation R. When it acts on **A**, it must return **B**. Let's define R

$$R=\begin{pmatrix}0&1\\-1&0\end{pmatrix}. \tag{1.23}$$

Then when it acts on **A** it must return **B**, i.e. $R\mathbf{A}=\mathbf{B}$. We can check easily

$$R\mathbf{A}=R\begin{pmatrix}-1\\1\end{pmatrix}=\begin{pmatrix}1\\1\end{pmatrix}=\mathbf{B}\ , \tag{1.24}$$

and you can compute that $R\mathbf{B}=\mathbf{C}$, $R\mathbf{C}=\mathbf{D}$ and $R\mathbf{D}=\mathbf{A}$.

We can also represent the reflection m_1 operator as

$$m_1=\begin{pmatrix}-1&0\\0&1\end{pmatrix}\ , \tag{1.25}$$

and you can easily check that $m_1\mathbf{A}=\mathbf{B}$, $m_1\mathbf{B}=\mathbf{A}$, $m_1\mathbf{C}=\mathbf{D}$ and $m_1\mathbf{D}=\mathbf{C}$.

What about the rest of the operators? We can now use the fact that we can *generate* them using R and m_1, and calculate:

$$R^2=RR=\begin{pmatrix}0&1\\-1&0\end{pmatrix}\begin{pmatrix}0&1\\-1&0\end{pmatrix}=\begin{pmatrix}-1&0\\0&-1\end{pmatrix}\ ,\ R^3=RR^2=\begin{pmatrix}0&-1\\1&0\end{pmatrix}\ , \tag{1.26}$$

and

$$m_2=Rm_1=\begin{pmatrix}0&1\\1&0\end{pmatrix}\ ,\ m_3=R^2m_1=Rm_2=\begin{pmatrix}1&0\\0&-1\end{pmatrix}\ ,\ m_4=Rm_3=\begin{pmatrix}0&-1\\-1&0\end{pmatrix}. \tag{1.27}$$

In other words, the composition law of the group can be described by matrix multiplication too! Indeed, we use the word "multiplication" to describe composition of group operators just now, and when represented by matrices, the group operators literally are multiplied.

Since these are the elements of the group D_4 being represented by matrices, we say that these set of 8 matrices *and* their target space form a **Group Representation**, and more specifically we say that they "form a Linear Group Representation on a 2 dimensional real vector space". Of course, although the corners of the square clearly belong to this vector space, they are but a small subset of points of the (infinitely) many points on this vector space. Our operators D_4 are democratic – not only will they act on the corners of the square, they will also act on *any* vector in this space e.g.

$$R\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}y\\-x\end{pmatrix}\ ,\ \forall\ x,y\in\mathbb{R}. \tag{1.28}$$

We have introduced the notation $\in$, which means "in" or "belonging to", and $\mathbb{R}$ which denotes the set of all real numbers. In words we say "x and y in 'R'." We will come back and talk about sets and mathematical structure in the next Chapter 2.

1.2.1 Summary

We have introduced the idea of **symmetries** which are properties of some object which remain "the same" under some operation. Relying on your mathematical intuition, we showed that the set of operations which leave the object invariant naturally form an algebraic structure called a **Group**. Stripping away the object from the discussion, we assert that this algebraic structure can "lives on its own" and once we figure out all the algebraic properties this structure can be applied to any physical system or any object which possesses the same symmetry.

For example, we will show that the set of operations known as "rotation about an axis" will form a **continuous abelian group** called $U(1)$. Clearly, there are many physical systems which are invariant under rotation about an axis – e.g. the Coulomb potential, or more interestingly, the *gauge field* of the electromagnetic field. Indeed, for half our lectures, we will cast away the "extra" baggage of the "object" and study Groups on their own right.

However, since we are physicists, we ultimately want to study actual physical systems – the "objects" such as the square. In this chapter, we show that if these objects can be **represented** by vector spaces, then the group elements can be represented by matrices. Collectively, the vector spaces of the objects and the matrices representation of the group elements are called **Group Representations**.

Finally, focusing on particle physics, we wade into the study of **Lie Groups** and **Lie Algebras** – one of the key pillars of modern physics. You might have heard the words "The Standard Model of Particle Physics is $SU(3) \times SU(2) \times U(1)$". What this means is that the particles can be represented by vector spaces (think of them as column matrices) which are operated upon by the group elements of $SU(3) \times SU(2) \times U(1)$. This may look mysterious to you – what do we mean by $SU(3)$ and by "$\times$"?

The goal of this set of lectures is to demystify that statement. Of course, we will not be studying particle physics. However, as a first course in symmetries, we aim to build up your mathematical vocabulary of the language of symmetries which you can bring into your more advanced classes.

Chapter 2

Mathematical Preliminaries

I speak from some sort of
protection of learning,
Even tho' I make it up as I go on.

Yes, *New Language*

In this section we will review some basic mathematics. The goal of this section is to (re)-introduce you to some mathematical language so that you become comfortable with thinking abstractly, and on the flipside, to challenge you into rethinking some things which you might have gotten "comfortable" with. We will not be rigorous in our treatment of the mathematics as this is a course in physics. Having said that, we will be very deliberate in using faux math language, and sound a bit pretentious along the way.

2.1 Sets

A **set** is a collection of objects. For example

$$S = \{a, b, c\} \tag{2.1}$$

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 in S". The simplest set is the **empty set** and is denoted by

$$\emptyset \tag{2.2}$$

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 integers[1]

$$\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\} \tag{2.3}$$

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

$$\mathbb{N} = \{x | x \in \mathbb{Z}, x \geq 0\} \ , \tag{2.4}$$

so $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$.

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

[1] $\mathbb{Z}$ is for *Zahlen*, German for "integers".

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

$$\mathbb{Z} = \{\ldots, -4123, 69, 794, 0, 66, -23, \ldots\}. \tag{2.5}$$

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-theoretic[2] 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

$$A \subseteq B \tag{2.6}$$

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

$$\{a, b, c\} = \{a, a, a, a, b, c, c, a\} = \{a, b, c, c\} \tag{2.7}$$

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

$$0 = \emptyset \ , \ 1 = \{\emptyset\} \ , \ 2 = \{\emptyset, \{\emptyset\}\} \ , \ 3 = \{\emptyset, \{\emptyset, \{\emptyset\}\}\} \tag{2.8}$$

[2]You 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. It's good to think of "obvious" ideas in not so obvious ways.

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 $0 \leq 1 \leq 2 \ldots$, or you can also write $0 < 1 < 2 \ldots$ by replacing $\subseteq$ with $\subset$ (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 an 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.

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

$$A \cup B = \{x | x \in A \text{ or } x \in B\}. \tag{2.9}$$

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

$$A \cap B = \{x | x \in A \ , \ x \in B\}. \tag{2.10}$$

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**

$$A \cup B = B \cup A \ , \ A \cap B = B \cap A. \tag{2.11}$$

This looks like the commutativity of the "+" operator you have seen before $x + y = y + x$ in high skool. So you can think of a union as "addition".

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

$$(x + y) - (y + x) = 0. \tag{2.12}$$

We have not invented the notion of the operation of "subtraction" for sets. Let's invent it now using properties

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

$$A \setminus B = \{x | x \in A, x \notin B\}. \tag{2.13}$$

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

$$(A \cup B) \cup C = A \cup (B \cup C)\ ,\ (A \cap B) \cap C = A \cap (B \cap C). \tag{2.14}$$

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

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)\ ,\ A \cup (B \cap C) = (A \cup B) \cap (A \cup C). \tag{2.15}$$

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

$$\mathcal{P}(A) = \{x | x \subseteq A\}. \tag{2.16}$$

So if $A = \{a, b\}$, then $\mathcal{P}(A) = \{\emptyset, \{a\}, \{b\}, \{a, b\}\}$.

2.1.2 Cartesian Product and Coordinate Systems

Finally, intuitively, we know that pairs of objects exist – say for example a coordinate system (x, y) or $(Justin, Selena)$ 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

$$A \times B = \{(a, b) | a \in A, b \in B\} \tag{2.17}$$

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 $a = b$ and $c = d$ (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

$$\{(x, y) | x \in \mathbb{R}, y \in \mathbb{R}\}. \tag{2.18}$$

Since x and y are both in $\mathbb{R}$, we often write such a coordinate system in shorthand as

$$(x, y) \in \mathbb{R} \times \mathbb{R} \tag{2.19}$$

or even shorter hand by

$$(x, y) \in \mathbb{R}^2. \tag{2.20}$$

You can keep making ordered "triplets". For example, suppose A, B and C are three sets

$$B \times C \times A = \{(b, c, a) | b \in B, c \in C, a \in A\}. \tag{2.21}$$

We can also make ordered pairs from ordered paris. For example, if $B \times C$ is an ordered pair, and A is another set, then

$$(B \times C) \times A = \{((b, c), a) | b \in B, c \in C, a \in A\}. \tag{2.22}$$

In general, $(B \times C) \times A$ is not the same as $B \times (C \times A)$, and both will not the same as $A \times B \times C$ because the bracket structure will be different.

A three dimensional coordinate system is then $\mathbb{R}^3$ etc. These are known as **Cartesian Coordinate Systems**. In general, an ordered list $(a, b, c, \dots)$ is called a n**-tuple**. So an ordered triplet would be a 3-tuple.

2.2 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

$$f : A \to B; \text{(Rules)}. \tag{2.23}$$

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

$$f : A \to B;\ f : x \mapsto x^2\ \forall\ x \in \mathbb{R} \tag{2.24}$$

where we have distinguished the arrows $\to$ 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 $\to$. 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

$$f : A \to B;\ f(x) = x^2\ \forall\ x \in \mathbb{R}. \tag{2.25}$$

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 $x^2 \geq 0$, $f(x)$ maps only to the **positive semi-definite** part[3] of B. 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 $f : A \to B$ is a map, then the image of f is

$$\text{im}(f) = \{x | x \in B, x = f(a)\ \forall\ a \in A\}, \tag{2.26}$$

or even shorter form

$$\text{im}(f) = \{f(a) | a \in A\}. \tag{2.27}$$

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,

$$\text{Id} : A \to A;\ \text{Id}(a) = a\ \forall\ a \in A. \tag{2.28}$$

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).

[3] The range of B which is $x > 0$ (not including zero) is called the **positive definite** range.

2.2.1 Surjective, Injective and Bijective

In the map $f(x) = x^2$ we described above, there exists points in cod(f) that has no "partner" in 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.

$$\text{im}(f) = \text{cod}(f) \tag{2.29}$$

is called a **surjective** (Latin) or **onto** (English) or **epic** (Hellenic) map. I personally prefer "onto" but epic is quite epic. If im(f) is continuous[4] then such a map is also called a **covering**. We will encounter covering maps when we discuss Lie groups in the future.

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

$$g : A \to B;\; a, a' \in A \text{ and if } a \neq a' \text{ then } f(a) \neq f(a'). \tag{2.30}$$

Such a map is called an **injective** (Latin) or **one-to-one** (or **into**) (English) or **monic** (Hellenic) map. I personally prefer one-to-one, since it is rather descriptive.

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)

$$\begin{aligned} \text{dom}(f) &= \text{dom}(g), \\ \text{cod}(f) &= \text{cod}(g), \\ \forall x \in \text{dom}(f)\ ,\ f(x) &= g(x). \end{aligned} \tag{2.31}$$

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 dom(f) $= \mathbb{Z}$ while dom(g) $= \mathbb{R}$, $f \neq g$.

2.2.2 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 $f : A \to B$ and $g : B \to C$, then the composition map $g \circ f$ is described by

$$g \circ f : A \to C;\; g(f(x))\ \forall x \in A. \tag{2.32}$$

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 $h : C \to D$ we can construct

$$h \circ g \circ f : A \to D;\; h(g(f(x)))\ \forall x \in A. \tag{2.33}$$

We will now state three theorems.

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

Proof: Suppose $f : A \to B$ and $g : B \to C$ 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 $g(f(a)) \neq g(f(a'))$. $\square$.

[4] We have not defined continuous, but now we are going to accelerate and rely on your intuition.

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

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

Proof: Suppose $f : A \to B$ and $g : B \to C$ 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 $g(b) = c \ \forall c \in C$. Now f is also surjective, so $\exists a \in A$ such that $f(a) = b \ \forall b \in B$, hence $g(f(a)) = c$, i.e. there exist a map from a to *every* element in C hence $g \circ f$ is surjective. □.

(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 $f : X \to Y$. 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, we can't find an inverse because it is not unique. 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 $x \in X$, there is no "partner" in X to map to.

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 $g(f(x)) = \mathrm{Id_X}$ (where we have appended the subscript X to indicate that it is the identity map in X). Formally,

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

f is bijective $\Leftrightarrow$ There exists an inverse map $g : Y \to X$ such that $f \circ g = \mathrm{Id_Y}$ and $g \circ f = \mathrm{Id_X}$.

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.

2.2.3 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

$$\mathrm{map}_{A,B} = \{f | f : A \to B\} \tag{2.34}$$

so $\mathrm{map}_{A,B}$ 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

$$\mathrm{bij}_{A,B} = \{g | g \in \mathrm{map}_{A,B}\ ,\ g \text{ is bijective}\}. \tag{2.35}$$

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

2.3 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. Consider the set of all animals, let's call it P. Let's say we want to define an **equivalence relationship**, let's call it $\sim$, based on the animal species. For example, we want to be able to say things like "the chihuahua is equivalent to the bulldog" because both chihuahuas and bulldogs are dogs so Chihuahua$\sim$bulldog. Naturally, bulldog$\sim$chihuahua (symmetric), and furthermore pomeranian$\sim$chihuahua and hence bulldog$\sim$pomeranian (transitive). Obviously Chihuahua$\sim$Chihuahua.

(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" , and "with glasses" and "without glasses", in the set of members of the Symmetric in Physics students.)

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 fruit types" partitions fruits into "apples" subset A, "oranges" subset B, "pears" subset C etc, and it's clear that $A \cap B = \emptyset$ etc. 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 people with glasses and people with no glasses. 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.

2.3.1 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 $a + a = 2a$, $a \cdot (b + c) = a \cdot b + a \cdot c$ 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

$$\text{or } s_1 \star s_2 \text{ or } s_1 \cdot s_2 \text{ or simply } s_1 s_2. \tag{2.36}$$

We say that we want a **product** or **composition** rule[5].

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

$$s_1 s_1 = s_1 \ , \ s_1 s_2 = s_2 \ , \ s_2 s_1 = s_2 \ , \ s_2 s_2 = s_1. \tag{2.37}$$

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. (2.37) *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 the specifying certain **algebraic axioms**. There are several well known algebras, and we will discuss two of them in the following.

2.3.2 Fields and its Algebra

Instead of exact rules, there are some general classification of rules which we can list down. As a specific example, consider something you learned in high school: the algebra of a set S equipped with two binary operators **addition** $+$ and multiplication $\cdot$ which have the following rules (for any $a, b, c, \dots \in S$).

Field Axioms:

- **Closure**: $a \star b \in S$ for $\star = \{\cdot, +\}$.

[5] You have seen the word "composition" before when we discuss maps – this is of course not coincidental, as the elements in a set can be maps.

- **Commutative** if $a \star b = b \star a$ for $\star = \{\cdot, +\}$.
- **Associative** $a \star (b \star c) = (a \star b) \star c$ for $\star = \{\cdot, +\}$.
- **Distributive** $a \cdot (b + c) = a \cdot b + a \cdot c$. (But not the other way : $a + (b \cdot c) \neq (a + b) \cdot (a + c)$.)
- **Identities**: $a + 0 = a = 0 + a$ (0 is the **additive identity**) and $a \cdot 1 = a = 1 \cdot a$ (1 is the **multiplicative identity**).
- **Inverses**: $a + (-a) = 0 = (-a) + a$ (**additive inverse**) and $a \cdot a^{-1} = 1 = a^{-1} \cdot a$ for $a \neq 0$ (**multiplicative inverse**).

Notice that the Identities for $+$ and $\cdot$ are written down as different elements called 0 and 1 – and in general they are different (think of $S = \mathbb{R}$ the set of real numbers) but they can be the same.

Now you may legitimately ask (and if you do, pat yourself at the back for thinking abstractly): what are these new symbols $^{-1}$ and $-$ that has crept in? One way to think about them is that they are simply labels, so $-a$ can be simply a "different" element which, when added to a give us the element 0. We say that $-a$ is the **additive inverse** of a, and a^{-1} is the **multiplicative inverse** of a.

A set S which possesses the binary operators which obey the above rules is called a **Field** by mathematicians. You are no doubt familiar with it when $S = \mathbb{R}$, i.e. the set of real numbers, and $S = \mathbb{C}$ the field of complex numbers.

2.3.3 Groups and its Algebra

As you can see above, Fields have a fair amount of algebraic structure built into it. It turns out that we can perfectly invent an algebra with less rules. One such algebra is that of a **Group**, which we now state as follows.

(Definition) Groups: A group G is a set of elements $G = \{a, b, c, \dots\}$ with a composition law (multiplication) which obeys the following rules

- **Closure**: $ab \in G \ \forall a, b, \in G$.
- **Associative** $a(bc) = (ab)c \ \forall a, b, c \in G$.
- **Identity**: $\exists\ e$ such that $ae = ea = a$. e is called the *identity element* and is *unique*[6].
- **Inverses**: For every element $a \in G \ \exists\ a^{-1}$ such that $a^{-1}a = aa^{-1} = e$.

Notice that by this set of rules, the identity element e *must* be its own inverse, i.e. $e = e^{-1}$ – can you prove this? Also, an important formula is

$$\boxed{(ab)^{-1} = b^{-1}a^{-1} \ \forall\ a, b \in G}. \tag{2.38}$$

Proof: Since $ab \in G$, there is an inverse $(ab)^{-1}$ such that $(ab)^{-1}(ab) = e$. Multiplying from the right by b^{-1} we get $(ab)^{-1}a(bb^{-1}) = b^{-1}$, and then multiplying from the right again by a^{-1} we obtain Eq. (2.38) $\square$.

Comparing this to Fields, we have chugged away one of the binary operator, and dropped the Commutativity and Distributive axioms. If we restore commutativity, we get a very special kind of group called **Abelian Groups**.

We will be spending the rest of the lectures studying Groups, and why they are so important in describing symmetries in physics. But let us end with an example. Recall the set $S = \{s_1, s_2\}$ which

[6] Uniqueness is usually not stated as part of the axioms, but actually proven if you want to be completely hardcore about things.

we wrote down above with the composition rules Eq. (2.37). Let's give it a special name $Z_2 = \{s_1, s_2\}$, which forms a group as we can now prove.

Proposition: Z_2 with composition rules Eq. (2.37) forms a group.
Proof:

- **Closure**: it's clear that Eq. (2.37) imply that the group is close under the composition rule.
- **Associative**: You can show element by element that $(s_i s_j)s_k = s_i(s_j s_k)$ for $i, j, k = 1, 2$. Example: $(s_1 s_2)s_1 = s_2 s_1 = s_2$ and $s_1(s_2 s_1) = s_1 s_2 = s_2$ etc.
- **Identity**: Since $s_1 s_2 = s_2$ and $s_1 s_1 = s_1$, and we have no other elements, we identify $s_1 = e$ as the Identity element.
- **Inverse**: s_1 is the identity hence is its own inverse i.e. $s_1 s_1 = s_1$. Now we need to find an element a the inverse for s_2, i.e. $s_2 a = a s_2 = s_1$, which by eye we can see that's $a = s_2$, i.e. s_2 is its own inverse.

and since the set with the multiplicative rule obey all the Group axioms, it forms a group algebra. $\square$.

Can you see that Z_2 is also an Abelian group?

Chapter 3

Discrete Groups

Five bananas. Six bananas.
SEVEN BANANAS!

Count von Count, *Sing yourself Silly*

In this Chapter, we will begin our proper study of Group Theory. Groups can roughly be divided into whether they are **discrete** or **continuous**. Roughly speaking, discrete groups have finite number of elements or infinite but **countable** number of elements, while continuous groups are "continuously infinite" (which is at the moment are just words). We will focus on discrete groups in this Chapter, focusing on getting the definitions and jargon right. And then we will study several important discrete groups to get the basic ideas on how to manipulate them.

Let's now get back to the notion of *countability* we mentioned.

(Definition) Countability: A set S is said to be *countable* if there exists a bijective map $f : S \to \mathbb{N}$, 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. $0, 1, 2, 3, \ldots$ and you will be sure that you have not missed any numbers in between[1]. After all, we are taught to count in terms of the Natural numbers!

Figure 3.1: The Count of Sesame Street likes to count the natural numbers, which is countably infinite.

[1] Try "counting" the real numbers and you will see what I mean.

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 *infinitisimal distances*. In fact, since $\mathbb{N}$ does not possess such a structure, it does not make sense to talk about continuity. To discuss it, we need additional mathematical structure, and so we'll postpone the discussion until Chapter 4.

Let's begin by restating the group axioms and clean up on some loose threads from the last Chapter.

(Definition) Groups: A group G is a set of elements $G = \{a, b, c, \dots\}$ with a group composition law (or simply group law for short) which obeys the following rules

- **Closure**: $ab \in G \ \forall a, b, \in G$.
- **Associative** $a(bc) = (ab)c \ \forall a, b, c \in G$.
- **Identity**: $\exists \ e$ such that $ae = ea = a$. e is called the *identity element* and is *unique*[2].
- **Inverses**: For every element $a \in G \ \exists \ a^{-1}$ such that $a^{-1}a = aa^{-1} = e$. The inverse is unique.

We will prove the two statements cited above.

Uniqueness of identity e. Suppose there exist two elements e and e' with the property of the identity, i.e.

$$ae = ea = a \ , \ ae' = e'a = a. \tag{3.1}$$

Let $a = e'$, for the first equation we get $e'e = ee' = e'$. Let $a = e$, for the second equation we get $ee' = ee' = e$. Since the LHS of both equations are the same, $e = e'$ and hence the identity is unique $\square$.

Uniqueness of inverses. Suppose that h and k are inverses of the element g, then by the axioms

$$gh = e \ , \ kg = e \tag{3.2}$$

but now $k(gh) = k$ and by the associativity axiom we shift the brackets to get $(kg)h = k$ or using the 2nd equation above $eh = k$, we obtain $h = k$ $\square$.

Now that we have come clean on our assertions, we are ready to discuss some groups.

3.1 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 $|G| < \infty$.

3.1.1 Order 2 Group Z_2

We have seen this group in Chapter 1 before. As it turns out, it is the *only* possible order two group.

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

[2] Uniqueness is usually not stated as part of the axioms, but actually proven if you want to be completely hardcore about things.

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 $\Box$.

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.

$$P\psi(x) = \psi(-x). \tag{3.3}$$

We can also consider the "do nothing operator", let's call it e,

$$e\psi(x) = \psi(x). \tag{3.4}$$

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

$$P(P\psi(x)) = P\psi(-x) = \psi(x) \tag{3.5}$$

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.

$$P\uparrow=\downarrow \ , \ P\downarrow=\uparrow \tag{3.6}$$

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

$$e\downarrow=\downarrow \ , \ e\uparrow=\uparrow . \tag{3.7}$$

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

$$\uparrow=\begin{pmatrix} 1 \\ 0 \end{pmatrix} \ , \ \downarrow=\begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{3.8}$$

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

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \ , \ e = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{3.9}$$

and you can check that Eq. (3.6) and Eq. (3.7) are obeyed. We say that the matrices Eq. (3.8) and Eq. (3.9) form a Linear Group Representation of Z_2. We threw in "linear" because it is clear that $P(\uparrow + \downarrow) = P\uparrow + P\downarrow$ etc. Notice that the matrices $\uparrow$ and $\downarrow$ form a **Vector Space**. We will have a bit more to say about Vector Spaces when we study representation theory in Chapter 4, but in the meantime it shouldn't be hard to convince yourself that it obeys the usual rules you know and love from thinking in terms of a 2-dimensional vector $\mathbf{A} = x\mathbf{i} + y\mathbf{j} = x\uparrow + y\downarrow$ etc.

Of course, now that we have represented P in terms of 2×2 matrices, there is nothing to stop us from operating it on some other 2×1 column matrix e.g.

$$P\begin{pmatrix} 5 \\ -3 \end{pmatrix} = \begin{pmatrix} -3 \\ 5 \end{pmatrix} \tag{3.10}$$

and so on, just like in the case of the D_4 group we studied in Chapter 1. The take home message is that *groups exists on its own without the requirement of a symmetry (or symmetries) in the underlying space.* Having said that, it is when the underlying space possess symmetries, its power become magnified.

3.1.2 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 $a^2 b = eb = b$, 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 $a^2 a^{-1} = aa^{-1} = e$, 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 symmetricity 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

e	a	b
a	b	e
b	e	a

(3.11)

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

3.1.3 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

$$D_4 = \{e, R, R^2, R^3, m_1, R^2m_1, R^3m_1, Rm_1\}. \tag{3.12}$$

Example: Consider the Z_3 group. We can see that $a^2 = b$, so we can generate b from a, and hence the generating set of Z_3 is $\{a\}$.

The second example above can be generalized to n number of elements.

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

$$Z_n = \{e, a, a^2, a^3, \ldots, a^{n-1}\} \ , \ a^n = e. \tag{3.13}$$

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

$$gf = fg \tag{3.14}$$

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 "look" 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 $\star$. They are said to be *isomorphic*, and written $G_1 \cong G_2$, if there exist a bijective map $i : G_1 \to G_2$ such that it preserves the group composition law in the following way

$$i(g_1 \cdot g_1') = i(g_1) \star i(g_1') \ , \ \forall \ g_1, g_1' \in G_1. \tag{3.15}$$

Since i is a bijection, the following relation follows

$$i^{-1}(g_2 \star g_2') = i^{-1}(g_2) \cdot i^{-1}(g_2') \ , \ \forall \ g_2, g_2' \in G_2. \tag{3.16}$$

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

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 structure as that of G_1, and hence it "doesn't matter" which set of elements you choose to work with, all that has changed between G_1 and G_2 is the names you call the elements. This is clearer with an example.

Example: Let $N_n = \{0, 1, 2, 3, \ldots, n-1\}$ 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

$$i : N_n \to Z_n \ ; \ i(k) = a^k. \tag{3.17}$$

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*, and denoted $|g|$ (c.f. the notation to the order of the group $|G|$) 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.

3.1.4 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*[3].

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

$$a^2 = b^2 = c^2 = e \ , \ ab = ba = c \ , \ bc = cb = a \ , \ ca = ac = b \tag{3.18}$$

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 $\star$. The *product* (or sometimes *Cartesian product*) of G_1 and G_2 is defined to be the set

$$G_1 \times G_2 = \text{The set of all possible pairs } (g_1, g_2) \text{ where } g_1 \in G_1 \ , \ g_2 \in G_2 \tag{3.19}$$

[3]Literally, German for "four-group", and it seems that calling things in German make you sound more high-class, maybe except to the Germans.

with the new group composition law

$$(g_1, g_2)(g'_1, g'_2) = (g_1 \cdot g'_1, g_2 \star g'_2) \tag{3.20}$$

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

$$Z_2 \times Z_2 \cong V_4. \tag{3.21}$$

Let's prove this statement by explicitly constructing the bijective map. On the left of Eq. (3.21), 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

$$Z_2 \times Z_2 = \{(e_1, e_2), (e_1, \mu_2), (\mu_1, e_2), (\mu_1, \mu_2)\}. \tag{3.22}$$

The group composition law is given by the rule Eq. (3.20), which depends on the composition laws of the individual component group Z_2 i.e. $\mu_i^2 = e_i$ for $i = 1, 2$ etc. There are a lot of terms, so for convenient, let's construct a multiplication table

$$\begin{array}{c|ccc} (e_1, e_2) & (e_1, \mu_2) & (\mu_1, e_2) & (\mu_1, \mu_2) \\ \hline (e_1, \mu_2) & (e_1, e_2) & (\mu_1, \mu_2) & (\mu_1, e_2) \\ (\mu_1, e_2) & (\mu_1, \mu_2) & (e_1, e_2) & (e_1, \mu_2) \\ (\mu_1, \mu_2) & (\mu_1, e_2) & (e_1, \mu_2) & (e_1, e_2) \end{array} \tag{3.23}$$

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

$$i : Z_2 \times Z_2 \to V_4 \ ; \ (e_1, e_1) \mapsto e \ , \ (e_1, \mu_2) \mapsto a \ , \ (\mu_1, e_2) \mapsto b \ , \ (\mu_1, \mu_2) \mapsto c \tag{3.24}$$

then you can check that the group composition of the elements in Eq. (3.24) is the same as the V_4 composition laws Eq. (3.18), 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 $(g_1, g_2, g_3)(g'_1, g'_2, g'_3) = (g_1 g'_1, g_2 g'_2, g_3 g'_3)$ etc. We have thus taken the first steps in parsing the statement "The Standard Model is described by the group $SU(3) \times SU(2) \times U(1)$".

3.1.5 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 $h_1 h_2 = g$ where $g \in G$. In general g does not have to belong to H. However, suppose that we can find a subset H where $h_1 h_2 \in H$ $\forall h_1, h_2 \in H$, 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**[4] 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

[4] If 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.

- The identity of G is in H. (Can you see why?)
- For every $h_1, h_2 \in H$, $h_1 h_2 \in H$. 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 if 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 new term, **proper subgroups** of G, which is the set of all possible subgroups of G minus the trivial subgroup and G, and denoted $H \subset G$.

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 $H \subset G$. 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 $a^2 a^2 = a(a^3) = ae = a \notin \{e, a^2\}$ 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, $H = \{e, a, a^2\}$, $H = \{e, a, a^3\}$, $H = \{e, a^2, a^3\}$ – the first and third you can quickly rule out using the requirement that all the elements must be closed under inversion, while for the second set $a^3 a^3 = a^2$ is not in the set so hence all three are not subgroups. 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 $a^2 a^2 = e$, 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 $G = \{e, g_1, g_2, \ldots\}$ is a group and $X = \{x_1, x_2, x_3, \ldots\}$ 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.

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

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 $g_i^{-1} : X \to X$ also, this mean that the maps $g_i : X \to X$ must be *bijective*.

Instead of the clunky $g_1 : X \to X$ 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 $g_i(x)$, $i = 0, 1, 2, \ldots$, which is a perfectly fine notation. Instead of labeling with i, we write

$$\text{Acting on single element } g_i(x) = L_{g_i}(x) \text{ or } g(x) = L_g(x) \tag{3.25}$$

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

$$g_1 g_2 = L_{g_1} L_{g_2} = L_{g_1 g_2}. \tag{3.26}$$

Sometimes, instead of just an elemenet 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

$$\text{Acting on Sets of elements } L_g(Y) = \{L_g(x_1), L_g(x_2)\}. \tag{3.27}$$

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.

$$\boxed{O_x = \{L_g(x), \forall\, g \in G\}}. \tag{3.28}$$

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.

3.2.2 Action of Group on Itself, Partitioning

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

$$L_g(g'). \tag{3.29}$$

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

$$\boxed{L_g(g') = gg'}. \tag{3.30}$$

With left, there is right:

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

$$\boxed{R_g(g') = g'g}. \tag{3.31}$$

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.

$$L_g : G \to G. \tag{3.32}$$

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)

$$L_g(H) = H' \subseteq G, \tag{3.33}$$

where H' is some other subset of G. H' is also termed the **orbit** of the element g on H, and sometimes written $O_H = H'$. As an explicit example, suppose $G = \{e, g_1, g_2, \ldots\}$ 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 it 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 $H = \{e, h_1, h_2, \ldots\}$ 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.

$$\boxed{L_g(H) = gH \equiv \{ge, gh_1, gh_2, \ldots\}.} \tag{3.34}$$

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

$$\boxed{R_g(H) = Hg \equiv \{eg, h_1g, h_2g, \ldots\}.} \tag{3.35}$$

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 $h_1, h_2 \in H$, such that $g_1 = g_2h_2h_1^{-1}$. Since from closure of H, $h_2h_1^{-1} \equiv h_3 \in H$, 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*: From $g_1^{-1}g_1H = g_1^{-1}g_2H$, then $H = g_1^{-1}g_2H$ and hence $g_1^{-1}g_2 \in H$. 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 $g \in G$ belong 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 left[5] 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 remember 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 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*

$$\boxed{|G| = |G/H||H|.} \tag{3.36}$$

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

$$|G| = \text{Total Disjoint cosets } |G/H| \times \text{Cardinality of each coset } |H| \tag{3.37}$$

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 the identity and an element whose inverse is itself, i.e. $Z_2 = \{e, m_i\}$ and $Z_2 = \{e, R^2\}$ with $|Z_2| = 2$. Clearly, both 4 and 2 subdivides 8.

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

3.2.3 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

$$\boxed{\text{Conjugation } L_g(g') \equiv gg'g^{-1}}. \tag{3.38}$$

(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

$$L_g(g_1) = g_2 \text{ , i.e. } g_2 = gg_1g^{-1} \text{ for any } g \tag{3.39}$$

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.

Conjugacy allows us to define an extremely powerful and important type of subgroups as follows.

(Definition) Normal Subgroups: A *normal subgroup* H of G is a subgroup whose elements are conjugate to each other. Suppose H is a normal subgroup of G, then

$$L_g(h) = ghg^{-1} \in H \ \forall \ h \in H \text{ and } g \in G \ . \tag{3.40}$$

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

$$gH = Hg \ . \tag{3.41}$$

You can show this easily by noting that since $L_g(h) = ghg^{-1} \forall \ h \in H$, then $gHg^{-1} \in H$, or $gH = Hg$. 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 **quotient**[6]

[6]Unfortunately this is a reuse of both notation and terminology for subtraction from the set theory – however, if you think hard about it, it will actually eventually make sense.

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

$$G/H = \{gH; g \in G\}. \tag{3.42}$$

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 (inherity from G), we can rewrite

$$\begin{aligned}
(g_1H)(g_2H) &= g_1(Hg_2)H \\
&= g_1(g_2H)H \\
&= (g_1g_2)\underbrace{HH}_{H} \\
&= (g_1g_2)H \\
&= g_3H \in G/H
\end{aligned} \tag{3.43}$$

- (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 is[7]

$$|M| = 2^{46} \cdot 3^{20} \cdot 5^9 \cdot 7^6 \cdot 11^2 \cdot 13^3 \cdot 17 \cdot 19 \cdot 23 \cdot 29 \cdot 31 \cdot 41 \cdot 47 \cdot 59 \cdot 71. \tag{3.44}$$

3.3 Dihedral D_n Group

In the Introduction Chapter 1, we discussed the symmetries of the square in the context of the Dihedral-4 group D_4. But obviously, there exist an entire class of polygons each with its own symmetry group. Indeed, the word "Dihedral" means "two-sided", referring to the fact that not only we can rotate the polygon, we can *flip* over it around an axis through the center – that's the reflection action of course.

[7]A further amazing fact is that the dimensions of the irreducible representations of M turns out to be related to the coefficients of a fourier expansion of some obscure modular function, the proof which actually requires knowledge of string theory. A lot of exciting conjectures are spewing out of this curious fact – a fact so crazy that it is now known as "Monstrous Moonshine".

Dihedral groups are part of an ever larger class of groups called **Point Groups** since all the symmetry operations is pivoted around a single point (the geometrical center).

When we study D_4, we showed that it is generated by two elements $\{R, m_1\}$. We could easily replace m_1 with any of the other reflection operators, so let's drop the subscript 1. Now given an n-gon, we can define a clockwise rotation about $2\pi/n$ by an operator R, and then we can generate all the possible rotations using this single generator. Given this, it is not hard to see that any D_n group are generated by just this operator and a reflection m, $\{R, m\}$, just like D_4. Note that you can generate the entire group with two different reflection operators (m_i, m_j) for $i \neq j$ – you will show this in a homework problem. But for now, let's consider the rotations and reflections separately.

- **Rotation**: Any dihedral group have a subset of operators consisting of the identity and the rotation operators,

$$\{e, R, R^2, R^3, \ldots, R^{n-1}\} \subseteq D_n \tag{3.45}$$

which looks like the cyclic group we studied in section 3.1.3, which indeed it is so the Cyclic group Z_n is a subgroup of D_n. As Z_n is generated by an element $\{a\}$ of order n, the *cyclic subgroup* of D_n is generated by the single rotation operator R of order n.

Since all rotations are generated by the single element R, it is clear that they all commute with each other so this subgroup is Abelian (as Z_n is also Abelian).

If n is even, then $n/2$ is an integer and hence the operator $R^{n/2}$ exists. This is simply a rotation "half-way", or 180°. Despite its humble origins, this operator will play a crucial role in distinguishing the difference between odd and even groups as we will see.

- **Reflection**: It is clear that all the reflection operators are their own inverse, so it is of order 2, viz

$$m_i^2 = e, \tag{3.46}$$

so it is clear that to generate all the reflection operators, we need help from the rotation generator R to "move the n-gon into the correct position" so to speak. It is clear that, in general, the rotation and reflection operators do not commute

$$\boxed{m_i R^j \neq R^j m_i \ , \text{ except for } j = -j.} \tag{3.47}$$

Let's prove this, which we will also demonstrate the power of speaking in terms of generators. We *choose* $\{R, m_1\}$ to be the set of generators, then it is clear that we can generate rotations by $R^j = \Pi_j R$, and reflections by

$$\boxed{m_i = R^{i-1} m_1}, \tag{3.48}$$

with $R^0 = e$. Now, recalling that $m_i = m_i^{-1}$, we have from Eq. (3.48) $m_1^{-1} m_i = m_1^{-1} R^{i-1} m_1$. Choosing $i - 1 = j$, we get

$$m_1^{-1} m_{j+1} = m_1^{-1} R^j m_1 \ , \tag{3.49}$$

Meanwhile, taking the inverse of both sides of Eq. (3.48), we have $m_i^{-1} = m_i = m_1 R^{1-i}$, or $m_{j+1} = m_1 R^{-j}$. Plugging this back into Eq. (3.51), we get the relation

$$\boxed{m_1 R^j m_1^{-1} = R^{-j} \text{ or } m_1 R^j = R^{-j} m_1}. \tag{3.50}$$

Geometrically, Eq. (3.50) can be read as follow : to rotate counterclockwise R^{-j}, we first reflect with m_1, then *rotate clockwise*, and this should be the same as rotating clockwise and then reflecting since $m_1 = m_1^{-1}$.

Since it doesn't matter which m we use in Eq. (3.50) (can you see why?), the following fundamental formula in computing operations in D_n is always true

$$\boxed{m_i R^j m_i^{-1} = R^{-j}} \tag{3.51}$$

and you should memorize it. This relationship proves our assertion Eq. (3.47).

Recall now that two elements $g_1 \sim g_2$ (conjugate) if there exist *any* element g such that $gg_1g^{-1} = g_2$, Eq. (3.51) means that $R^j \sim R^{-j}$, i.e. R^j is conjugate to its inverse and both belong to the same conjugacy class. What other elements are in this conjugacy class?

The way to check this is to basically via bruteforce calculation: start with R^j, and act on it with the conjugation operator over all the elements of $g \in D_n$. We have already saw the result if g is the reflection operator – this is Eq. (3.51) and says that $R^j \sim R^{-j}$. On other hand, if g are the rotation operators:

$$L_{R^i}(R^j) = R^i R^j R^{-i} = R^j \text{ since all rotation operators commute} \tag{3.52}$$

and hence we get back the same element R^j, so the pair $\{R^j, R^{-j}\}$ forms a conjugacy class, i.e. the rotation operators form a class with its own inverse, and there is no other elements in this class.

What about the reflection operators? Let's conjugate m_1 with the rotation operators R^j and see what we get:

$$\begin{aligned} L_{R^j}(m_1) &= R^j m_1 R^{-j} \\ &= R^j(R^j m_1) \\ &= R^{2j} m_1 \end{aligned} \tag{3.53}$$

where we have used the fundamental relationship $m_1 R^j = R^{-j} m_1$ in the 2nd line. Conjugating it with the reflection operators we get, using $m_{j+1} = R^j m_1$, (using m_{j+1} instead of m_j simplifies our task)

$$\begin{aligned} L_{m_{j+1}}(m_1) &= m_{j+1} m_1 m_{j+1}^{-1} \\ &= (R^j m_1) m_1 (R^j m_1)^{-1} \\ &= (R^j m_1) m_1 m_1^{-1} R^{-j} \\ &= R^{2j} m_1 \end{aligned} \tag{3.54}$$

where in the 3rd line we have used the identity Eq. (2.38) $(ab)^{-1} = b^{-1}a^{-1}$. So conjugating m_1 with both rotation and reflection operators gets us the same relation $R^{2j}m_1$. This defines a set of reflection operators, since it is of the form $R^k m_i$.

Given the set of possible values for $j = 0, 1, 2, 3 \ldots, n-1$, we can calculate the reflection operators that are conjugate to m_1. However since $R^{2j} = R^{2j \bmod n}$, this means that what the set is depends on whether n is odd or even:

- If n is odd then every even integer $2j$ mod n is a multiple of 1, and hence *all* the reflection operators are in the same conjugacy class

  $$\text{Reflection conjugacy class for } n \text{ odd} = \{m_1, m_2, \ldots, m_n\}. \tag{3.55}$$

  For example: if $n = 5$, then $j = 0, 1, 2, 3, 4 \Rightarrow 2j \bmod 5 = 0, 2, 4, 1, 3$.

- If n is even, we only get half the relection operators,

  $$\text{Odd Reflection conjugacy class for } n \text{ even} = \{m_1, m_3, \ldots, m_{n-1}\}. \tag{3.56}$$

  For example: if $n = 6$, then $j = 0, 1, 2, 3, 4, 5 \Rightarrow 2j \bmod 6 = 0, 2, 4$. What about the $m_2, m_4, \ldots$ in this case? You can (and should) show that by computing the general conjugation of m_i with

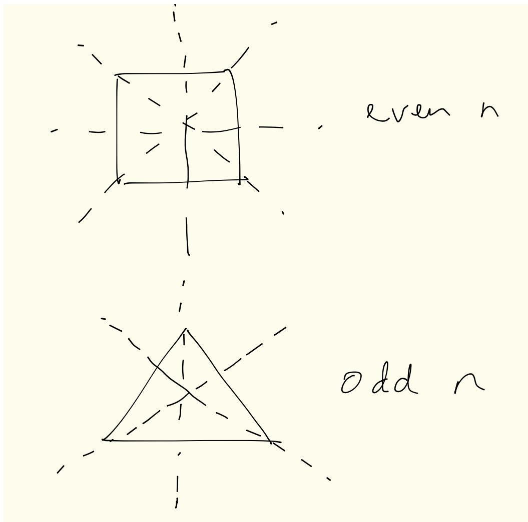


Figure 3.2: Reflection symmetries of the odd and even dihedral-n groups, illustrated by a square (D_4) and an equilateral triangle (D_3). For even n, there are two distinct reflections, while for odd n there is only one kind of reflection. This "distinct"-ness is captured by the conjugacy classes of the group.

the rotation and reflection operators that they obey $L_g(m_i) = R^{2j}m_2$ for $j = 0, 1, 2, 3 \ldots n-1$. Thus choosing $i = 2$ and using the same argument as the odd case, they form a conjugacy class by themselves

$$\text{Even Reflection conjugacy class for } n \text{ even} = \{m_2, m_4, \ldots, m_n\}. \tag{3.57}$$

In other words, for n even, odd and even reflection operators form separate conjugacy classes.

Finally the identity element always is its own (and lonely) conjugacy class, and hence we have shown that the conjugacy partitions D_n into disjoint sets.

Now after all these abstract math, let's pause see how the conjugacy classes have picked up on a property of the reflection operators that you might have noticed. Geometrically we can see that there is a qualitative difference between the reflections of odd and even Dihedral-n groups. For n odd, one edge of the reflection axis goes through a vertex and the midpoint of an edge, while for n even there are two distinct reflections – reflection axes which go through two vertexes, and reflection axes which go through two midpoints. This is why for n even, there exist two distinct conjugacy classes for the reflection operators! Notice further that, the reflection conjugacy class(es) holds all the elements of reflections which are "equivalent". When we said the words "it doesn't matter which of the m's we choose to be the one where the reflection is across the axis that runs from top to bottom", we are secretly using the fact that all these m's are in the same conjugacy class. For the rotation operators, the fact that R^j and R^{-j} are in the same class of course reflects the fact that it doesn't matter whether we call the rotation "clockwise" or "counterclockwise" which is a matter of convention. We will see other examples of "sameness" or "conjugacy" in other groups we will study – indeed one of the first thing you should do when confronted with a new group is to calculate its conjugacy classes since it immediately gives you an insight into some notion of "sameness".

3.3.1 The Permutation Group and the Symmetric Group

Suppose you are given 3 balls, and you are asked to put the 3 balls in three boxes. What is the number of possible ways you can do that? Suppose now you label the balls with numbers 1,2,3, then you can arrange the balls in the following ways

$$(1,2,3)\ ,\ (1,3,2)\ ,\ (3,1,2)\ ,\ (3,2,1)\ ,\ (2,3,1)\ ,\ (2,1,3)$$

for a total of 3! ("three factorial") ways.

Suppose that you have chosen to arrange it in the order $(1, 2, 3)$, but your BFF comes along and asked you rearrange it to become $(3, 1, 2)$. No problem you say, you just switch $1 \to 3$, $2 \to 1$, $3 \to 2$ and you are done. Saying the previous sentence is a mouthful, so let's invent some mathematical notation for "switching" in the following way

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} \tag{3.58}$$

where the top row indicites the original ordering, and the bottom row indicates the rearrangement. It is clear that, starting from $(1, 2, 3)$ you have 6 different ways of rearranging the balls, counting the "do nothing rearrangment" where you just keep the balls where they are originally. They are

$$P_1 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix} , \; P_2 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} , \; P_3 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} \tag{3.59}$$

$$P_4 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} , \; P_5 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} , \; P_6 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} \tag{3.60}$$

where we have given them names P_i and used the "=" sign indiscriminately. Mathematically, such a "rearrangement" when you swap the objects around is called a **permutation**, and we say that P_i's are **permutation operators**. Let's now imagine that you start with the balls in $(1, 2, 3)$, do a P_3 permutation, and then immediately do a P_4 permutation. In box 1, the ball goes from $1 \to 3 \to 1$, in box 2, it goes from $2 \to 1 \to 3$, and finally in box 4 we have $3 \to 2 \to 2$. Hence the final configuration is $(1, 3, 2)$.

We *define* the rule of successive permutation, or composition of operations, mathematically by writing $P_4 \star P_3$ where the **binary operator** $\star$ (remember this?) is defined to act from the right to the left. So $P_i \star P_j$ means that we permute with P_j and then with P_i. Let's drop the $\star$ from now to simplify our notation, i.e. $P_i \star P_j = P_i P_j$. Note that $P_4 P_3$ is

$$P_4 P_3 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} = P_2 \tag{3.61}$$

i.e. it is simply the same operation as P_2. Indeed, as you can check for yourself, any composition of two permutations will result in another permutation operator i.e. $P_i P_j = P_k$, where $i, j, k = 1, 2, \ldots, 6$. We say that permutations are *closed* under composition. You can also easily check that they are also *associative*

$$(P_i P_j) P_k = P_i (P_j P_k). \tag{3.62}$$

You can probably guess where we are headed – we are in the process of showing that the set of permutations $\{P_1, P_2, P_3, P_4, P_5, P_6\}$ forms a group. We have shown closure and associativity. Which leaves us to show the existence of the identity element and the inverses. The identity element is simply the "do nothing" operator P_1. For inverses, it is not hard to see that $P_2 P_2 = P_1$, $P_3 P_5 = P_1$, $P_4 P_4 = P_1$ and $P_6 P_6 = P_1$, so each element possess an inverse. Hence we have proved that the set of permutations of 3 balls forms a group, called the **Permutation Group** of 3 balls. Note that however $P_2 P_3 \neq P_3 P_2$ – in general $P_i P_j \neq P_j P_i$ so it is not Abelian.

We don't have to stop with 3 balls of course, we can have as many balls as we want. Note that you can think of a permutation as a *map* from the set back to itself (with some swapping among its elements). Notice that if we have just P_1 and P_2, then since P_2 is its own inverse, the subset $\{P_1, P_2\}$ actually forms a subgroup – in this case Ball 1 does not move at all, but that's OK, P_1 an P_2 are still permutations. This is of course Z_2. With that in mind, let's now properly define **Permutation Group**.

(Definition) Permutation Group Perm(S): Suppose S is a set, and $P : S \to S$ is a bijective map from S back to itself, then the set of bijective maps $\text{Perm}(S) = \{P_1, P_2, \ldots\}$ from S to S which are closed, with a composition operator defined as two successive permutations, forms a group in the following way:

- Closure: This is by construction.

- Identity: the trivial map $P(s) = s \forall s \in S$.

- Inverse: since P is bijective, there exist an inverse for all $P_i \in \text{Perm}(()S)$.

- Associative: Let $F(s) = P(P(s))$, but then $P(F(s)) = F(P(s))$ since $P(s)$ is a injective. This follows from the fact that all map composition are associative.

In the case where we consider *all possible permutations of the set* S, this become the **Symmetric Group**.

(Definition) Symmetric Group $\mathbb{S}_n$: Suppose that S is a finite order set, $|S| = n < \infty$, then the set of all possible permutations of S is $\mathbb{S}(S)$, and has order $|S|!$.

So in our case of the 3 balls above, $S = \{1, 2, 3\}$ is the set of 3 balls labeled by 1,2,3, and hence is known as either the permutation group of 3 balls or the symmetric group $\mathbb{S}_3$. The symmetric group is of crucial importance in group theory due to the theorem we will discuss next. In a way, the "permutation group" is a misnomer – as we will see, every finite group can be cast as a permutation – we already saw the Z_2 subgroup from $\mathbb{S}_3$ above. We'll prove a theorem about this at the end, but let's plow on.

Let's introduce an even slicker notation, called the **cyclic notation**[8]. Suppose we have the following permutation for an element of an $\mathbb{S}_5$ symmetric group,

$$P = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 3 & 1 & 2 & 4 \end{pmatrix}. \tag{3.63}$$

The rules for writing it in the cyclic notation is as follows. Start with any number, conventionally we choose 1. Now we follow the trail of permutations from $1 \to 5 \to 4 \to 2 \to 3$, and we write

$$P = (15423). \tag{3.64}$$

If we "return" to the starting point without going through all the numbers, we start a new series. So

$$P = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 1 & 2 & 5 & 4 \end{pmatrix} \tag{3.65}$$

becomes

$$P = (132)(45), \tag{3.66}$$

i.e. we have $1 \to 3 \to 2 \to 1$, so we choose 4 to start a new series and it goes $4 \to 5$. And if there are elements not permuted, by convention we leave them out. I.e.

$$P = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 1 & 2 & 4 & 5 \end{pmatrix} \tag{3.67}$$

becomes

$$P = (132). \tag{3.68}$$

Note that any *cyclic* rearrangement of P does not change the elements, so $(15423) = (42315) = (31542)$ etc. We can cycle the terms inside the brackets individually, so for $(132)(45) = (132)(54) = (321)(54)$ etc. We call the brackets an n**-cycle** where n denotes the number of elements. So (12) is a 2-cycle, (1342) is a 4-cycle, and (132)(45) is a product of a 3-cycle and a 2-cycle etc.

[8]This has nothing to do with the Cyclic group, cyclic is yet another overused word.

You can, for fun, calculate the conjugacy classes for the Symmetric Group. We won't do it here, as it will take us too far afield. So we will simply assert that elements with the same cycle structure $(\ldots)(\ldots)(\ldots)$ belong to the same conjugacy class. So for example, for $\mathbb{S}_4$ the conjugacy classes are

$$\begin{aligned} &\{e\} \\ \{(ab)\} &= \{(12),(13),(14),(23),(24),(34)\} \\ \{(ab)(cd)\} &= \{(12)(34),(13)(24),(14)(23)\} \\ \{(abc)\} &= \{(123),(124),(132),(134),(142),(143),(234),(243)\} \\ \{(abcd)\} &= \{(1234),(1243),(1324),(1342),(1423),(1432)\}. \end{aligned}$$

We won't spend too much time with the Symmetric group, although it has very interesting applications, for example in the studies of the identical particle problem in Quantum Mechanics (it is a crucial ingredient in showing that half integer spin particles obey the Pauli Exclusion Principle).

However, it is even more crucial in the standpoint of the classification of finite groups (which, as physicists, we don't care that much but it is one of the great achievements of modern mathematics). Its power comes from a famous theorem we will study next. But first, a bit more definition.

3.3.2 Homomorphism and its Kernel

When we discussed isomorphism in sect 3.1.3, we have constructed a bijective map j between two groups which preserves the group law. If we relax the condition of bijectiveness, then we obtain a more general map that simply preserves the group laws. Such a map is called a **homomorphism**.

(Definition) Homomorphism: Suppose G_1 and G_2 are groups with binary operators $\cdot$ and $\star$, then a map $f : G_1 \to G_2$ is a *homomorphism* if

$$f(g \cdot g') = f(g) \star f(g') \ , \ \forall \ g, g' \in G_1. \tag{3.69}$$

In order to preserve the group law in G_2, each element in G_1 is mapped to a *unique* element in G_2 *but the reverse is not true*. In other words, for every element $g_1 \in G_1$, $f(g_1)$ is an element in G_2, but there may exist more than one element in G_1 which is mapped to the same element in G_2, see Fig. 3.3. You will be asked to prove this in the homework.

Furthermore, by using the rule Eq. (3.69) and the group axioms, the identity in G_1 has to be mapped to the identity of G_2. *Proof*: Suppose, let $g_1, e_1 \in G_1$, so $f(g_1 e_1) = f(g_1) \star f(e_1)$, but $f(g_1 e_1) = f(g_1)$ hence $f(g_1) = f(g_1) \star f(e_1)$ and $f(e_1) = e_2$ $\square$.

Also, similarly, the inverse of g_1, g_1^{-1} is mapped to the inverse in G_2. *Proof*: Let g_1^{-1} be the inverse of g_1, so $f(g_1 g_1^{-1}) = f(g_1) \star f(g_1^{-1})$ but $f(g_1 g_1^{-1}) = f(e_1)$ so $f(g_1) \star f(g_1^{-1}) = e_2$ hence $f(g_1^{-1})$ must be the inverse of $f(g_1)$ $\square$.

Now we have several cases depending on the domain and image of the map in the codomain (see Figure 3.4.

- The map is bijective: i.e. $\text{im}(f) = \text{cod}(f)$ and one-to-one, then the homomorphism is an *isomorphism* which we have studied previously. In other words, isomorphisms are special cases of homomorphisms.

- The map is one-to-one, but not onto: the homomorphism is called an **injective/one-to-one homomorphism**. We encountered this case in the proof of Cayley's Theorem – the map j is an injective homomorphism. This means that the map f defines a so-called **isomorphic embedding** of the group G_1 into a subset of G_2 such that G_1 is subgroup of G_2.

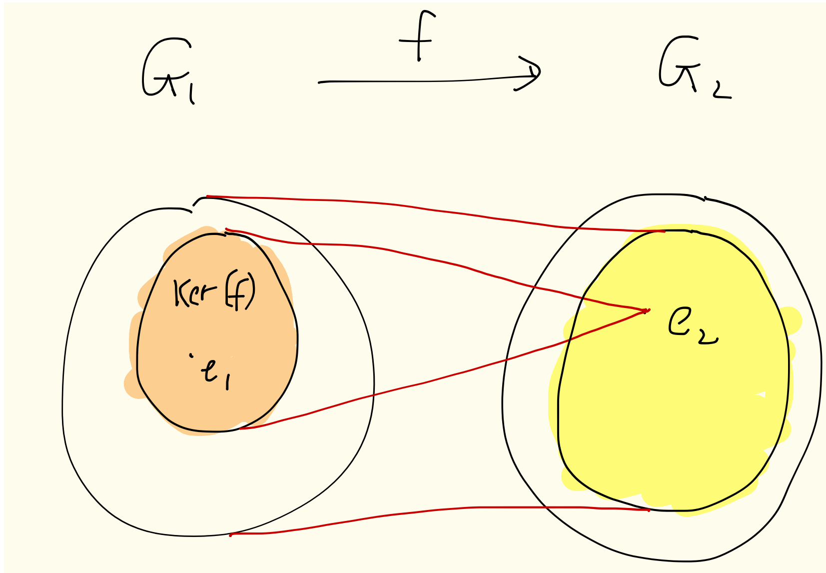


Figure 3.3: A homomorphism $f : G_1 \to G_2$. The Kernel of f is shaded red, and is mapped to e_2 the identity of G_2. The entire group G_1 is mapped to a subset of G_2 (yellow). If f maps to the entirety of G_2, and the kernel of f is trivial, then f is an *isomorphism*.

- The map is neither onto nor one-to-one: each element in G_2 is mapped to more than one element in G_1. This is the case which will be the most important in our study of groups, so we will expand the discussion below.

In the discussion below, we will drop the explicit writing of the binary operators $\cdot$ and $\star$ in from now on (restoring them only to clarify things) – it should be clear "in context".

(Definition) Kernel: Suppose G_1 and G_2 are groups, and the homomorphism $f : G_1 \to G_2$ maps more than one element of G_1 into G_2. Since the map is not necessarily onto, the **image** of f is a subset of G_2, i.e.

$$\text{im}(f) \subseteq G_2. \tag{3.70}$$

In addition, the subset of G_1 which is mapped to the identity element $e_2 \in G_2$ in G_2 is called the *kernel*

$$\text{Ker}(f) = \{g \in G_1 \text{ such that } f(g) = e_2\}. \tag{3.71}$$

This definition implies that *the kernel* Ker(f) *forms a* normal subgroup *of* G_1. *Proof*: Let's prove that it is a subgroup first, then we prove that it is also normal. Since the kernel Ker$(f) \subseteq G_1$, all we need to show is that it obeys the group laws. Let $K = \text{Ker}(f)$, then

- Identity: It must possess the identity e_1 of G_1, since f maps to the identity of G_2.
- Closure: If $k_1, k_2 \in K$, then by the defining property of the homomorphism $f(k_1)f(k_2) = f(k_1k_2) = e_2e_2 = e_2$, hence $k_1k_2 \in K$.
- Inverse: If $k \in K$, then again by defining property of f, $k^{-1} \in K$.
- Associativity: Since G_1 is a group, K inherits the associativity property from G_1.

Now to prove that it is also normal, we need to show that for every $k \in K$, $gkg^{-1} \in K\ \forall g \in G_1$. Using the defining property for the homomorphism $f(gkg^{-1}) = f(gk)f(g^{-1}) = f(g)f(k)f(g^{-1}) = f(g)e_2[f(g)]^{-1} = e_2$, and hence $gkg^{-1} \in K$. $\square$

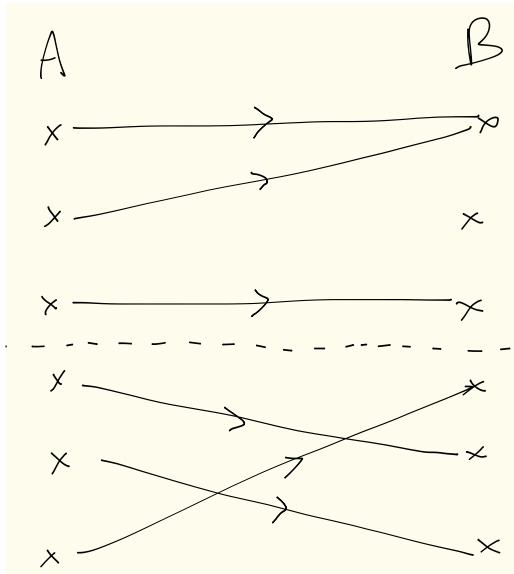


Figure 3.4: Possible homomorphisms from A to B. The top map is *one-to-one* but not *onto*, while the bottom map is bijective. Note that there is there is no map from two elements of B to an element of A.

3.3.3 Cayley's Theorem

(**You should know the theorem, but not the proof.**)

We have discussed subgroups previously. It is an interesting question to now ask "is there a group in mathematics where all other groups are subgroups of?" The answer is yes, and it was proven by Cayley in 1854, in which will now discuss.

(Theorem) Cayley: *Any group G is isomorphic to a subgroup of Symmetric Group $\mathbb{S}$ for some choice of S.*

In other words, every single group you can think of is embedded (i.e. an isomorphic embedding) in the Symmetric Group, so in vulgar language it is the mother of all groups known to humans.

The proof of this theorem is long compared to the others in these lectures, so buckle you seat-belts and read slowly. I encourage you to fully understand the proof, as you will learn some of the abstract thinking and tricks that will serve you well in your future career as a physicist.

Proof: Consider a group G with elements $g_i \in G$ and group composition law $g_i g_j \in G$. To prove this theorem, we want to show we can always construct an injective homomorphism j which maps G to $\mathbb{S}_n$. In equations, we want to construct

$$j : G \to \mathbb{S} \text{ such that } j(g_1 g_2) = j(g_1) \star j(g_2) \tag{3.72}$$

where we have used $\star$ to denote the group composition law of $\mathbb{S}$ group (which at this stage is different). There are two things we need to show:

- We want to show that $j(g_1) \star j(g_2)$ does result in an element of $\mathbb{S}$ for some S.
- For a given set $G = \{g_1, g_2, \ldots, g_n\}$, its image set $\{j(g_1), j(g_2), \ldots, j(g_n)\}$ forms a group under the group law of $\mathbb{S}$.

If these two requirements are fulfilled, then j is an isomorphism into a subset of $\mathbb{S}$ which is also a group (i.e. a subgroup of $\mathbb{S}$) then we are done. Let's begin with the first item.

Since we are free to choose S, let's choose $S = G$ – this is the sneaky but crucial trick! The map j then becomes

$$j : G \to \mathbb{S}(G). \tag{3.73}$$

$\mathbb{S}(G)$ is the set of all permutations on the elements of G. There are a total of $|G|! = n!$, this is (for $n > 2$) more than $|G|$, and since j is one-to-one, we get to choose which of $\mathbb{S}$ gets mapped to. Let's choose the following, for each $g \in G$,

$$j(g) = P_g \ , \text{ where } P_g(g') = gg' \ \forall g' \in G. \tag{3.74}$$

The above Eq. (3.74) is quite dense. Let's parse it in human language. The first part $j(g) = P_g$ tells us that j maps each element of g to $P_g \in \mathbb{S}(G)$, which at the moment we are free to choose. Each element of the $\mathbb{S}(G)$ permutes the entire set of G. We choose P_g to be the element that permutes G by **left action** (or left multiplication) with g. To see this, remember that acting on every element $g' \in G$ by an element of g results in another element in G so it essentially permutes it, viz

$$P_g(g') \equiv \begin{pmatrix} g_1 & g_2 & g_3 & g_4 & \cdots \\ gg_1 & gg_2 & gg_3 & gg_4 & \cdots \end{pmatrix} , \tag{3.75}$$

recalling that by closure $gg_1, gg_2, \ldots$ are also elements of G. So we have chosen j to map g into an element of $\mathbb{S}(G)$ whose action is to permute the entire set of G by left multiplication with g. Note that this map automatically maps the identity e of G into the trivial permutation P_e of $\mathbb{S}(G)$. (This is a lot of words to describe j, but it is crucial that you understand it as this is the key step in the proof.)

Now, we want to show that j preserves the group law, in the following way. There is some trickery with notation so read carefully. For any two elements $g_1, g_2 \in G$, and for all g' of G, we find

$$\begin{aligned} j(g_1g_2)(g') &= P_{g_1g_2}(g') \\ &= g_1g_2g' \\ &= g_1P_{g_2}(g') \\ &= (P_{g_1} \star P_{g_2})(g') \\ &= (j(g_1) \star j(g_2))(g') \end{aligned} \tag{3.76}$$

where g_1g' is to be read as Eq. (3.75) above. The final step is to show that the subset $\{P_{g_1}, P_{g_2}, \ldots, P_{g_n}\}$ forms a group.

- Identity: $j(e) = P_e$ is clearly the identity.
- Inverse: Since for any $g \in G$, $j(gg^{-1}) = j(e) = P_e = P_g \star P_{g^{-1}}$, the inverse exists for all elements P_g.
- Associativity follows from the associativity of the group laws of G and $\mathbb{S}(G)$.
- Closure of follows from the closure of group law of G and the preservation of the group law of $\mathbb{S}(G)$ under j.

Hence we have found an isomorphism j that maps G to a subgroup of $\mathbb{G}$ as required $\square$. Note that in some books, Cayley's Theorem is stated to be valid for the symmetric group. But as you just saw, the proof does not rely on the finiteness of G and hence Cayley's Theorem is also valid for infinite ordered groups. I don't quite know why some books restrict the theorem to finite ordered groups. The only reason I can think of is history as Cayley first proved it for finite ordered groups but it was later extended by other people for infinite ordered groups.

Chapter 4

Representation Theory for Finite Groups

I've got a bad feeling about this.

Various *Star Wars* characters.

In this chapter, we will discuss matrix groups and representations.

Previously, in Chapter 1, when we discussed the D_4 group, we showed that we can *represent* group operators with matrices. For each such matrix representation of the group operators, the target space (i.e. the objects that the operators act on, like the square) can then be similarly represented by matrices. Each such pair of sets of matrices is called a **Linear Representation**, or just "representations" for short.

Why would we want to do that? Leaving aside the purely mathematical reasons (mathematicians are like wizards – they just do whatever they want to do), there are several reasons. Operationally, we know a lot about matrix algebra or **linear algebra**, so that makes life easy when we want to calculate stuff. Also as we have mentioned, group operators as **abstractions** – indeed the group algebra we have learned in the previous Chapter 3 is called **abstract algebra**, but obviously each group can describe the symmetries of many different things. For example, the group of permutation of 4 objects S_4 can equally describe the symmetry of the methane molecule or the possible ways of putting 4 balls into four cups. So different representations can "represent" (see why they call it representations?) different physical systems. Finally, most importantly, as we will soon see, *representations are not unique* – in other words, each group has many *inequivalent* ways (indeed, an infinite number of ways) of being represented. This added structure adds a lot of richness mathematically speaking, but also vastly increase its usefulness to physics.

4.1 Matrix Groups

Since we can represent group operators as matrices, it's not surprising that the matrices themselves must also form a group under their own composition law, which is simply matrix multiplication. For example, as we have seen in Chapter 1, the set of 8 2×2 square matrices that represent that D_4 group Eq. (1.22), Eq. (1.23), Eq. (1.26) and Eq. (1.27), form a group under matrix multiplication. As it turns out, this set of matrices is actually a subgroup of a much, much larger group of 2×2 matrices, $GL(2, \mathbb{R})$. Unlike the discrete group we have studied in Chapter 3, this group is a **continuous group**. Let's now define what we mean by a continuous group.

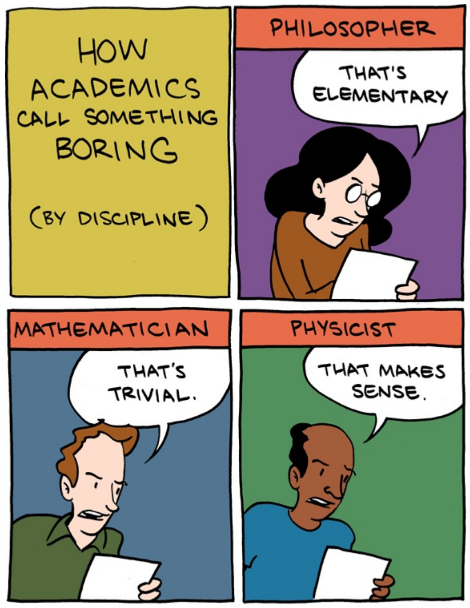


Figure 4.1: Stolen from `http://www.smbc-comics.com/`.

4.1.1 Continuous Groups

Since they are groups, continuous groups also obey the group axioms we laid out in the beginning of Chapter 3. Let's consider a continuous group that you know and love.

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.

If you have been paying attention, you will notice that way back in Chapter 2 we have defined a Field of real numbers with addition and multiplication. As it turns out, if we take away multiplication, then the algebra of addition on the set of real numbers form a group. 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**. We discuss a little more about continuity later in section 5.1, but for now let's make a hand-wavy distinction. We can intuitively think that if all elements in a set are said to be "infinitisimally" close to at least another element in the set, then we can "continuously" move from one element to some other elements in the set[1], then the set is continuous. But this would require us to define a notion of "closeness". A set equipped with this definition is known as a **topological space**. The set of real numbers $\mathbb{R}$ is such as space, where the notion of "closeness" is given by its distance L defined by[2]

$$L \equiv |x - y| \ \forall x, y \in \mathbb{R} \ . \tag{4.1}$$

[1]We have been careful not to say *all* other elements in the set since continuous spaces can also be not simply connected.
[2]It is also a metric space, but this is already too much math.

So two elements x and $x + \Delta x$ are "infinitisimally close" if $\Delta x \to 0$. It's clear that $x \in \mathbb{R}$, this exists. Anyhow, in lieu of a long digression, in these lectures, we will cheat and *assert* the following:

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

4.1.2 Linear/Matrix Groups

Let's now talk about more interesting continuous groups. We begin with the group that string theorists love[3].

(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

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \ , \ a, b, c, d \in \mathbb{C} \tag{4.2}$$

Since all the matrices must have $\det(A) = 1$ (the S in $\mathbb{C}$), this means that

$$ad - bc = 1. \tag{4.3}$$

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
$$e = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{4.4}$$

- 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 numbers[4], 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 (i.e. complex numbers equipped with the Field axioms we studied in Chapter 2). The word *Special* means that the determinant of the matrices is 1, while *Linear* simply means that the elements are matrices, and obviously 2 means the elements are 2×2 matrices. So you can also have $SL(4,\mathbb{R})$ the group of 4×4 matrices with real entries and unit determinant etc.

$SL(2,\mathbb{C})$ 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.

[3] In string theory, this group is also known as the **Global Conformal Group** in 2D, and it describes the symmetries of a world sheet, and as you will show in an homework problem, it preserves the angle of any two lines on the complex plane.

[4] Complex numbers are really an ordered pair of real numbers $\mathbb{C} = \mathbb{R} \times \mathbb{R}$ equipped with the usual additional and multiplication operators, and additionally possess a *conjugation* action in the following way if $x, y \in \mathbb{R}$ then the conjugation relates $(x + iy) = (x - iy)^*$. In case you are wondering – this conjugation action is an equivalent relationship.

(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.)

As you might have guessed, the word *General* means that the determinants of the matrices can be any value except zero (which is required for invertibility of the matrices). Obviously, $GL(N,\mathbb{R})$ would be a similar group, but with real entries.

4.1.3 Special Orthogonal Group $SO(N)$ and Rotations

Another subgroup of $GL(N,\mathbb{R})$ is the special orthogonal group $SO(N)$, which as we will see, describes the symmetry of rotations in N dimensions.

(Definition) Special Orthogonal Group $SO(N)$: Let $SO(N)$ be a set of $N \times N$ matrices R such that $\det R = 1$, and the additional condition $R^T R = e$, where R^T is the **Transpose**[5] of the matrix R. Then $SO(N)$ 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 $R_1, R_2 \in SO(N)$, then

$$\begin{aligned}(R_1R_2)^T(R_1R_2) &= R_2^T R_1^T R_1 R_2 \\ &= R_2^T(R_1^T R_1)R_2 \\ &= R_2^T R_2 \\ &= e \end{aligned} \tag{4.6}$$

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). You might remember from your first year linear algebra class that an orthogonal matrix is a matrix whose inverse is its transpose This group is a subgroup of the **Orthogonal Group** $O(N)$, which relaxes the condition on the determinant to be ± 1 (instead of just $+1$).

Why is this the group of rotations in N dimensions? Let's look at the simplest $N = 2$ case, $SO(2)$, whose elemenst have the form

$$R = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \ , \ a, b, c, d \in \mathbb{R} \tag{4.7}$$

with unit determinant

$$ad - bc = 1. \tag{4.8}$$

Meanwhile the orthogonality condition

$$R^T R = e \tag{4.9}$$

gives us three additional equations

$$a^2 + b^2 = 1 \ , \ ac + bd = 0 \ , c^2 + d^2 = 1 \tag{4.10}$$

A solution to these set of equations Eq. (4.8) and Eq. (4.10) is

$$a = \cos\theta \ , \ b = -\sin\theta \ , \ c = \sin\theta \ , \ d = \cos\theta \tag{4.11}$$

[5] I.e. For a 2×2 matrix A,

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \ , \ A^T = \begin{pmatrix} a & c \\ b & d \end{pmatrix}. \tag{4.5}$$

i.e., the matrix element is *parameterized by a single real parameter*[6] θ.

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}. \tag{4.12}$$

Note that since the trigonometric functions are periodic the domain of $0 \leq \theta < 2\pi$. We say that the domain is **compact** and $SO(2)$ is a **compact group**[7].

For those who have seen it before, it's clear that Eq. (4.12) is a rotation matrix in 2 dimesions, in the following way. Recall that group elements can *act* on sets, as we have studied in Chapter 3 in the following way.

Let $SO(N)$ act on the points belonging to the 2 dimensional Cartesian coordinates space[8] $\mathbb{R}\times\mathbb{R} = \mathbb{R}^2$, i.e. whose elements are the ordered pairs (x, y) with $x, y \in \mathbb{R}$. How do we want R_θ to act $\mathbb{R}^2$? Since the operators R_θ are 2×2 matrices, we can arrange (x, y) as a 2×1 column matrix V, and *define* the action as matrix multiplication from the left, which takes an element of $\mathbb{R}^2$ into another element of $\mathbb{R}^2$, i.e.

$$\text{Action of } SO(N) \text{ on } \mathbb{R}^2 \quad R_\theta : \mathbb{R}^2 \to \mathbb{R}^2;\ R_\theta V = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x' \\ y' \end{pmatrix} \tag{4.13}$$

where

$$x' = x\cos\theta - y\sin\theta\ ,\ y' = x\sin\theta + y\cos\theta. \tag{4.14}$$

As Figure 4.2 shows, this is a rotation of the vector (x, y) into a (x', y') counter-clockwise by an angle θ.

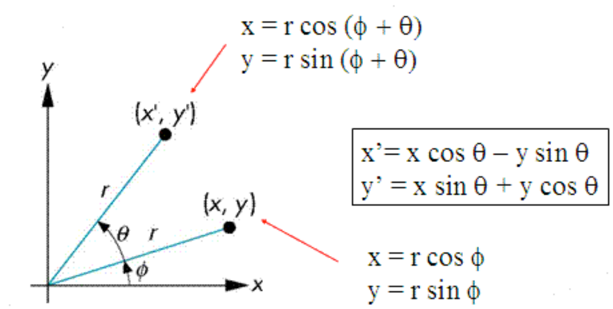


Figure 4.2: Action of an operator of $R_\theta \in SO(N)$ on an element of $(x, y) \in \mathbb{R}^2$ is a counterclockwise rotation by θ.

[6]You might be puzzled why 4 equations do not yield the solutions to 4 unknowns a, b, c, d – it turns out that one of the equations is redundant.

[7]$\mathbb{R}$ under addition is an example of a **non-compact** group. Compactness is a topological property that is a bit technical to define, and will not really reveal a lot of insight. Roughly speaking, a compact space is something which we can reach the "boundary" by successively taking smaller and smaller steps (called a sequence) towards it. So for example, the interval $[0, 1]$ is compact because we can reach 0 by taking the sequence of steps $1, 1/2, 1/4, 1/8, \ldots$ towards it, and reach 1 by taking the sequence of steps $0, 1/2, 3/4, 7/8 \ldots$.

[8]See Chapter 1 to refresh your memory on product sets.

We can compose two rotation operators R_{θ_1} and R_{θ_2}, by using matrix multiplication

$$\begin{aligned}
R_{\theta_1}R_{\theta_2}V &= \begin{pmatrix} \cos\theta_1 & -\sin\theta_1 \\ \sin\theta_1 & \cos\theta_1 \end{pmatrix} \begin{pmatrix} \cos\theta_2 & -\sin\theta_2 \\ \sin\theta_2 & \cos\theta_2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\
&= \begin{pmatrix} \cos(\theta_1)\cos(\theta_2) - \sin(\theta_1)\sin(\theta_2) & -\cos(\theta_2)\sin(\theta_1) - \cos(\theta_1)\sin(\theta_2) \\ \cos(\theta_2)\sin(\theta_1) + \cos(\theta_1)\sin(\theta_2) & \cos(\theta_1)\cos(\theta_2) - \sin(\theta_1)\sin(\theta_2) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\
&= \begin{pmatrix} \cos(\theta_1+\theta_2) & -\sin(\theta_1+\theta_2) \\ \sin(\theta_1+\theta_2) & \cos(\theta_1+\theta_2) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \qquad (4.15) \\
&= R_{\theta_1+\theta_2} \begin{pmatrix} x \\ y \end{pmatrix} \qquad (4.16)
\end{aligned}$$

i.e. successive operations of two rotation matrices is cumulative. As practice, you can show that this group is abelian, i.e.

$$R_{\theta_1}R_{\theta_2} = R_{\theta_2}R_{\theta_1} = R_{\theta_1+\theta_2}. \qquad (4.17)$$

Note that $SO(N)$ for $N > 2$ is generally *non-abelian* – can you see why?

In the introduction Chapter 1, we said that if we represented the group elements with square matrices, and objects which they act on as column matrices, then together they form a **Group Representation**. Here, $SO(2)$ *is already a matrix*, and $\mathbb{R}$ have been represented as column matrices V, so together they form a Group Representation of $SO(2)$. We say "*a* representation" and not "*the* representation" because, in general, there are many representations for the same group. We will formally discuss Group Representations in the next section.

4.1.4 Unitary Groups $U(N)$ and Special Unitary Groups $SU(N)$

Another matrix group that is very important in physics is the Special Unitary Group $SU(N)$.

(Definition) Unitary Group $U(N)$: Let $U(N)$ be a set of $N \times N$ square matrices with non-zero determinant, complex entries and the **Unitary condition** $U^\dagger U = UU^\dagger = e$, where $U^\dagger = (U^*)^T$ is the **Hermitian Conjugate** of U. Then $U(N)$ forms a group under usual matrix multiplication.

Like its close brethren, the $SO(N)$, the only tricky part of the proof is to show closure, which follows the same prescriptoin

$$\begin{aligned}
(U_1U_2)^\dagger(U_1U_2) &= U_2^\dagger U_1^\dagger U_1 U_2 \\
&= U_2^\dagger(U_1^\dagger U_1)U_2 \\
&= U_2^\dagger U_2 \\
&= e \qquad (4.18)
\end{aligned}$$

An equivalent way of writing the Unitary condition $U^\dagger U = e$ is $U^\dagger = U^{-1}$, where in very high brow language we say "its inverse is its **adjoint**" – we'll come back to the often-seen word "adjoint" later.

(Definition) Special Unitary Group $SU(N)$: The Special Unitary Group $SU(N)$ are $U(N)$ groups with the additional condition that the determinant of the matrices is 1.

Let's take a slight detour off group theory. You might have seen Hermitian Conjugation in your Quantum Mechanics class, and have studied operators. This is exactly the same Hermitian Conjugation – in Quantum Mechanics the space of states, called a **Hilbert Space** is a Complex Vector Space. Operators in Quantum are *linear maps* that acts on a state to take it to another state in the Hilbert Space. For example, suppose $\hat{O}$ is an operator, and ψ is some normalizable state, then

$$\hat{O} : \psi \to \phi \qquad (4.19)$$

where ϕ is some other state which is normalizable. Furthermore, if f is a **Hermitian Operator**, then it obeys $f = f^\dagger$, and f is associated with an **observable**. For example, the momentum operator $\hat{\mathbf{p}}$ is a Hermitian Operator. Sometimes you might also hear people say that Hermitian Operators are **Self-adjoint**[9]. We will talk about Vector Spaces and Linear Maps in Section 4.2.2.

In group theory language, the group $SU(N)$ operators matrices and the Hilbert Space form a Group representation. Physical symmetries are then associated with states in the Hilbert space which are *equivalent* under the group action. These are dense words for the moment, but we will come back to study group theory in Quantum Mechanics in Chapter **??** when you will see that the symmetries associated with the $U(N)$ and $SU(N)$ operators are associated with conservation of charges and spins.

For now, let's consider the simplest of the unitary group, $U(1)$. We will now show that it is actually isomorphic to $SO(2)$ which we just studied, i.e. $U(1) \cong SO(2)$.

Proposition: $SO(2) \cong U(1)$

Proof: A literal reading of the definition for the Unitary Groups in Section 4.1.4 tells us that $U(1)$ is the set of 1×1 complex matrix, which of course is just a complex number. Let $u \in \mathbb{C}$ be an element of $U(1)$, then the unitary condition implies

$$u^* u = 1 \text{ or } |u|^2 = 1. \tag{4.20}$$

Parameterizing the complex number u using the de Moivre formula

$$u = r \exp(i\theta) \ , \ r, \theta \in \mathbb{R} \tag{4.21}$$

where r is the radial distance in the Argand diagram and θ is the **argument** of u. But Eq. (4.20) implies that $r = 1$ and $0 \leq \theta < 2\pi$, so we can write the set of the elements of $U(1)$ as

$$U(1) = \{\exp(i\theta) \mid 0 \leq \theta < 2\pi\}. \tag{4.22}$$

Then it is easy to see that the group law implies

$$u_{\theta_1} u_{\theta_2} = \exp(i\theta_1 + i\theta_2) = u_{\theta_1 + \theta_2} \tag{4.23}$$

is obeyed. $U(1)$ is clearly also an abelian group.

But look! Comparing Eq. (4.23) to the group law for Eq. (4.17), we can find the isomorphism

$$i : U(1) \to SO(2) \ ; \ i(\exp(i\theta)) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \tag{4.24}$$

and with both sets of elements parameterized by $0 \leq \theta < 2\pi$, this means that this map is injective. So despite their different names, *they are the same group*, i.e. $SO(2) \cong U(1)$. $\square$.

[9] What is the meaning of "adjoint"? In fact, you will hear the word "adjoint" used in apparently very different contexts in these lectures: adjoint operators, adjoint representations, self-adjoint etc. Adjoint-ness is an important abstract concept which is unfortunately not discussed more widely in physics. Suppose X is a set, and we are given an *action* $f : X \to X$, i.e. maps X into itself. Furthermore there exist a notion of "taking the inner product" (think of it as a way to define distance between objects in a set) of the elements of X defined to be (x_1, x_2) for $x_1, x_2 \in X$, then if there exist another action g such that

$$(gx_1, x_2) = (x_1, fx_2)$$

then we say that *g is the adjoint of f*. Of course, what the adjoint is depends on the set X, and the action, and the notion of "inner product". For example, in quantum mechanics, given two wavefunctions $\psi(x)$ and $\phi(x)$, then the inner product $(\phi(x), \psi(x)) \equiv \int_{-\infty}^{\infty} \phi(x)^* \psi(x)\, dx$. So an operator $\hat{O}$ is *self-adjoint* if $(\hat{O}\phi, \psi) = (\phi, \hat{O}\psi)$, which you should be able to show yourself that implies $\hat{O}^\dagger = \hat{O}$. The usual notation physicists used to indicate the adjoint an operator $\hat{O}$ is with the dagger $\dagger$ symbol – can you see why the raising and lowering operators $\hat{a}$ and $\hat{a}^\dagger$ of the Simple Harmonic Oscillator are adjoints of each other but they are not self-adjoint?

What about the representation of $U(1)$? Obviously, the representations of $SO(2)$ are 2×1 matrices so there is no natural action one can define for $u = \exp(i\theta)$ to act on them. On the other hand, one can define a 1×1 "matrix" v, i.e. another complex number $v \in \mathbb{C}$. Let $v = r \exp i\alpha$ with $0 < r < \infty$ and $0 \leq \alpha < 2\pi$, then the action of u on v is simply

$$u_\theta v = r \exp(i\alpha + i\theta) \tag{4.25}$$

which is to *rotate* v in the Argand diagram clockwise by θ. This is exactly the same action as the action of $SO(2)$, R_θ, on a vector in $\mathbb{R}^2$ space. The coincident is not surprising of course.

Now notice that while $SO(2) \cong U(1)$ are isomorphic to each other, their *Group representations* are different. Hence we learn one very crucial lesson here: *for any group, there may exist more than one group representations.*

4.2 Matrix Representions

4.2.1 Matrix (linear) Representations

We have seen before that we can represent group elements with matrices. For example in the introduction Chapter 1, we represented D_4 by a set of 2×2 matrices. Let us now formalize this. We will be discussing *linear representations.*

Suppose we have a group G. To represent a group G as a set of matrices, we identify each element of the group $g \in G$ with a matrix $U \in GL(N, \mathbb{C})$, i.e. the matrix is an element of the set of $n \times n$ matrices with non-zero determinant and complex entries. But look! We know from our study of continuous groups that $GL(N, \mathbb{C})$ *is also a group* under the usual matrix multiplication. This means that we can define a representation by *a map between the group G and $GL(N, \mathbb{C})$* in the following way.

(Definition) Linear Representations: A group G is said to be linearly represented by $U(g)$ with $g \in G$, where U is a *homomorphism* from G to $GL(N, \mathbb{C})$

$$\text{Linear Representation } U : G \to GL(N, \mathbb{C});\ U(g_1)U(g_2) = U(g_1 g_2). \tag{4.26}$$

Here think of the U's as square $n \times n$ matrices and $U(g_1)U(g_2)$ as simply matrix multiplication.

The first example you have seen is D_4 in Chapter 1. You also have seen the representation for Z_2 in Chapter 3. All the continuous groups that we discussed earlier in this chapter are already conveniently matrices, and hence are "represented by themselves". On the other hand, don't let this simple and intuitively notion lull you into complacency, particularly on the case of the continuous groups!

There are two key points here.

- The first easy key point is that we have left the dimensionality of N in the above definition as a free parameter – indeed this means in general *there exist more than one linear representation for any group G.*

  Example: For the group $Z_2 = \{e, a\}$, we have seen the 2×2 representation previously

  $$T(e) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \ ,\ T(a) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{4.27}$$

  But there is also a 1×1 representation

  $$U(e) = 1\ ,\ U(a) = -1 \tag{4.28}$$

  where you can easily show that U is a homomorphism between Z_2 and $GL(1, \mathbb{C})$.

Example: For the cyclic group $Z_4 = \{e, a, a^2, a^3\}$, an "obvious" representation by a set of 4×4 matrices U. Let's see how we can construct this representation by first constructing the multiplication table

$$\begin{array}{c|ccc} e & a & a^2 & a^3 \\ \hline a & a^2 & a^3 & e \\ a^2 & a^3 & e & a \\ a^3 & e & a & a^2 \end{array} \tag{4.29}$$

Let's start with the element e, which is obviously mapped to the identity in the matrix representation (we use spaces for zeroes for legibility)

$$U(e) = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix} \tag{4.30}$$

Now look at the 2nd row $\{a, a^2, a^3, e\}$. We see that it is a **complete rearrangement** in the ordering from the first row $e \to a$, $a \to a^2$, $a^2 \to a^3$ and $a^3 \to e$. Similarly, the 3rd row $\{a^2, a^3, e, a\}$ is also a complete rearrangement of the first row and so on. We can hence easily construct the rest of the representations

$$U(a) = \begin{pmatrix} & 1 & & \\ & & 1 & \\ & & & 1 \\ 1 & & & \end{pmatrix}, \; U(a^2) = \begin{pmatrix} & & 1 & \\ & & & 1 \\ 1 & & & \\ & 1 & & \end{pmatrix}, \; U(a^3) = \begin{pmatrix} & & & 1 \\ 1 & & & \\ & 1 & & \\ & & 1 & \end{pmatrix} \tag{4.31}$$

which you can check that obey the group laws, e.g. $U(a)U(a^2) = U(a^3)$ etc. In fact, while this is a rather trivial example, it is a theorem[10] that *given any finite order group with order $|G| = n$, one can always construct a multiplication table and follow the above prescription to construct an $N \times N$ representation.* Such a representation is called a **Regular Representation**. We will use this construction in section 4.2.8 when we discuss how to decompose a reducible representation into irreps. (You will be asked to construct the regular representation for a D_3 group in the homework.)

One can also find a 3×3 representation of Z_4. There are many ways to find such representations (and indeed even 3×3 representations are *not unique*). A way is to use the fact that a is the generator of Z_4, with the cyclic condition $a^4 = e$. For example, one can find

$$T(a) = \begin{pmatrix} 1 & & \\ & & -1 \\ & 1 & \end{pmatrix} \text{ such that } T(a^4) = (T(a))^4 = T(e) = \begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix} \tag{4.32}$$

so the rest of the elements can be generated by multiplying $T(a)$ with itself

$$T(a^2) = T(a)T(a) = \begin{pmatrix} 1 & & \\ & -1 & \\ & & -1 \end{pmatrix}, \; T(a^3) = T(a)T(a)T(a) = \begin{pmatrix} 1 & & \\ & & 1 \\ & -1 & \end{pmatrix} \tag{4.33}$$

- The second much more subtle and important key point is that we have used *homomorphisms* to construct a representation, and not isomorphisms. *This means that one can find representations where the matrices represent more than one group element.* Let's see how this works in practice by

[10] You might be able to see that this theorem is actually a specialized version of Cayley's Theorem we proved earlier. Mathematicians usually just say "by Cayley's Theorem" blah blah blah...and leave the rest of us scratching our heads.

considering a 2×2 representations of the Z_4 group we have been discussing above. Consider the generator element

$$U(a) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \text{ such that } U(a^4) = U(e) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{4.34}$$

then

$$U(a^2) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \ , \ U(a^3) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{4.35}$$

Notice that both a^2 and e are mapped to the 2×2 identity matrix $\mathbb{I}$, so the kernel $\text{Ker}(U) = \{e, a^2\}$ has two elements[11]. This means that this 2×2 representation do not "fully represent" all the intricacies of the Z_4 group. Such a representation is called an **Unfaithful Representation**. On the other hand, if the representation is also an *isomorphic embedding*, i.e. for each element $g \in G$ there is a unique matrix in $GL(N, \mathbb{C})$, then such a representation is called a **Faithful Representation**, so all the representations that we have considered up to now but not including the 2×2 representation of Z_4 are all faithful representations.

Now here is the fun bit: although continuous groups are matrices themselves, we can represent them with other matrices! In fact, you have seen a toy version of this previously in the guise of $U(1)$ and $SO(2)$ – although they are isomorphic to each other we could have easily say that we have represented the 2×2 matrices of $SO(2)$ with the 1×1 matrices of $U(1)$ and *vice versa*, since both of them are after all subgroups of $GL(N, \mathbb{C})$.

4.2.2 Vector Spaces and Linear Actions on Vector Spaces

We have been slightly careless in the past section in describing the matrix representation of groups as "the representation". But as we have emphasised way back in Chapter 1, a Group Representation for a group G consists of both the matrix representation of the group elements *and* the underlying vector space that the matrices *act on*.

Let us now formalize the notion. We begin by reminding you that an (left) action of a mathematical object on an another (usually class of) object can be written as (Eq. (1.14))

$$\text{result} = \text{operator} \times \text{object}. \tag{4.36}$$

Now if the "operator" above is a $N \times N$ square matrix, then the "object" can be any $N \times M$. We will focus on the case where $M = 1$, i.e. the object is a column matrix.

(Definition) Complex Vector Space: Consider three such objects,

$$V = \begin{pmatrix} v_1 \\ v_2 \\ . \\ v_N \end{pmatrix} \ , \ U = \begin{pmatrix} u_1 \\ u_2 \\ . \\ u_N \end{pmatrix} \ , \ W = \begin{pmatrix} w_1 \\ w_2 \\ . \\ w_N \end{pmatrix} \tag{4.37}$$

where $v_i, u_i, w_i \in \mathbb{C}$ are in general complex numbers. Analogous to the **vector** that you have learned in high school, one can think of such complex column matrices as elements of an **N-dimensional Complex Vector Space**. From your prior knowledge on matrices, the following conditions must be familiar to you. For any $a, b \in \mathbb{C}$

- Associativity: $U + (V + W) = (U + V) + W$.

[11] You should also check that the Kernel is also a normal subgroup of Z_4, and that it is isomorphic to Z_2. Also the fact that a and a^3 are mapped to the same matrix is simply a coincident.

- Closure: $U + V \in \mathbb{V}$.
- Identity: $U + 0 = U$, where 0 here denotes a vector where all the entries are zeroes.
- Inverse: For every U, there exist an inverse element $-U$ such that $U + (-U) = 0$.
- $a(V + U) = aV + aU$,
- $(a + b)V = aV + bV$,
- $a(bV) = (ab)V$.

Notice that there are two distinct operations – a "vector addition" and a "scalar multiplication". In high-brow mathematical language, we say that $\mathbb{V}$ *is a vector space over the Field of complex numbers.* Harking back to Chapter 2, "Field" means that the vector space is armed with two binary operators, the usual vector addition "+" and the *scalar multiplication* aV, i.e. it is an **Algebra**.

Now let us consider some examples of vector spaces, some which are familiar to you but perhaps you are not aware are actually secretly vector spaces.

Example: "Ordinary" 3-dimensional real vectors.

Example: The N-dimensional abelian group of complex numbers $\mathbb{C}^N$, with the additional scalar multiplication binary operation (and hence promoting the group algebra into a Field algebra). Let $(a_1, a_2, a_3, \ldots, a_N)$ be any element of $\mathbb{C}^N$ – you can hardly resist the temptation of thinking of them as vectors already, but they are not vectors *until* we define the following scalar multiplication operation

$$\mu(a_1, a_2, a_3, \ldots, a_N) = (\mu a_1, \mu a_2, \mu a_3, \ldots, \mu a_N). \tag{4.38}$$

Example: **Quantum Mechanics and Function Spaces**. Consider the set of all possible complex functions over a compact real interval $x = [0, 1]$ with the following conditions

$$F = \{f(x)|x \in [0, 1], f(0) = f(1) = 0\}\ ,\ f : [0, 1] \to \mathbb{C}. \tag{4.39}$$

i.e. the values of the function vanishes at the boundaries $x = 0, 1$. We can add two functions together in the usual way

$$f(x) + g(x) = (f + g)(x). \tag{4.40}$$

Don't let the strange notation confuse you – Eq. (4.40) simply says that, for any $x_0 \in [0, 1]$, either adding the functions up in their functional form and then evaluating the composite function at x_0 *or* evaluating the functions separately at x_0 and then adding their results together, give the same numerical result. (Read again.)

Scalar multiplication can also be defined by

$$af(x) = (af)(x). \tag{4.41}$$

Given the definitions Eq. (4.40) and Eq. (4.41), F forms a vector space. Let's be a pedant and quickly check through the axioms. The addition definition Eq. (4.40) clearly obeys Associativity. For Closure, we need to show the addition operation results in functions which obey the boundary condition, i.e. for any two functions $f_1(x), f_2(x)$, $f_1(0) + f_2(0) = f_3(0) = 0$, and the same for $x = 1$. The identity element is clearly the function which is always zero $f_0 = 0$, while the inverse of any function is simply minus of itself. The scalar multiplication axioms follow from the fact that the functions f itself *is a map from* $[0, 1]$ *to the set of complex values*, and we use the scalar multiplication of real values to prove the rest.

In your Quantum Mechanics class, you might have studied the infinite square well. There you solved the time-independent Schrodinger's equation

$$-\frac{\hbar^2}{2m}\frac{d^2\psi_n(x)}{dx^2}+U(x)\psi_n(x)=E_n\psi_n(x) \tag{4.42}$$

where $U(x)$ is the potential shown in figure 4.3, E_n are the **Energy Eigenvalues** and $\psi_n(x)$ are the **Energy Eigenfunctions** or **Stationary States**. Since the potential $U(0)=U(1)=\infty$, the eigefunctions $\psi_n(0)=\psi_n(1)=0$ vanish at the boundaries, obeying the condition we wrote down in our definition of a function space Eq. (4.39). Hence *the set of all eigenfunctions forms a vector space*, i.e.

$$\psi_n(x)=\sin(n\pi x)\ ,\ n=1,2,\ldots,N\ ,\ F=\{\psi_n(x)\}. \tag{4.43}$$

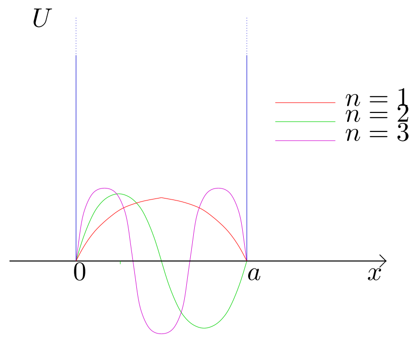


Figure 4.3: An infinite square well potential and the first few eigenfunctions of a quantum mechanical particle.

Given these examples, one can see that if we multiply an N-th dimensional vector with an $N\times N$ square matrix group matrix operator, what the operator does is to map the vector into another vector in the same vector space. For example, for a $N=3$ vector, acted upon by a 3×3 matrix representation of some group element $U(g)$ is simply

$$\begin{pmatrix} v_1' \\ v_2' \\ v_3' \end{pmatrix}=U(g)\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}. \tag{4.44}$$

We can use the more slick compact notation of indices i.e.

$$V_j'=\sum_i M_{ji}V_i. \tag{4.45}$$

We can also introduce the **Einstein Summation Convention**, whereby we automatically summed over the repeated indices. I.e. on the right, we can simply write

$$\sum_i M_{ji}V_i \to M_{ji}V_i\ , \tag{4.46}$$

dropping the $\sum_i$ altogether. In particular, we see that, for $a,b\in\mathbb{C}$

$$M_{ji}(aV_i+bW_i)=aV_j'+bW_j' \text{ where } V_j'=M_{ji}V_i\ ,\ W_j'=M_{ji}W_i. \tag{4.47}$$

Such a map that obey the Eq. (4.47) is called a **Linear Map**, and the matrix that executes this map a **Linear Operator**. If $\mathbb{V}$ is a finite dimensional vector space, i.e. $N<\infty$, it is clear that we can represent linear maps as $N \times N$ matrices. But $N \times N$ matrices are exactly the form that group elements are represented! This is what we mean by "Group Representations consists of the matrix representations of groups, and the underlying vector space which they act on".

Linear maps do not have to be matrices – although they can be represented as such (which is why the study of matrices is called **Linear Algebra**). Abstracting, a linear operator is defined to be a map that takes an element of a vector space to another in the following way

$$f(aV + bU) = af(V) + bf(U) \ \forall \ a, b \in \mathbb{C}. \tag{4.48}$$

Such a mapping is sometimes also called a **Linear Transformation**. A common example of a linear transformation or linear map is the *derivative operator* d/dx acting on the space of all possible *normalizable* functions in x – you can check that for any normalizable functions $f(x)$ and $g(x)$,

$$\frac{d}{dx}(af(x) + bg(x)) = a\frac{df}{dx} + b\frac{dg}{dx}. \tag{4.49}$$

Can you think of a *non-linear transformation?*[12]

Finally, we end this section on Vector Spaces with a few definitions and results which will be useful for the future. From your study of Quantum Mechanics, you learned that any wavefunction can be described by a linear superposition of eigenfunctions (here we ignore the fact that there maybe time evolution involved), i.e.

$$\Psi(x) = \sum_{n=1}^{N} a_n \psi_n(x) \ , \ a_n \in \mathbb{C} \tag{4.50}$$

i.e. a_n are complex coefficients. Compare this to the "ordinary three vector" $\mathbf{V} = v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}$ where $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$ are **basis vectors**, you can see that the eigenfunctions similarly *form a basis.* But since N is infinite[13], *this function space is infinitely dimensional.*

One thing which you have learned from Quantum Mechanics, but is a general property of linear vector spaces, is **Linear Independence**.

(Definition) Linear Independence. Suppose $\{v_i\}$ is a set of vectors. We say that this set is *linearly independent* if there exist no set of non-zero complex values λ_i such that

$$\sum_i \lambda_i v_i = 0. \tag{4.51}$$

This is a very precise definition for something you intuitively have grasped. For example, the basis vectors for ordinary three vectors are linearly independent of each other, while $2\mathbf{i} + 3\mathbf{j}$, $-2\mathbf{i} + 5\mathbf{j}$, $4\mathbf{j}$ are clearly not linearly indepedent to each other.

Don't confuse linear independence with the highly related but not exactly equivalent **orthogonality**. Orthogonality imply linear independence, but not the other way around. But you ask "what is orthogonality"? To define orthogonality requires a little bit of extra structure on the Vector space – the notion of an **Inner Product**. In usual ordinary 3 vectors, you remember that there exist an operation call the "dot product" of two vectors $\mathbf{u} \cdot \mathbf{v} = u_1v_1 + u_2v_2 + u_3v_3$ which is simply a (real) number. It has no direction, so it is a **scalar**. In high-brow math language, such an operation is called a **Scalar Product** or **Inner Product** or (obviously) **Dot Product** of two vectors, and the idea generalizes to the Vector Space we are studying.

(Definition) Scalar Product: Suppose $U, V \in \mathbb{V}$ are elements of a Vector Space over the complex numbers, then the *scalar product* $U \cdot V$ is a map $\mathbb{V} \times \mathbb{V} \to \mathbb{C}$ such that it is

[12] Example of a non-linear transformation let $N : x \to x^2$, so $N(ax) = a^2x^2 \neq aN(x) = ax^2$.

[13] A great joke to tell at a physics party is to ask people whether N is countably or uncountably infinite and then watch the ensuing argument.

- **Hermitian**: $U \cdot V = (V \cdot U)^*$.
- **Linear**: Given any $W \in \mathbb{V}$, $U \cdot (aV + bW) = a(U \cdot V) + b(U \cdot W)$ for $a, b \in \mathbb{C}$.
- **Positive Norm**: $U \cdot U \geq 0$, with $U \cdot U = 0$ iff $U = 0$.

The Hermitian condition imply that if V, U are complex vectors i.e. $V = (v_1, v_2, \dots)$ and $U = (u_1, u_2, \dots)$ where v_i and u_i are complex numbers, then their inner product is explicitly

$$\boxed{U \cdot V = U^\dagger V = u_1^* v_1 + u_2^* v_2 + \dots}. \tag{4.52}$$

With this definition, we can finally define Orthogonality.

(Definition) Orthogonality: Two vectors U and V are *orthogonal* to each other if their inner product vanishes $U \cdot V = 0$.

In your study of Quantum Mechanics, you have encountered the scalar product of two state vectors $|\Psi\rangle$ and $|\Phi\rangle$ of a quantum mechanic particle in x space

$$\langle \Phi | \Psi \rangle \equiv \int_{-\infty}^{\infty} dx \ \Phi(x)^* \Psi(x). \tag{4.53}$$

So often you hear people say the words "the state $|\Phi\rangle$ in x-representation is the wavefunction $\Phi(x)$". The use of the word "representation" is not a coincident – we can represent the state in any representation we like. For example, when you quantize the Harmonic Oscillator, you have used the E-representation $|n\rangle$ where n refers to the energy level of the state. But it is clear that once we have chosen a representation, the operators that act on it takes a different "mathematical form", even though the physics is the same. For example, the momentum operator in x-representation is $-i\hbar\partial_x$.

4.2.3 Reducible and Irreducible Representations

You should be convinced by now that there is a large number of representations for each group, and each representation often do not even have the same matrix dimensionality (as opposed to the number of parameters). This promiscuity poses a problem for us – which is the "best" representation for a physical problem at hand? A related problem is that many of the representations of the group are actually **equivalent** to each other in the sense that they describe the same physical thing.

What we want is a notion of breaking down all the representations into its most fundamental "building blocks". In this section, we will study this. It turns out that representations can be broadly classified as two types, **reducible** and **irreducible** representations.

4.2.4 Similarity Transforms and Equivalence of Representations

Let's begin by attacking the idea that many of the representation are actually "equivalent" to each other.

Suppose we have an element $g \in G$, and its matrix representation A_{ij} acting on some vector V_j to get another vector

$$U_i = A_{ij} V_j. \tag{4.54}$$

Now let's suppose that we are given a non-singular matrix[14] B_{ij} which for the moment *is not necessary the matrix representation any group*. This matrix can create a new vector by

$$V_i' = B_{ij} V_j \ , \ U_i' = B_{ij} U_j \tag{4.55}$$

[14] Non-singular matrices M are those whose determinants are non-zero $\det(M) \neq 0$ and hence possess inverses.

The question is now: given B, what is the new matrix representation A'_{ij}, such that

$$U'_i = A'_{ij}V'_j \ ? \tag{4.56}$$

For clarity in what follows, we will now drop the indices for now, and give the operators hats $A_{ij} = \hat{A}$ just to be clear. We write

$$\hat{B}U = \hat{B}(\hat{A}V) = \hat{B}\hat{A}\underbrace{\hat{B}^{-1}\hat{B}}_{\hat{1}}V \tag{4.57}$$

or by using $\hat{B}U = U'$ and $\hat{B}V = V'$, we get

$$U' = \hat{B}\hat{A}\hat{B}^{-1}V' \equiv \hat{A}'V'. \tag{4.58}$$

Hence, if we define $\hat{A}'$ to be

$$\boxed{\hat{A}' = \hat{B}\hat{A}\hat{B}^{-1}} \tag{4.59}$$

we find that, for any given non-singular $\hat{B}$,

$$U = \hat{A}V \longrightarrow U' = \hat{A}'V'. \tag{4.60}$$

A transformation like Eq. (4.59) is called a **Similarity Transform**. Representations which can be related to each other by a similarity transforms are **equivalent**, otherwise they are **inequivalent**.

What's the point of all this, and what is so similar about $\hat{A}$ and $\hat{A}'$? This depends on what $\hat{B}$ is of course. This is best illustrated by a common operation which you do a lot: $\hat{B}$ is a **Change of Basis Transformation**.

As an example, let's go back to the ordinary 3 vectors, so $V = \mathbf{v} = v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}$ and $U = \mathbf{u} = u_1\mathbf{i} + u_2\mathbf{j} + u_3\mathbf{k}$, with **Basis Vectors** $(\mathbf{i}, \mathbf{j}, \mathbf{k})$. Suppose now we want to execute a *linear* change of basis to $(\mathbf{i}', \mathbf{j}', \mathbf{k}')$. Such a change of basis can be written as

$$\mathbf{i}' = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k} \ , \ a_1, a_2, a_3 \in \mathbb{R} \tag{4.61}$$

and so on for $\mathbf{j}'$ and $\mathbf{k}'$. This should be familiar to you – for example when you are doing a coordinate transformation from x basis to x' basis (see for example Fig. 4.2.) In general, a linear basis transform is defined as follows.

(Definition) Linear Basis Transformation: Suppose $\{\mathbf{e}_i\}$ and $\{\mathbf{e}'_i\}$ are two different basis vectors, then a *non-singular* linear basis transform B_{ij} is defined as

$$\boxed{\mathbf{e}'_i = B_{ij}\mathbf{e}_j}. \tag{4.62}$$

This form should look familiar to you: compare this to the Conjugacy Equivalence Relation for two group elements $g_1 \sim g_2$, which we encountered in our study of Conjugacy Classes in Section 3.2.3. There we say that if $g_1 = gg_2g^{-1}$ for *any* $g \in G$, then $g_1 \sim g_2$. The "colloquial" explanation there makes even more visual sense here: in a similarity transform, we can write $\hat{A} = \hat{B}^{-1}\hat{A}'\hat{B}$, so now instead of acting on the basis vectors with $\hat{A}$, we first "move to a new basis by $\hat{B}$", act there with $\hat{A}'$, and then move back to the original basis with $\hat{B}^{-1}$.

Notice however, that here $\hat{B}$ is any non-singular matrix, and does not have to be the matrix representation of any group element whereas for the conjugacy equivalence relation, the "rotation" to the new basis is an element of the group. But, as you might imagine, the power of the similarity transform will blossom when $\hat{B}$ is actually an element of the group representation. This we will study in the coming sections.

Consider now the **trace** of the matrix

$$\boxed{\mathrm{Tr}A_{ij} = \sum_i A_{ii}}. \tag{4.63}$$

You will be asked to prove the famous **Trace Identity** in the homework, for any two matrices A, B,

$$\boxed{\mathrm{Tr}(AB) = \mathrm{Tr}(BA)}. \tag{4.64}$$

Using the identity Eq. (4.64), we can easily show that

$$\mathrm{Tr}(\hat{B}\hat{A}\hat{B}^{-1}) = \mathrm{Tr}((\hat{B}\hat{A})\hat{B}^{-1}) = \mathrm{Tr}(\hat{B}^{-1}\hat{B}\hat{A}) = \mathrm{Tr}\hat{A}. \tag{4.65}$$

In other words, the *trace of any matrix representation is invariant under a similarity transform.* In physics or math, when you see invariance under some operation, it usually means something important so we should give it a name. Traces of representations are called **Characters**. They are one of the important tools we will use to classify group representations.

(Definition) Characters: The *character* of a matrix representation D_α for any group element $g \in G$ is defined to be its trace, and we call it χ_α:

$$\boxed{\chi_\alpha = \mathrm{Tr}D_\alpha}. \tag{4.66}$$

The character is invariant under a similarity transform.

Some further remarks on characters.

- Characters can be complex or real-valued.
- Characters of representations in the same conjugacy class are equal. *Proof*: Suppose $g_1, g_2 \in G$ and $g_1 \sim g_2$, hence $\exists g$ such that $g_1 = gg_2g^{-1}$. Suppose now $D(g)$ is a representation for G so $D(g_1) = D(gg_2g^{-1})$. Taking the trace and using the trace identity $\mathrm{Tr}D(g_1) = \mathrm{Tr}D(g_2)$. $\square$.
- Characters of the *different inequivalent* representations of the same group are not necessarily equal.

We will come back and discuss characters and their uses when we discuss classification of representations.

4.2.5 Constructing Representations

Suppose a group G has two matrix representations (which may or may not be faithful) $D_1(g)$ and $D_2(g)$, which has dimensionality n and m respectively. In other words, $D_1(g)$ are $n \times n$ square matrices while $D_2(g)$ are $m \times m$ square matrices. We can now construct a *new* representation by taking the **Direct Sum** in the following way

$$D_3(g) = D_1(g) \oplus D_2(g) = \begin{pmatrix} D_1(g) & 0 \\ 0 & D_2(g) \end{pmatrix} \tag{4.67}$$

i.e. the top left corner of $D_3(g)$ is a $n \times n$ matrix, while the bottom right is a $m \times m$ matrix, and the rest are zeroes. For example, suppose

$$D_1(g) = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, \; D_2(g) = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \tag{4.68}$$

then

$$D_3(g) = \begin{pmatrix} a_{11} & a_{12} & a_{13} & & \\ a_{21} & a_{22} & a_{23} & & \\ a_{31} & a_{32} & a_{33} & & \\ & & & b_{11} & b_{12} \\ & & & b_{21} & b_{22} \end{pmatrix} \tag{4.69}$$

where empty spaces are zeroes. Such a matrix is called a **block diagonal matrix** – i.e. the diagonal of th matrix are blocks of square matrices.

It is easy to show that if $D_1(g)$ and $D_2(g)$ is a matrix representation for the group element $g \in G$, so is $D_3(g)$, viz for any two elements $g, g' \in G$, then

$$\begin{aligned} D_3(g)D_3(g') &= \begin{pmatrix} D_1(g) & 0 \\ 0 & D_2(g) \end{pmatrix} \begin{pmatrix} D_1(g) & 0 \\ 0 & D_2(g) \end{pmatrix} \\ &= \begin{pmatrix} D_1(g)D_1(g') & 0 \\ 0 & D_2(g)D_2(g') \end{pmatrix} \\ &= \begin{pmatrix} D_1(gg') & 0 \\ 0 & D_2(gg') \end{pmatrix} \\ &= D_3(gg') \end{aligned} \tag{4.70}$$

where we have used the fact that $D_1(g)D_1(g') = D_1(gg')$ since $D_1(g)$ is a matrix representation, and similarly for D_2, going from line 2 and 3.

In fact, we can keep going by adding more representations together, i.e. $D_4(g) = D_3(g) \oplus D_2(g) \oplus D_1(g)$. So in general, we can construct all the representations to our heart's content with the following relation

$$\mathcal{D}(g) = \sum_{\oplus} D_i(g) \tag{4.71}$$

where the $\sum_{\oplus}$ means a direct sum of the representations $D_i(g)$.

But, like many things in life, plenty does not equal good. Here we are simply generating bloated representations without learning anything new. However, our new found ability to make new representations out of old ones begs the question: *are the old representations themselves made out of even older ones?*. Ask another way : *are there representations which cannot be constructed using the direct sum method that we just learned?* The answer to these questions is a resounding yes, there exist "building blocks" representations. Such building block representations are called **irreducible representations** or **irreps** for short. On the other hand, the representations that can be *reduced* into a block-diagonal form by some suitable *similarity transforms* are called **reducible representations** — note that some reducible representations do not "look" block-diagonal initially but can be made into a block-diagonal form with similarity transforms.

Let's say that again since it is so important: suppose you are given a reducible representation that is not in block-diagonal form, then there exists a similarity transform that will bring the representation into the block-diagonal form, with each "block" being an irreducible representation. Note that this definition means that *all one dimensional representations are irreducible by definition.*

So now our question becomes: *given some representations, how do we know that it is reducible or irreducible?* We can try a large number of similarity transforms and hope by trial and error we can reduce the representation (or die trying). But that's not efficient – fortunately we have some theorems and lemmas to help us and we will now study them. However, a good question is to ask "why are irreducible representations" so important? The answer depends on whether you are an mathematician or a physicist. A mathematician's answer would be "because they are the fundamental building blocks of which all other representations are built upon". In particular, irreducible representations of any group can be classified completely by their dimensionality and their characters – a powerful restriction.

On the other hand, a physicist answer would depends on the application of group theory to the physical problem at hand – for example, in particle physics irreps are states of fundamental particles, and a reducible representation will mix them up in some complicated way.

Before we start on that journey, let's look at a concrete case for reducible and irreducible reps.

Example: Reducible and Irreducible representations of D_3.

The Dihedral $D_3 = \{e, R, R^2, m_1, m_2, m_3\}$ is an order 6 non-abelian group[15]. Consider 2×2 representations U with real entries

$$U : D_3 \to SO(2, \mathbb{R}). \tag{4.72}$$

The identity element is always mapped to the identity matrix

$$U(e) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{4.73}$$

but now we have a plethora of choices for the other 5 elements. One such representation which is *irreducible* is the following

$$U(m_1) = \frac{1}{2}\begin{pmatrix} 1 & -\sqrt{3} \\ -\sqrt{3} & -1 \end{pmatrix}\ ,\ U(m_2) = \frac{1}{2}\begin{pmatrix} 1 & \sqrt{3} \\ \sqrt{3} & -1 \end{pmatrix}\ ,\ U(m_3) = \frac{1}{2}\begin{pmatrix} -2 & 0 \\ 0 & 2 \end{pmatrix} \tag{4.74}$$

$$U(R) = \frac{1}{2}\begin{pmatrix} -1 & -\sqrt{3} \\ \sqrt{3} & -1 \end{pmatrix}\ ,\ U(R^2) = \frac{1}{2}\begin{pmatrix} -1 & \sqrt{3} \\ -\sqrt{3} & -1 \end{pmatrix}. \tag{4.75}$$

This representation has a nice geometrical explanation as rotation and reflection matrices on a triangle with edges of unit length and centroid at $(0, 0)$ as shown in Fig. 4.4.

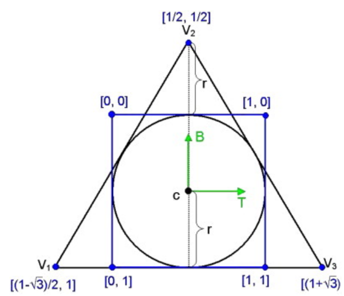


Figure 4.4: An equilateral triangle. The 3 vertices labeled V_1, V_2 and V_3 form a vector space on which the 2 dimensional matrix representation U act on to keep the symmetry of the triangle invariant.

We have simply asserted that it is irreducible of course and have not proven it. Let's leave the general proof aside for later when we have developed more technology. But for 2×2 matrices, there is an easy way to check – suppose that U is reducible, then it is clear that the only way it can be reduced is to of the form

$$U(g) = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}\ ,\ a, b \in \mathbb{C} \tag{4.76}$$

i.e. a, b are simply some complex numbers (recall that similarity transforms can be executed by complex matrices). But matrices of the form Eq. (4.76) clearly *commute* with each other. Furthermore, similarity transforms *do not change the commutativity of the representations* which you can easily prove to yourself. Hence if the representation above have been reducible, they will also commute with each other. You can check for yourself that they do not commute, so they are hence irreducible. Note that this "test" of reducibility only works for 2×2 representations – can you see why?

On the other hand, a representation which is *inequivalent* to the above is given by

$$T(m_1) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\ ,\ T(m_2) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\ ,\ T(m_3) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{4.77}$$

[15] It is also isomorphic to the permutation group of 3 objects.

$$T(R) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \ , \ T(R^2) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} . \tag{4.78}$$

This representation is clearly *reducible* since it is already in block-diagonal form

$$T(g) = \begin{pmatrix} D_1(g) & 0 \\ 0 & D_2(g) \end{pmatrix} \tag{4.79}$$

where $D_1(g) = \{1, 1, 1, 1, 1, 1\}$ and $D_2(g) = \{1, -1, -1, -1, 1, 1\}$ are two *inequivalent* 1×1 real representations of D_3. It is also *unfaithful*, since $\text{Ker}(T) = \{e, R, R^2\}$.

Finally, one thing you can check for yourself is that both representations U and T are *unitary*, i.e.

$$U^\dagger U = e \ , \ T^\dagger T = e \tag{4.80}$$

The group representations do not have to be unitary of course, but it will turn out that *every representation can be brought into unitary form by a similarity transform.* This theorem is one of the tools that will allow us to begin restricting the number of representations which we have to study. And we will begin our journey on how to do that by proving this theorem.

4.2.6 Unitary Representations

At the start of this chapter, in section 4.1.1, we discussed several properties of matrices which I hope you have learned and remembered from a class on linear algebra. Let's summarize here. For any matrice with complex entries A_{ij}, it is

- **Special** if $\det A = 1$
- **Trace-free** if $\text{Tr}A = 0$
- **Orthogonal** if $A^T A = e$, where T denotes **Transposition**, $(A^T)_{ji} = A_{ij}$.
- **Hermitian** if $A^\dagger = A$, where $A^\dagger = (A^T)^*$ denotes **Hermitian Conjugation**
- **Unitary** if $A^\dagger A = AA^\dagger = e$.

Further more, you can prove to yourself that

$$\boxed{(AB)^\dagger = B^\dagger A^\dagger \ , \ (AB)^T = B^T A^T .} \tag{4.81}$$

Since the title of this section is Unitary Representations, let's focus on unitary matrices.

Proposition: Unitary Matrix Operators preserve the inner product of two complex vectors.

Proof: Suppose we have two complex vectors U, V, with the inner product

$$U \cdot V = \mathcal{N} \tag{4.82}$$

where $\mathcal{N} \in \mathbb{C}$. Let $\hat{A}$ be a unitary matrix, such that $V' = \hat{A}V$ and $U' = \hat{A}U$, then

$$U' \cdot V' = (\hat{A}U)^\dagger (\hat{A}V) = U^\dagger \hat{A}^\dagger \hat{A} V = U \cdot V = \mathcal{N} \ \ \square . \tag{4.83}$$

Corollary: Norms of complex vectors are preserved under unitary matrix operations $U \cdot U = U' \cdot U'$.

In ordinary 3 vectors, the norm of a vector $\mathbf{a} \cdot \mathbf{a} = |a|^2$ denotes its length. If we rotate this vector in any direction, the length remains the same. Hence one can think of unitary matrices as "generalized rotations".

We will now state an important theorem.

(Theorem) Unitary Representations: *Every representation can be made unitary by a suitable similarity transformation.*

Proof: To prove this theorem requires two lemmas.

(Lemma 1) : Suppose $\hat{H}$ is a Hermitian Matrix with *normalized* eigenvectors $\{v_i\}$ which are column matrices, and *real*[16] eigenvalues λ_i, i.e.

$$\hat{H}v_i = \lambda_i v_i \tag{4.84}$$

then we can construct a unitary matrix

$$U = (v_1\ v_2\ v_3\ \ldots). \tag{4.85}$$

Proof: Since the eigenvectors v_i are normalized, and hence *orthonormal* to each other i.e.

$$v_i \cdot v_j = v_i^\dagger v_j = \delta_{ij} \tag{4.86}$$

then

$$U^\dagger U = UU^\dagger = e \tag{4.87}$$

and hence U *is a unitary matrix.* □.

(Lemma 2) : We can diagonalize $\hat{H}$ by a similarity transform using U.

Proof: Acting with $\hat{H}$ on U, we get

$$\hat{H}U = U\Lambda \tag{4.88}$$

where Λ is a diagonal matrix consisting of the eigenvalues of $\hat{H}$

$$\Lambda = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix} \tag{4.89}$$

□.

With these Lemmas in hand, we can now prove the Theorem.

Suppose the group G is represented by a set of square $n \times n$ matrices $\{D_\alpha(g)\}$, where α labels the matrices. For clarity, we will write D_α and drop the (g) for now. We can construct a Hermitian matrix H in the following way

$$H = \sum_\alpha D_\alpha D_\alpha^\dagger \tag{4.90}$$

since $H^\dagger = \sum_\alpha (D_\alpha D_\alpha^\dagger)^\dagger = H$. This Hermitian matrix has a set of eigenvalues $\{\lambda_i\}$ and eigenvectors $\{v_i\}$ which we can use to construct the unitary matrix using Lemma 1

$$U = (v_1\ v_2\ v_3\ \ldots) \tag{4.91}$$

and

$$\Lambda = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}. \tag{4.92}$$

Using Lemma 2, we can diagonalize H to get

$$\Lambda = U^{-1}HU = U^{-1}(\sum_\alpha D_\alpha D_\alpha^\dagger)U. \tag{4.93}$$

[16] Hopefully you have learned that Hermitian matrices have real eiggenvalues somewhere in your linear algebra or quantum mechanics classes.

Now inserting $UU^\dagger = e$ in the middle, and using $U^\dagger = U^{-1}$ we get

$$\Lambda = U^{-1}HU = \sum_\alpha (U^\dagger D_\alpha U)(U^\dagger D_\alpha^\dagger U) \tag{4.94}$$

and by defining the new matrix

$$\tilde{D}_\alpha \equiv U^\dagger D_\alpha U \tag{4.95}$$

this become[17]

$$\Lambda = \sum_\alpha \tilde{D}_\alpha \tilde{D}_\alpha^\dagger. \tag{4.96}$$

Now we want to claim that the diagonal elements of Λ are *positive definite*. We can see this by noting that the $i-i$ element of Λ is given by (restoring the indices of the matrices)

$$\begin{aligned}\Lambda_{ii} &= \sum_\alpha \sum_j (\tilde{D}_\alpha)_{ij}(\tilde{D}_\alpha)^\dagger_{ji} \\ &= \sum_\alpha \sum_j |\tilde{D}_{\alpha ij}|^2 > 0 \end{aligned} \tag{4.97}$$

where we have restored the summation of the matrix elements over j for clarity. In the last line, since α is summed over the identity matrix $\tilde{D}(e) = U^\dagger D(e) U = e$, it must always be non-zero[18]

Since Λ has non-zero positive entries, it must possess an inverse Λ^{-1}, and since the entries are positive definite, $\Lambda^{1/2}$ must also possess real and non-zero diagonal entries

$$\Lambda^{1/2} = \begin{pmatrix} \lambda_{11}^{1/2} & & & \\ & \lambda_{22}^{1/2} & & \\ & & \ddots & \\ & & & \lambda_{ii}^{1/2} \end{pmatrix}, \tag{4.98}$$

and similarly for $\Lambda^{-1/2}$. With all these construction, we now assert that the new representation T_α

$$\boxed{T_\alpha \equiv \Lambda^{-1/2}\tilde{D}_\alpha \Lambda^{1/2}} \tag{4.99}$$

is *unitary* and equivalent to D_α. The fact that it is equivalent to D_α should be clear, since

$$\begin{aligned} T_\alpha &= (\Lambda^{-1/2}U^\dagger)D_\alpha(U\Lambda^{1/2}) \\ &= (\Lambda^{-1/2}U^{-1})D_\alpha(U\Lambda^{1/2}) \\ &= (U\Lambda^{1/2})^{-1}D_\alpha(U\Lambda^{1/2}) \end{aligned} \tag{4.100}$$

i.e. the similarity transform is executed by the matrix $U\Lambda^{1/2}$.

To show that it is unitary is a bit more tricky, using $\Lambda^{1/2} = (\Lambda^{1/2})^\dagger$ since it is real,

$$\begin{aligned} T_\alpha T_\alpha^\dagger &= \Lambda^{-1/2}\tilde{D}_\alpha\Lambda^{1/2}(\Lambda^{-1/2}\tilde{D}_\alpha\Lambda^{1/2})^\dagger \\ &= \Lambda^{-1/2}\tilde{D}_\alpha\Lambda\tilde{D}_\alpha^\dagger\Lambda^{-1/2} \end{aligned} \tag{4.101}$$

and now using $\Lambda = \sum_\beta \tilde{D}_\beta \tilde{D}_\beta^\dagger$, we get

$$\begin{aligned} T_\alpha T_\alpha^\dagger &= \sum_\beta \Lambda^{-1/2}(\tilde{D}_\alpha\tilde{D}_\beta)(\tilde{D}_\beta^\dagger\tilde{D}_\alpha^\dagger)\Lambda^{-1/2} \\ &= \sum_\beta \Lambda^{-1/2}(\tilde{D}_\alpha\tilde{D}_\beta)(\tilde{D}_\alpha\tilde{D}_\beta)^\dagger\Lambda^{-1/2} \end{aligned} \tag{4.102}$$

[17]Can you see why $D_\alpha \sim \tilde{D}_\alpha$?

[18]Not so fast, some of you might say: what if $|G| = \infty$, i.e. the order of G is countably infinite, so we have to sum α to infinite number of elements. How sure are we in this case that Λ_{ii} is not infinity, and hence the diagonal matrix is not well defined? It turns out that such pathological cases can indeed occur for *some* countably infinite order groups. In such cases, the theorem breaks down and all hell breaks lose and much angst is to be had. Fortunately, for the groups that physics is mostly concerned with, especially the Lie Groups, this does not happen.

where we have done some creative bracketing in the second line. But now, using closure of the group elements, this mean that

$$\tilde{D}_\alpha \tilde{D}_\beta = \tilde{D}_\gamma \tag{4.103}$$

and so for each incident of β in the summation, we get some element $\tilde{D}_\gamma$ – which is *different*. This is because for fixed $\tilde{D}_\alpha$, if there exist two elements $\tilde{D}_{\beta_1}$ and $\tilde{D}_{\beta_2}$ which gives us the same element $\tilde{D}_\gamma$, implies that

$$\tilde{D}_\gamma^{-1}\tilde{D}_\gamma = \tilde{D}_{\beta_1}^{-1}\tilde{D}_\alpha^{-1}\tilde{D}_\alpha\tilde{D}_{\beta_2} = \tilde{D}_{\beta_1}^{-1}\tilde{D}_{\beta_2} = e \tag{4.104}$$

or by uniqueness of inverses $\tilde{D}_{\beta_1} = \tilde{D}_{\beta_2}$. (Sometimes this fact is called the **Rearrangement Theorem**.) Hence we can sum over γ in Eq. (4.102) instead of β to get

$$T_\alpha T_\alpha^\dagger = \sum_\gamma \Lambda^{-1/2}\tilde{D}_\gamma\tilde{D}_\gamma^\dagger\Lambda^{-1/2} = \Lambda^{-1/2}\Lambda\Lambda^{-1/2} = 1 \tag{4.105}$$

and proving that T_α is unitary, and completing the proof of our theorem. □.

4.2.7 Schur's Lemmas and the Orthogonality Theorem of Representations

(**You should know the Lemmans and Theorems, but not the proofs**.)

After the last theorem on unitary representations, we finally have garnered enough infrastructure to discuss the tools we need to identify and classify irreps. These tools are encapsulated in two lemmas called **Schur's first and second Lemmas** and the **Orthogonality Theorem of Representations**. Let's state these theorems, and then prove them later.

Schur's First Lemma: *Suppose $D_\alpha(g)$ is an* irreducible *representation of the group G with elements g and order $|G|$. If any $n \times n$ non-zero matrix M which commutes with every D_α, i.e*

$$MD_\alpha = D_\alpha M \tag{4.106}$$

then M is the identity matrix e times some constant $M = ce$ for some constant c.

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*

$$MD_\alpha = D'_\alpha M \tag{4.107}$$

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 $M = 0$.

Since inequivalent representations of the same group G are still related by the group laws, one can imagine a relationship between the matrix elements of of these representations. This relationship is encoded in the so-called **Fundamental Orthogonality Theorem of Representations**, which is in two parts.

Orthogonality Theorem (Single irreps): *Suppose D_α is an irreducible representations with dimensions n of the group G with order $|G|$. Then*

$$\boxed{\sum_\alpha (D_\alpha)^*_{ij}(D_\alpha)_{kl} = \frac{|G|}{n}\delta_{ik}\delta_{jl}.} \tag{4.108}$$

Orthogonality Theorem (Different irreps): *Suppose D_α and D'_α are two* inequivalent *irreducible representations with dimensions n and n' of the group G with order $|G|$, respectively. Then*

$$\boxed{\sum_\alpha (D_\alpha)^*_{ij}(D'_\alpha)_{kl} = 0,} \tag{4.109}$$

and if D_α and D'_α are equivalent, it reduces to the single irrep case above.

Suppose now we have $\mathcal{M}$ number of inequivalent irreps of the group G each with dimensionality n_m we can label the irreps by D^m_α with $m = 1, 2, \ldots, \mathcal{M}$, and then we can combine both versions of the Orthogonality theorem in the following way

$$\boxed{\sum_\alpha (D^m_\alpha)^*_{ij}(D^{m'}_\alpha)_{kl} = \frac{|G|}{n_m}\delta_{ik}\delta_{jl}\delta^{mm'}.} \tag{4.110}$$

The Orthogonality theorem invites us to think of the representations and their matrix elements as some gigantic Vector Space. For example, in Eq. (4.110), if we fix the index $m = m'$, then one can think of the set $\{D^m_{1ij}, D^m_{2ij}, \ldots, D^m_{\alpha ij}\}$ as an α-dimensional column vector, and since we can run $i, j = 1, 2, \ldots, n_m$, there are n_m^2 of these vectors, while Eq. (4.110) expresses the fact that each of these vectors are orthogonal to each other.

Now if we allow m and m' to be free, the $\delta^{mm'}$ on the RHS of Eq. (4.110) means that any vectors which are formed by D^m and $D^{m'}$ will be mutually orthogonal to each other, so the total number of vectors in the whole set is $\sum_{m=1}^{\mathcal{M}} n_m^2$ (i.e. each representation labeled m has n_m^2 vectors as we described in the previous paragraph). But since the order of the group $|G|$ gives us the restriction on the total number of such vectors, the following inequality holds

$$\sum_{m=1}^{\mathcal{M}} n_m^2 \leq |G|. \tag{4.111}$$

In fact, Eq. (4.111) is actually an *equality*, i.e.

$$\boxed{\sum_{m=1}^{\mathcal{M}} n_m^2 = |G|}, \tag{4.112}$$

a relation which is *an incredibly powerful restriction on the total number of irreducible representations $\mathcal{M}$ of the group.* We will discuss this when we study Characters in section 4.2.9. This is one of the most important formula so you should memorize it.

Now that we have reviewed our theorems, we shall prove them.

Proof (Schur's First Lemma): Let D_α be an unitary[19] irreducible representation of G. Suppose there exist a matrix M such that

$$MD_\alpha = D_\alpha M \tag{4.113}$$

Taking the Hermitian conjugation, and then using the fact that D_α is unitary we get

$$M^\dagger D_\alpha = D_\alpha M^\dagger. \tag{4.114}$$

Though M is not necessary Hermitian, we can construct Hermitian matrices from M by

$$H = M + M^\dagger \ , \ H' = i(M - M^\dagger) \tag{4.115}$$

[19] Which we can take WLOG since we have proven that any representation can be made unitary with a suitable similarity trasnform.

so that $H^\dagger = H$ and $H'^\dagger = H'$ are true trivially, and that $2M = H - iH'$. The reason we want to use Hermitian matrices instead of M is because we can apply Lemma 1 from section 4.2.6, i.e. we can find a unitary matrix U such that H is diagonlized $\Lambda = U^{-1}HU$. Now, instead of M we want to prove the lemma for

$$HD_\alpha = D_\alpha H \tag{4.116}$$

and then since M is some linear combination of H and H', Schur's first Lemma will follow. Applying a unitary transform with U on Eq. (4.116), we get

$$U^{-1}HUU^{-1}D_\alpha U = U^{-1}D_\alpha UU^{-1}HU \tag{4.117}$$

and defining $\tilde{D}_\alpha = U^{-1}D_\alpha U$, we get

$$\Lambda\tilde{D}_\alpha = \tilde{D}_\alpha\Lambda. \tag{4.118}$$

Restoring the labels for the matrix elements, and rewriting we get

$$\Lambda_{ij}(\tilde{D}_\alpha)_{jk} - (\tilde{D}_\alpha)_{ij}\Lambda_{jk} = 0 \tag{4.119}$$

and now since Λ_{ij} is a diagonal matrix, we get (dropping the α for simplicity)

$$\Lambda_{ij}(\tilde{D}_\alpha)_{jk} = \begin{pmatrix} \Lambda_{11} & 0 & \dots \\ 0 & \Lambda_{22} & \dots \\ \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} \tilde{D}_{11} & \tilde{D}_{12} & \dots \\ \tilde{D}_{21} & \tilde{D}_{22} & \dots \\ \vdots & \vdots & \ddots \end{pmatrix} = \begin{pmatrix} \Lambda_{11}\tilde{D}_{11} & \Lambda_{11}\tilde{D}_{12} & \dots \\ \Lambda_{22}\tilde{D}_{21} & \Lambda_{22}\tilde{D}_{22} & \dots \\ \vdots & \vdots & \ddots \end{pmatrix} \tag{4.120}$$

and

$$(\tilde{D}_\alpha)_{ij}\Lambda_{jk} = \begin{pmatrix} \tilde{D}_{11} & \tilde{D}_{12} & \dots \\ \tilde{D}_{21} & \tilde{D}_{22} & \dots \\ \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} \Lambda_{11} & 0 & \dots \\ 0 & \Lambda_{22} & \dots \\ \vdots & \vdots & \ddots \end{pmatrix} = \begin{pmatrix} \Lambda_{11}\tilde{D}_{11} & \Lambda_{22}\tilde{D}_{12} & \dots \\ \Lambda_{11}\tilde{D}_{21} & \Lambda_{22}\tilde{D}_{22} & \dots \\ \vdots & \vdots & \ddots \end{pmatrix}. \tag{4.121}$$

We can easily see that the diagonal terms in Eq. (4.119) hence are automatically satisfied, leaving the following relation

$$(\tilde{D}_\alpha)_{ij}(\Lambda_{ii} - \Lambda_{jj}) = 0. \tag{4.122}$$

Now consider three cases

- **Case 1**: If Λ_{ii} are equal, then Eq. (4.122) vanishes automatically and hence $\tilde{D}_{ij}$ can have any value, so it can form an irreducible representation. But since $\Lambda_{ii} = c\forall\ i$, we can write $\Lambda_{ij} = c\delta_{ij}$ or $\Lambda = ce$, a unit matrix times a constant. If Λ is proportional to the unit matri, then so is H and hence so is M.

- **Case 2**: If $\Lambda_{ii} \neq \Lambda_{jj}$ for $i \neq j$ (i.e. H has distinct eigenvalues), then $\tilde{D}_{ij} = 0$ for all the off-diagonal elements. Hence $\tilde{D}_{ij}$ is a diagonal matrix. But we know that diagonal representations are clearly also block-diagonal, and hence is reducible.

- **Case 3**: If Λ_{ii} is equal for only *some* $i = 1, 2, \dots, k$ where $k < n$, then following Case 2, $\tilde{D}_{ij}$ must vanish in the diagonal for any paris of i, j which do not have equal values for Λ_{ii}. So this follows that that part of the matrix must be diagonal just like in Case 2, and hence is block-diagonal.

Hence we have showed that if M is not proportional to the unit matrix, D_α must necessarily be reducible. Only in Case 1, if a non-zero matrix commutes with the elements of an irreducible representation, it must be proportional to the unit matrix. □.

Proof (Schur's Second Lemma): Let D_α and D'_α be two inequivalent unitary irreducible representations of group G with dimenions n and n' respectively. Suppose there exist a $n \times n'$ matrix M such that

$$MD_\alpha = D'_\alpha M. \tag{4.123}$$

Following the same steps as the 1st Lemma, we take the Hermitian Conjugation and using the fact the D_α and D'_α are unitary, $D^\dagger = D^{-1}$, we get

$$D_\alpha^{-1} M^\dagger = M^\dagger D'^{-1}_\alpha. \tag{4.124}$$

Multiplying both sides by M from the left, and using $D = D^{-1}$ in Eq. (4.123) to replace D with D with D', we get

$$D'_\alpha{}^{-1}(MM^\dagger) = (MM^\dagger)D'_\alpha{}^{-1}. \tag{4.125}$$

The matrix $MM^\dagger$ is a $n' \times n'$ *square* matrix, but this has the form of Schur's First Lemma Eq. (4.106), this means that since D'_α is an element of a unitary *irreducible* representation,$MM^\dagger$ must have the form

$$MM^\dagger = ce \ , \ c \in \mathbb{C}. \tag{4.126}$$

We now want to investigate M. Suppose $n = n'$ and $c \neq 0$, Eq. (4.126) can be rearranged to become $M^{-1} = (1/c)M^\dagger$, giving us a relationship between M^{-1} and $M^\dagger$ – and since our assumption is that M is non-zero, then M^{-1} is non-zero. This implies that M *is invertible*, so it can act to make D and D' equivalent via a similarity transform

$$D_\alpha = M^{-1} D'_\alpha M. \tag{4.127}$$

Suppose again $n = n'$ but $c = 0$, this means that $MM^\dagger = 0$, and the only matrix M that satisfies this equation is M itself is a zero matrix, i.e. $M = 0$.

In the case where $n \neq n'$, equation $MM^\dagger = ce$ implies that $M^\dagger$ is the *inverse* of M. But we know from our study of matrices that *non-square matrices do not possess an inverse*[20], and hence the only way we can satisfy this equation is $c = 0$, or that $M = 0$.

This completes our proof of Schur's second Lemma. $\Box$.

Proof (Orthogonality Theorem): This is a long proof, so we will skip some algebra – you are encouraged to work through it.

Let D_α and D'_α be any two unitary representations of G with dimensions n and n'. Let B be any arbitrary non-singular matrix[21] of dimensions $n \times n'$, then we can construct a matrix M by summing up α in the following way

$$M \equiv \sum_\alpha D'_\alpha B D_\alpha^{-1}. \tag{4.128}$$

We now want to show that Eq. (4.128) is satisfies the equation

$$D'_\alpha M = M D_\alpha. \tag{4.129}$$

Multiplying Eq. (4.128) on the left by D'_β, and by $D_\beta{}^{-1} D_\beta = e$ on the right, and after some fancy footwork we get

$$D'_\beta M = \sum_\alpha (D'_\beta D'_\alpha) B (D_\alpha^{-1} D_\beta{}^{-1}) D_\beta \tag{4.130}$$

and now using the closure of the representation $D'_\beta D'_\alpha = D'_\gamma$ and relabeling the summation from α to γ (see section 4.2.6 for the proof), then we get Eq. (4.129) as asserted.

[20]You might argue that it may possess a *left inverse*, but one can also show that $M^\dagger M = ce$ so it must possess both left and right inverses for the equation to be true.

[21]What is B? Secretly, it is the linear mapping between the vector spaces acted on by D and that acted on by D' – i.e. it is a **linear basis transformation** as we have studied in 4.2.4.

Now Eq. (4.129) has the form of which Schur's Second Lemma applies, which tells us that for $M = 0$, either $n \neq n'$ or $n = n'$ and D and D' are inequivalent. In these condition, then from Eq. (4.128) is (restoring the matrix indices)

$$M_{ij} = \sum_{\alpha} \sum_{k,l} (D'_\alpha)_{ik} B_{kl} (D_\alpha^{-1})_{lj} = 0. \tag{4.131}$$

Considering each (i, j) element, and since B_{kl} is arbitrary, this means that

$$\sum_{\alpha} \sum_{k,l} B_{kl} [(D'_\alpha)_{ik} (D_\alpha^{-1})_{lj}] = 0. \tag{4.132}$$

or that each element k, l in this sum must vanish (since B_{kl} is arbitrary)

$$\sum_{\alpha} (D'_\alpha)_{ik} (D_\alpha^{-1})_{lj} = 0. \tag{4.133}$$

and using the unitary condition $(D^{-1})_{lj} = D^\dagger_{lj} = D^*_{jl}$ which is the Orthogonality Theorem for different Irreps

$$\sum_{\alpha} (D'_\alpha)_{ik} (D^*_\alpha)_{jl} = 0. \tag{4.134}$$

For the case when $n = n'$ and D is equivalent to D' so we will drop the prime, this is the case of Schur's First Lemma, so M must be proportional to the identity matrix $M = ce$. Eq. (4.128) the becomes

$$M_{ij} = \sum_{\alpha} \sum_{k,l} (D_\alpha)_{ik} B_{kl} (D_\alpha^{-1})_{lj} = c\delta_{ij} \tag{4.135}$$

Taking the trace of this equation, we get $\mathrm{Tr}\delta_{ij} = n$, and using the Trace Identity, we get

$$cn = \sum_{\alpha} \mathrm{Tr}B = |G|\mathrm{Tr}B \tag{4.136}$$

where the second equality we have used the fact that there are $|G|$ copies of B_{kl} in the summation. This gives us c as a function of the properties of the group representations and B

$$c = \frac{|G|}{n}\mathrm{Tr}B = \frac{|G|}{n} \sum_{k,l} B_{kl}\delta_{kl}. \tag{4.137}$$

Plugging Eq. (4.137) back into Eq. (4.135) we get

$$\sum_{k,l} B_{kl} \left[\sum_{\alpha} (D_\alpha)_{ik} (D_\alpha^{-1})_{lj} - \frac{|G|}{n}\delta_{kl}\delta_{ij} \right] = 0 \tag{4.138}$$

and again using the argument that since B is an arbitrary matrix, each element of the term in the square brackets must vanish

$$\left[\sum_{\alpha} (D_\alpha)_{ik} (D_\alpha^{-1})_{lj} - \frac{|G|}{n}\delta_{kl}\delta_{ij} \right] = 0 \tag{4.139}$$

and using $(D^{-1})_{lj} = D^\dagger_{lj} = D^*_{jl}$, we get the Orthogonality Theorem for equivalent unitary representations

$$\sum_{\alpha} (D_\alpha)_{ik} (D^*_\alpha)_{jl} = \frac{|G|}{n}\delta_{kl}\delta_{ij}, \tag{4.140}$$

and hence the proof is complete $\square$.

Finally, before we proceed, there is a "toy version" of the Orthogonality Theorem for characters, we studied previously in 4.2.3. There, we introduced *characters* of representations as mathematical constructs – it is the trace of the matrix representation

$$\chi_\alpha = \mathrm{Tr}\ D_\alpha. \tag{4.141}$$

In this section, we will combine this with the Orthogonality Theorem to come up with a set of important results on irreducible representations which will allow us a systematic way of determining the irreducibility of any representations – and hence answering the question that we posed a while ago: *how do we tell if a representation is irreducible or not.* Fortunately, while the mathematical infrastructure is somewhat complicated to prove, the results are quite easy to apply.

Recall, that again in section 4.2.3, we showed that elements of a group which are in the same conjugacy class has the same character since if $g_2 \sim g_1$ for $g_1, g_2 \in G$, this means that $\exists\, g$ such that $g_1 = g g_2 g^{-1}$

$$\mathrm{Tr}D(g_1) = \mathrm{Tr}D(gg_2g^{-1}) = \mathrm{Tr}D(gg^{-1}g_2) = \mathrm{Tr}D(g_2). \tag{4.142}$$

or $\chi(g_1) = \chi(g_2)$.

We can now use this fact to prove the simplified version of the Orthogonality Theorem for characters:

(Theorem) Orthogonality Theorem for Characters: *Suppose D^m_α and D'^m_α are the m-th and m'-th inequivalent irreducible unitary representation of the element α of a group G, then*

$$\boxed{\sum_\alpha (\chi^m_\alpha)^*(\chi^{m'}_\alpha) = |G|\delta_{mm'}}. \tag{4.143}$$

Proof: Let's begin with the Orthogonality Theorem for single irreps (see section 4.2.7)

$$\sum_\alpha (D_\alpha)^*_{ij}(D_\alpha)_{kl} = \frac{|G|}{n}\delta_{ik}\delta_{jl}. \tag{4.144}$$

Since the indices (i, j, k, l) are all different in Eq. (4.144), so one can think of the LHS as 4-index *tensor object*. We can now **contract** over the pair of indices (i, j) and (k, l), which means we *sum* over all the terms $i = j$ and $k = l$ so now the equation becomes (adding in the sum over i, k for emphasis)

$$\sum_\alpha \sum_{i,k} (D_\alpha)^*_{ii}(D_\alpha)_{kk} = \frac{|G|}{n}\sum_{i,k}\delta_{ik}\delta_{ki}. \tag{4.145}$$

But $\sum_i (D^*_\alpha)_{ii} = \mathrm{Tr}D^*_\alpha$, and $\sum_k (D_\alpha)_{kk} = \mathrm{Tr}D_\alpha$, i.e. *contraction* is nothing but selective "trace" over the indices of a multi-index object. Also, on the RHS

$$\sum_{ik}\delta_{ik}\delta_{ki} = \sum_i \delta_{ii} = n \tag{4.146}$$

which is the dimensions of the matrix representation for D. So Eq. (4.145) becomes

$$\boxed{\sum_\alpha \mathrm{Tr}D^*_\alpha \mathrm{Tr}D_\alpha = |G|} \tag{4.147}$$

If we now look at the General Orthogonality theorem with different representations, using Eq. (4.110) instead of Eq. (4.108) as our starting point,

$$\sum_\alpha (D^m_\alpha)^*_{ij}(D^{m'}_\alpha)_{kl} = \frac{|G|}{n_m}\delta_{ik}\delta_{jl}\delta^{mm'} \tag{4.148}$$

the $\delta_{mm'}$ just gets carried around, and we obtain the following

$$\sum_\alpha \mathrm{Tr}(D^m_\alpha)^*\mathrm{Tr}D^{m'}_\alpha = |G|\delta_{mm'} \tag{4.149}$$

and using Eq. (4.141) we get the theorem we intend to prove $\square$.

(Corollary): Recall that matrix representations belonging to the same conjugacy class have equal characters. Suppose group G has $\mathcal{C}$ conjugacy classes, and for each conjugacy class labeled by $\mu =$

$1, 2, 3, \ldots \mathcal{C}$ there are N_μ elements, then we can rewrite Eq. (4.143) as a sum over conjugacy classes instead of sum over all elements of G

$$\boxed{\sum_{\mu}^{\mathcal{C}} N_\mu (\chi_\mu^m)^* (\chi_\mu^{m'}) = |G|\delta_{mm'}}. \tag{4.150}$$

4.2.8 Decomposition of Reducible Representations

The previous section is long, and we proved a bunch of theorems. Now we are going to reap the first rewards of our diligence.

We have spent a lot of time discussing irreps, but what about reducible representations? In section 4.2.5, we show how, given any set of representations (regardless of whether they are reducible or irreps), we can construct reducible representations by *direct sum* of irreps $D^m(g)$

$$\mathcal{D}_\alpha = D^1_\alpha \oplus D^2_\alpha \oplus D^3_\alpha \oplus D^m_\alpha \oplus \cdots \oplus D^1_\alpha \oplus \ldots \tag{4.151}$$

where m labels inequivalent irreps where $m = 1, 2, 3, \ldots, \mathcal{M}$, where $\mathcal{M}$ is the total number of *inequivalent* irreps in the theory. As we have discussed, the fundamental building blocks for representations are the irreps. So, although we can build any other representations from reducible representations, let us restrict ourselves to the irreps as building blocks. Notice that we can have *repeated* irreps, as illustrated above, where $m = 1$ appears twice or more.

Now remember that characters of irreps are invariant under similarity transforms, so if we take the trace of Eq. (4.151) we get

$$\chi_\alpha = \sum_m a_m \chi_\alpha^m \tag{4.152}$$

where χ_α is the character for element α, and a_m is a *non-complex* integer which tells us *how many times the m-th irrep appear in the reducible form.* Next we use the Orthogonality Theorem for characters Eq. (4.150): multiply Eq. (4.152) by $\chi_\alpha^{m'*}$, and then summing over all α, we get on the RHS

xxxxxxxxxxx

$$\sum_m a_m \underbrace{\sum_\alpha \chi_\alpha^{m'*}\chi_\alpha^m}_{|G|\delta_{mm'}} = |G|a_{m'} \tag{4.153}$$

so plugging this back to Eq. (4.152) we get

$$\boxed{a_{m'} = \frac{1}{|G|}\sum_\alpha \chi_\alpha^{m'*}\chi_\alpha}. \tag{4.154}$$

The way to think about Eq. (4.154) is to think of the big reducible representation as a *linear sum* (the $\oplus$) over the irreps where the irreps are the "eigenfunctions", so Eq. (4.154) is the equation which *picks* out the *coefficients* of these irreps[22].

[22] Here is it in hardcore equation form. Suppose we have a function $\psi(x)$, and a *complete set* of eigenfunctions $u_i(x)$. We can then describe $\psi(x) = \sum_i a_i u_i(x)$ using the completeness property. The coefficient a_i can be obtained by taking the inner product of $\psi(x)$ with the eigenfunction $u_i(x)$ viz

$$u_i^*(x)\psi(x) = \sum_{i'} a_{i'} u_i^*(x) u_{i'}(x) = \delta_{ii'}a_{i'} = a_i \tag{4.155}$$

so the LHS of Eq. (4.155) is the linear algebra equivalent of the RHS of Eq. (4.154). Think of fourier transforms.

Let's now construct a *test* for whether a representation is an irrep or reducible rep. Starting with Eq. (4.152), and take the absolute square, then summing over all α :

$$\begin{aligned}\sum_\alpha |\chi_\alpha|^2 &= \sum_\alpha \chi_\alpha^* \chi_\alpha \\ &= \sum_\alpha \sum_{m,m'} a_m a_{m'} \chi_\alpha^{m*} \chi_\alpha^{m'} \\ &= \sum_{m,m'} a_m a_{m'} |G| \delta_{mm'} \\ &= \sum_m a_m^2 |G| \end{aligned} \tag{4.156}$$

where in the 2nd line we have used Eq. (4.150) again. Note that $a_m = a_m^*$ since it is simply a non-complex integer. This result gives us a direct result between the coefficients a_m, and the total trace of the (possibly) reducible representation:

- $\mathcal{D}$ is an *irreducible representation*: Recall that the a_m's are the coefficients that tell us *how many m-th representation* are there in $\mathcal{D}$, hence since $\mathcal{D}$ is one (and only one) of the irrep, say the m_0-th irrep, then $a_{m_0} = 1$ and every $a_m = 0$. Then Eq. (4.156) becomes the equality

$$\boxed{\sum_\alpha |\chi_\alpha|^2 = |G|}. \tag{4.157}$$

- $\mathcal{D}$ is a *reducible representation*: From the argument above, there must exist either more than one non-zero a_m's, or that at least one of the $a_m > 1$, or both. Hence Eq. (4.156) is the *inequality*

$$\boxed{\sum_\alpha |\chi_\alpha|^2 > |G|}. \tag{4.158}$$

Taken together, these two results give us a *test* for whether a representation is reducible or now as long as you know $|G|$: calculate the sum on the LHS of Eq. (4.157), and if it is equal to $|G|$ then it is a irreducible representation, and if it is more than $|G|$ it is reducible. (If you get less than $|G|$, you have made a mistake.)

Example: Consider the representations of D_4. Let's start with 1-dimensional representations – recall that all one dimensional representations are irreducible. It turns out that there are four of them, which we will lable $i = 1, 2, 3, 4$

$$D^i : D_4 = \{e, R, R^2, R^3, m_1, m_2, m_3, m_4\} \to \begin{cases} \{1,1,1,1,1,1,1,1\} \\ \{1,1,1,1,-1,-1,-1,-1\} \\ \{1,-1,1,-1,1,1,-1,-1\} \\ \{1,-1,1,-1,-1,-1,1,1\} \end{cases} \tag{4.159}$$

Notice that the first representation is just all ones, i.e. $\text{Ker}(D^1) = D_4$. Such a representation is called the **Trivial Representation**. You can easily prove to yourself that *all groups possess a trivial representation.* It is clear that all the 1-D representations are *unfaithful*.

The trace of a 1×1 representation is simply the number itself i.e. $\chi_\alpha = D_\alpha$. You can now apply Eq. (4.157) to check that the 1-D representations are truly irreducible. For D^1, all $\chi_\alpha = 1$, so

$$\sum_\alpha |\chi_\alpha|^2 = 8 = |G| \ , \tag{4.160}$$

as claimed for all the 1-D reps

Now, let's consider 2×2 representations of D_4. We have seen *a* representation in the introduction Chapter 1 – we can check whether or not it is an irrep or not easily by applying Eq. (4.157). The characters are

$$\chi(e) = 2 \ , \ \chi(R^2) = -2 \ , \tag{4.161}$$

and everything else zeroes. Hence Eq. (4.157) then gives

$$\sum_\alpha |\chi_\alpha|^2 = 4 + 4 = 8 = |G| \tag{4.162}$$

which proves that it is an irrep.

Finally consider the following *unfaithful* representation of D_4

$$D(e) = D(R) = D(R^2) = D(R^3) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \ , \ D(m_1) = D(m_2) = D(m_3) = D(m_4) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{4.163}$$

which you can prove to yourself is a representation of D_4 by showing that the matrices obey the group composition laws (what is the kernel of this representation?). Now the characters are

$$\chi(e) = 2 \ , \ \chi(R) = 2 \ , \ \chi(R^2) = 2 \ , \ \chi(m_1) = 0 \ , \ \chi(m_2) = 0. \tag{4.164}$$

so applying Eq. (4.157) we get

$$\sum_\alpha N_\alpha |\chi_\alpha|^2 = (1 \times 4) + (2 \times 0) + (2 \times 0) + (1 \times 4) + (2 \times 4) = 16 > |G| \tag{4.165}$$

which proves that it is not an irrep, although the fact that it is already in block-diagonal form probably gives the game away – can you see which are the two 1-D representation that this is built out of?

Let's collect the characters of the irreps into a table. Calling the 2×2 irrep D^5, we have

	e	R	R^3	R^2	m_1	m_3	m_2	m_4
$\chi(D^1)$	1	1	1	1	1	1	1	1
$\chi(D^2)$	1	1	1	1	-1	-1	-1	-1
$\chi(D^3)$	1	-1	-1	1	1	1	-1	-1
$\chi(D^4)$	1	-1	-1	1	-1	-1	1	1
$\chi(D^5)$	2	0	0	-2	0	0	0	0

(4.166)

Such a table is called a **Character Table**. It doesn't really convey more information than those that one already obtained from applying all the theorems, but it does provide a good pedagogical device for us to see some patterns. Notice that, except for the Trivial representation D^1, *the characters sum to zero.* This is actually a consequence of the Orthogonality Theorem for different irreps Eq. (4.143)

$$\sum_\alpha (\chi_\alpha^m)^* (\chi_\alpha^{m'}) = |G| \delta_{mm'} \tag{4.167}$$

and now choosing $m = 1$ for the trivial irrep, so $\chi_\alpha^1 = 1$, and then for any other m', we have $\delta_{1m'} = 0$ for $m' \neq 1$, so

$$\sum_\alpha (\chi_\alpha^{m'}) = 0 \tag{4.168}$$

and hence we have shown that the *sum of all characters in any non-trivial irrep is zero* – yet another way for us to check for irreducibility.

Here for D_4, there are 5 conjugacy classes as we have shown, and we have found 5 irreps. Is there another irrep lurking away? In fact we can show from our results that there isn't any, recall Eq. (4.111)

$$\sum_{m=1}^{\mathcal{M}} n_m^2 \leq |G| \tag{4.169}$$

and we have already found 4 1-D and 1 2-D representations, so $1+1+1+1+(2)^2 = 8 \leq |G|$, so there is no more irreps. Of course, we asserted that the $\leq$ is really an equality – D_4 is evidence of it. We will come back to proving this in the next section.

There is another piece of information shown in the Character Table that it not manifest and which is apparently a coincident: *the number of conjugacy classes is the same as the number of irreps.* Let's look at the Character Table for $D_3 = \{e, R, R^2, m_1, m_2, m_3\}$, with three conjugacy classes $\mathcal{C}_1 = \{e\}$,$\mathcal{C}_2 = \{R, R^2\}$,$\mathcal{C}_3 = \{m_1, m_2, m_3\}$:

	e	R	R^2	m_1	m_2	m_3
$\chi(D^1)$	1	1	1	1	1	1
$\chi(D^2)$	1	1	1	-1	-1	-1
$\chi(D^3)$	2	-1	-1	0	0	0

(4.170)

where D^1 is the trivial representation, D^3 is the 2×2 representation we considered in the example in Section 4.2.5 – there we asserted that it is an irrep but now you can actually prove it by using Eq. (4.157) which we will leave to you to do. $D^2 = \{1, 1, 1, -1, -1, -1\}$ is a 1-D unfaithful irrep which is new to us. Since using Eq. (4.111), we have $1+1+4 = 6 = |G|$ there is no other irreps. Notice again that there are 3 conjugacy classes and 3 irreps. We will prove this theorem next.

4.2.9 Burnside's Theorem

We now end this long chapter with a final theorem. This is actually a couple of theorems, which we have asserted but didn't prove. The two theorems are collectively sometimes know as **Burnside Theorem**.

(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 $\mathcal{M}$ number of such irreps, then the following equality holds*

$$\boxed{\sum_{m=1}^{\mathcal{M}} n_m^2 = |G|} \tag{4.171}$$

- *(Part 2): The number of inequivalent irreducible representations is equal to the number of conjugacy classes in the group.*

Proof: We will prove part Part 1 first. Recall the construction of the *regular representation*, see section 4.2.1. Since it is constructed via a *complete rearrangement*, a regular representation has two main salient points (which you should spend a moment to convince yourself)

- The matrix representation is $|G|$-dimensional, i.e. it is made out of $|G| \times |G|$ square matrices.
- The only element that has no-zero trace is the identity, which is $\chi(e) = |G|$, and $\chi(g \neq e) = 0$.

One can easily check that the regular representation is *reducible* by applying Eq. (4.156) – the only non-zero matrix representation is the identity, which is also its own conjugacy class, so

$$\chi(e)^2 = |G|^2 > |G| \tag{4.172}$$

as long as $|G| > 1$ (i.e. it is not the trivial group). What are the building blocks of the regular representation, i.e. what are the irreps of which it is built out of? We can use the techniques of Decomposition 4.2.8 to find out in the following way. We first decompose the characters of the representation using Eq. (4.152)

$$\chi_\alpha = \sum_m a_m \chi_\alpha^m \tag{4.173}$$

where a_m are given by Eq. (4.154). Now set $\alpha = e$, i.e. we consider the representation for the identity. But now for every irrep labeled by m, the identity representation is simply the identity matrix of $n_m \times n_m$ dimensions, so $\chi_m(e) = n_m$. For the regular rep, $\chi(e) = |G|$, while it's clear that the characters for all other elements are zero $\chi_\alpha = 0$ for $\alpha \neq e$. Hence Eq. (4.154) becomes

$$\begin{aligned} a_m &= \frac{1}{|G|}\sum_\alpha \chi_\alpha^{m*}\chi_\alpha \\ &= \frac{1}{|G|} \times n_m \times |G| \end{aligned}$$

or

$$a_m = n_m. \tag{4.174}$$

Since a_m is the number of time the m-th irrep appear in the regular representation, this means that each m-th irrep appear n_m times. Now remember that the size of the regular representation is an $|G| \times |G|$ matrix, and this is constructed out of a_m number of $n_m \times n_m$ square matrices (after some suitable similarity transforms), this means that

$$\sum_m a_m n_m = |G| \tag{4.175}$$

or

$$\sum_{m=1}^{\mathcal{M}} n_m^2 = |G| \tag{4.176}$$

which proves Part 1 of Burnside's Theorem □.

(**You don't have to know the proof for Part 2**.)

Part 2 of Burnside's theorem is a more complicated and longer proof, so we will sketch this instead of going full monty – you are encouraged to fill in the steps – it has a very pretty ending though. We begin with the conclusion of the Part 1, which is that the Regular representation is built out of n_m number of inequivalent D^m representation, suitably similarly transformed. Let's call the regular representation $\mathcal{D}^{reg}$, so it is

$$\mathcal{D}^{reg}(g) = \sum_\oplus D^m(g). \tag{4.177}$$

Now let $g_1, g_2 \in G$. Via closure, $g_1 g_2^{-1}$ must be some element in G, so this becomes

$$\mathcal{D}^{reg}(g_1 g_2^{-1}) = \sum_\oplus D^m(g_1 g_2^{-1}). \tag{4.178}$$

Taking the trace of Eq. (4.178), we get on the LHS

$$\mathrm{Tr}\mathcal{D}^{reg}(g_1 g_2^{-1}) = \begin{cases} |G| \text{ if } g_1 = g_2 \\ 0 \text{ otherwise} \end{cases} \tag{4.179}$$

since we know that the only non-zero character of the regular representation is the identity, so the only way $g_1 g_2^{-1} = e$ is when $g_1 = g_2$, so we can write it as $\mathrm{Tr}\mathcal{D}^{reg}(g_1 g_2^{-1}) = |G|\delta_{g_1,g_2}$. Tracing over the RHS we get after some algebra

$$|G|\delta_{g_1,g_2} = \sum_m^{\mathcal{M}} n_m D^m_{ij}(g_1)(D^m_{ji}(g_2^{-1})) = \sum_m^{\mathcal{M}} n_m D^m_{ij}(g_1)(D^m_{ij}(g_2))^* \tag{4.180}$$

where we have used unitarity $D^m_{ij}(g_2^{-1}) = (D^m_{ji}(g_2))^*$.

Now here is some trickery: let us consider $g_1 \to g g_1 g^{-1}$ and sum over all $g \in G$ in Eq. (4.180),

$$|G|\sum_g \delta_{g g_1 g^{-1},g_2} = \sum_m^{\mathcal{M}}\sum_g n_m D^m_{ij}(g g_1 g^{-1})(D^m_{ij}(g_2))^* \tag{4.181}$$

and using $D^m(gg_1g^{-1}) = D^m(g)D^m(g_1)D^m(g^{-1})$, and after some index juggling and using the Orthogonal Theorem for different irreps Eq. (4.108), we get

$$|G|\sum_g \delta_{gg_1g^{-1},g_2} = |G|\sum_m \delta_{ij}\delta_{kl}D^m_{kl}(g_1)(D^m_{ij}(g_2))^* \tag{4.182}$$

but now using the deltas to contract the matrices, we get characters

$$\sum_g \delta_{gg_1g^{-1},g_2} = \sum_m^{\mathcal{M}} \chi^m(g_1)\chi^m(g_2)^*. \tag{4.183}$$

Now we fix g_1 and g_2, and let g runs over all the elements of the group G. We have two cases depending on the conjugacy class the element gg_1g^{-1} belongs to. Let the conjugacy class $\mathcal{C}_2$ be the class of which g_2 belongs to and $\mathcal{C}_1$ be the class g_1 belongs to, then we have two cases:

- If $gg_1g^{-1} \notin C_2$, then it is clear that gg_1g^{-1} cannot be g_2 so we get zero.
- If $gg_1g^{-1} \in C_2$, thus $\mathcal{C}_1 = \mathcal{C}_2$, then *some* terms are non-zero, because by conjugation some terms get sent from g_1 to g_2. Suppose $\tilde{g}$ and $\tilde{g}'$ are two such elements (which may be the same element), i.e. $g_2 = \tilde{g}g_1\tilde{g}^{-1}$ and $g_2 = \tilde{g}'g_1\tilde{g}'^{-1}$, then we can write $g_2 = hg_1h^{-1}$, where $h = \tilde{g}^{-1}\tilde{g}'$. Turns out that the set of all h, $H_{g_1} = \{h\}$ *is a subgroup of* G, which you will be asked to prove in the homework. This means that the set of $\{\tilde{g}, \tilde{g}', \tilde{g}'', \dots\}$ (which is *not* a subgroup of G!) has the same number of elements as H_{g_1}, or $|H_{g_1}|$. Since g_2 can be any element of $\mathcal{C}_2 = \mathcal{C}_1$, which means it can also be $g_1 \in \mathcal{C}_1$. Since the set $\{\tilde{g}, \tilde{g}', \tilde{g}'', \dots\}$ is, via $h = \tilde{g}^{-1}\tilde{g}'$ is simply the *left coset* of a subgroup H_{g_1}, it *partitions* G (see section 3.2.2) into $|\mathcal{C}_1|$ disjoint subsets each with $|H_{g_1}|$ number of elements, i.e.

$$|G| = |\mathcal{C}_1||H_{g_1}|. \tag{4.184}$$

  hence the LHS of Eq. (4.183) is then

$$\sum_g \delta_{gg_1g^{-1},g_2} = |G||H_{g_1}|\delta_{\mathcal{C}_1,\mathcal{C}_2} = \frac{|G|}{|\mathcal{C}_1|}\delta_{\mathcal{C}_1,\mathcal{C}_2} \tag{4.185}$$

  and we finally get the **Orthogonality Theorem for Classes**

$$\boxed{|G|\delta_{\mathcal{C}_1,\mathcal{C}_2} = |\mathcal{C}_1|\sum_m^{\mathcal{M}} \chi^m(g_1)\chi^m(g_2)^*}. \tag{4.186}$$

Eq. (4.186) says that the RHS is only non-zero for any given two g_1 and g_2, if $g_1 \sim g_2$ i.e. they belong to the same class. You can use this formula to calculate all sorts of stuff. But how does this show that the number of conjugacy classes is equal to $\mathcal{M}$ the number of irreps? Now if we write Eq. (4.186) as

$$|G|\delta_{\mathcal{C}_1,\mathcal{C}_2} = |\mathcal{C}_1|\sum_m^{\mathcal{M}} \chi^m_{\mathcal{C}_1}\chi^{m\,*}_{\mathcal{C}_2} \tag{4.187}$$

We can see this by thinking of $\mathcal{C}_1$, $\mathcal{C}_2$ and m on $\chi^m_{\mathcal{C}}$ as indices for a $N \times \mathcal{M}$ matrix, where N is the number of conjugacy class. But then Eq. (4.187) can be rewritten as

$$\mathbf{I} \propto \chi_{m\mathcal{C}}{\chi_{\mathcal{C}m}}^{-1} \tag{4.188}$$

which we have used unitarity again. But again we know that *only square matrices have inverses*! This means that $\chi_{m\mathcal{C}}$ *must be a square matrix*, and hence the dimension of m must be the same as the dimensions for $\mathcal{C}$. We have thus proven that the number of conjugacy classes must be equal to the number of irreps $\square$.

4.2.10 Summary

We shall summarize the important points on this chapter. The main thrust is to study *Group representations*. We have shown that group elements may be represented by square matrices, and the objects which square matrices naturally act on are column matrices which are also *Complex Vectors*. Then we showed that the matrix representation of group elements are *incredibly restricted* in their form – the order of the group $|G|$, the number of conjugacy classes severely restrict not just the number of irreps, but also the *individual matrix elements* of the matrix representation themselves! If you think about it, the only input we have entered is the group axioms – this is one of the most mind-blowing thing about group theory.

These set of restrictions are encoded in a set of **Orthogonality Theorems** which you should memorize:

- Orthogonality Theorem of matrix elements of group representations:

$$\sum_{\alpha}(D^m_\alpha)^*_{ij}(D^{m'}_\alpha)_{kl} = \frac{|G|}{n_m}\delta_{ik}\delta_{jl}\delta^{mm'}. \tag{4.189}$$

- Orthogonality Theorem of Characters:

$$\sum_{\alpha}(\chi^m_\alpha)^*(\chi^{m'}_\alpha) = |G|\delta_{mm'}. \tag{4.190}$$

- Orthogonality Theorem for Classes:

$$|G|\delta_{\mathcal{C}_1,\mathcal{C}_2} = |\mathcal{C}_1|\sum_{m}^{\mathcal{M}}\chi^m(g_1)\chi^m(g_2)^*. \tag{4.191}$$

- With these theorems, one can derive the following relations:

  Restriction on the values of characters:

$$\sum_{\alpha} N_\alpha|\chi_\alpha|^2 = |G|. \tag{4.192}$$

  Restriction on the dimensionality of irreps:

$$\boxed{\sum_{m=1}^{\mathcal{M}} n_m^2 = |G|} \tag{4.193}$$

  Restriction on the number of irreps.

$$\text{Number of conjugacy classes } \mathcal{M} = \text{Number of inequivalent irreps} \tag{4.194}$$

Chapter 5

Lie Groups and Lie Algebras

We move in space with minimum waste and maximum joy.

Sade, *Smooth Operator*

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 above[1]. 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.

5.1 Continuity, Smoothness and Analyticity

In the introduction to Chapter 3, we alluded to the fact that **continuous** groups needs additional mathematical structure, namely what do we mean for two elements in a set to be "infinitisimally" 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 (see figure 5.1)

$$f(x) = \begin{cases} x+1 \ , \ x < 0 \\ -x+1 \ , \ x \geq 0 \end{cases} \tag{5.1}$$

[1] This might sound trivial, but it flummoxed mathematicians for 2000 years.

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

$$\frac{df}{dx} = \begin{cases} 1 \ , \ x < 0 \\ -1 \ , \ x \geq 0 \end{cases} \tag{5.2}$$

At $x = 0$, the derivative d^2f/dx^2 is "undefined". In other words, while $f(x)$ is differentiable, its derivative

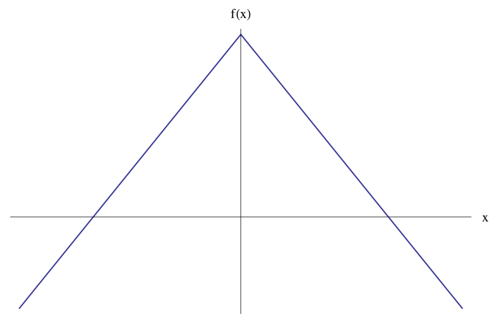


Figure 5.1: A kink is a continuous but non-differentiable (at $x = 0$) function.

$df(x)/dx$ 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 differentiability[2].

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.

$$f(x) = f(x_0) + \left.\frac{df}{dx}\right|_{x_0} (x - x_0) + \frac{1}{2!}\left.\frac{d^2f}{dx^2}\right|_{x_0} (x - x_0)^2 + \dots . \tag{5.3}$$

This series may or may not converge[3], 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.

5.2 Lie Groups

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, $SO(2)$. As we have seen in Chapter

[2] As we briefly alluded to in the beginning of Chapter 4, continuous sets with the notion of "distance" are called **topological spaces**. Each point on a topological space may be labeled with a **coordinate**, and if functions of this space is differentiable in these coordinates, then this space is called a **differential manifold**. Differential manifolds are crucial in physics for obvious reasons – most physical models of dynamical systems are expressed in terms of *time and space derivatives* of variables, and you will encounter them very often in the future. For example, in General Relativity, the curvature of spacetime **metric** can be described by a tensor field that lives on a differential manifold. Indeed, Lie Groups' natural "home" is also the differential manifold. For this class, we will not go that route since it will take us too far afield (and into the subfield of mathematics known as **Differential Geometry**).

[3] I.e. this means that not all smooth functions are analytic.

4, the 2D matrix representation for $SO(2)$ is given by the 1-parameter matrix Eq. (4.12)

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \ , \ 0 \leq \theta < 2\pi. \tag{5.4}$$

Similarly, the 1D irrep of $SO(2)$ is given by

$$R_\theta = e^{i\theta}. \tag{5.5}$$

Notice that both representations are parameterized by a single continuous parameter.

In general, the **dimension** of any Lie Group is the number of continuous parameters needed to parameterized it. Consider now a dimension 1 Lie Group G with some representation L, i.e.

$$L : G \to GL(N, \mathbb{C}). \tag{5.6}$$

Do not confuse the dimensionality of the matrix N with the dimensionality of the Lie Group n – as we have seen in the case $SO(2)$ above, $n \neq N$ in general. We parameterize the representation L with some real variable $\mathbb{R}$. From closure, then

$$L(c) = L(a)L(b) \ , \ a, b, c \in \mathbb{R}. \tag{5.7}$$

Now since G is a continuous group, then c *must be a continuous function of* a *and* b, i.e.

$$c = f(a, b) \ , \ f \text{ is a continuous function of } a \text{ and } b. \tag{5.8}$$

The function $f(a, b)$ must obey the inherited rules of the parent group.

$$\text{Associativity}: \ L(a)[L(b)L(c)] = [L(a)L(b)]L(c) \Rightarrow f(a, f(b, c)) = f(f(a, b), c) \tag{5.9}$$

$$\text{Identity}: \ \text{Let } a_0 \text{ labels the identity i.e. } L(a_0) = e$$

$$\text{then } L(a)L(a_0) = L(a_0)L(a) = L(a) \Rightarrow f(a, a_0) = f(a_0, a) = a. \tag{5.10}$$

The last equality, $f(a, a_0) = f(a_0, a) = a$ can be derived via $L(c) = L(a)L(b) \Rightarrow L(f(a, b)) = L(a)L(b)$ and substituting $b = a_0$ we get $L(f(a, a_0)) = L(a)L(a_0) = L(a)$ and hence $f(a, a_0) = a$. Finally there is the inverse

$$\text{Inverse}: \ \text{Let } L(b) \text{ be the inverse of } L(a) \text{ i.e. } (L(b))^{-1} = L(a) \text{ then} \tag{5.11}$$

$$L(a)L(b) = L(b)L(a) = L(a_0) \Rightarrow f(b, a) = f(a, b) = a_0. \tag{5.12}$$

So far we have recapped the usual story about continuous groups, with the group laws that we have now *isomorphically* mapped to some domain in $\mathbb{R}$, and expressed by the continuous function $f(a, b)$. Now, we have all that is required to describe Lie Groups.

(Definition) Lie Groups: Let G be a n-the dimensional Lie Group parameterized by some continuous n dimensional space M, $a, b \in M$ be points on the parameter space of G, and $f(a, b)$ be a continuous function which maps two points (a, b) in the Lie Group onto a third point c, then $f(a, b)$ is an *analytic* function in both variables.

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.

Example: Consider again $SO(2)$, and for simplicity let's work with the 1D irrep where the operators are paramterized as

$$R(\theta) = e^{i\theta}. \tag{5.13}$$

The group composition law is

$$R(\theta_3) = R(\theta_1)R(\theta_2) = R(\theta_1 + \theta_2) \tag{5.14}$$

which means that the function f for this group is given by the simple addition

$$f(\theta_1, \theta_2) = \theta_1 + \theta_2 \bmod 2\pi \tag{5.15}$$

This function is clearly C_∞ and analytic for both variables, which we can see by expanding in a Taylor series

$$\begin{aligned} f(\theta_1, \theta_2) &= f(\bar{\theta}_1, \bar{\theta}_2) + \left.\frac{\partial f}{\partial \theta_1}\right|_{\bar{\theta}_1} (\theta_1 - \bar{\theta}_1) + \left.\frac{\partial f}{\partial \theta_2}\right|_{\bar{\theta}_2} (\theta_2 - \bar{\theta}_2) + \ldots \\ &= \bar{\theta}_1 + \bar{\theta}_2 + (\theta_1 - \bar{\theta}_1) + (\theta_2 - \bar{\theta}_2) \\ &= \theta_1 + \theta_2, \end{aligned} \tag{5.16}$$

i.e. within the domain $0 \leq \theta < 2\pi$, the Taylor expansion converges to the functional value f.

5.3 Invariant transformations on Real Space

You have learned that one way to understand group operations is to leave some object invariant. For example the dihedral-4 group D_4 leaves the square invariant. We can abstract this class of objects from geometrical things to the more abstract notion of "equations". For example, consider the differential equation for $y(x)$

$$x\frac{dy}{dx} + y - xy^2 = 0. \tag{5.17}$$

If one perform the following transformation, for any non-zero $a \neq 0 \in \mathbb{R}$

$$x \to ax \ , \ y \to a^{-1}y \tag{5.18}$$

then

$$\frac{dy}{dx} \to a^{-2}\frac{dy}{dx}. \tag{5.19}$$

Substituting these back into Eq. (5.17) we recover the original equation – we say that the tranformations leave Eq. (5.18) invariant, hence it is a *symmetry* of the equation. Note that we have insisted that $a \neq 0$ else the transformation $y \to a^{-1}y$ is undefined – you will soon see that this requirement is intimately tied with the existence of inverses. Operationally this means that if $f(x, y) = 0$ is a solution of the differential equation Eq. (5.17), then the transformed equation $f(ax, a^{-1}y) = 0$ is also a solution.

The transformations Eq. (5.18) form a group – we can check that they obey the usual group laws as follows:

- Closure : Let $a, b \in \mathbb{R}$ and $a, b \neq 0$, and consider the two successive transformations

$$x' = ax \quad , \ y' = a^{-1}y \tag{5.20}$$

$$x'' = bx' \quad , \ y'' = b^{-1}y' \tag{5.21}$$

  and then it follows that $x'' = abx$ and $y'' = (ab)^{-1}y$. But let $c = ab$ and it's clear that $c \in \mathbb{R} \neq 0$, and hence we have demonstrated closure.

- Associativity : Follows from the associativity of multiplication of real numbers.
- Identity : It is trivial to see that 1 is the identity.
- Inverse : Again it is clear that the inverse for any transformation labeled a is its geometric inverse a^{-1}.

Hence we have proven that the set of transformations Eq. (5.18) is indeed a group[4]. In fact, it is a Lie group – it is the Lie group $\mathbb{R}^* = \mathbb{R} - \{0\}$, i.e. the group of reals (modulo the origin) under multiplication. To see that, we first construct the representation in a more familiar manner in the following way. First note that the "action" of the group is to transform the coordinates $(x, y) \to (ax, a^{-1}y)$, so we can write a 2D reducible representation as

$$T(a) = \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} \tag{5.22}$$

which act on the vector (x, y) such that

$$T(a) \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax \\ a^{-1}y \end{pmatrix}. \tag{5.23}$$

The group law as we have shown above is multiplicative, i.e.

$$T(c) = T(a)T(b) = T(ab) \tag{5.24}$$

so in function form $f(a, b) = ab$. This function is clearly analytic, and hence we have shown that this is a Lie Group.

The above example is something you will encounter very often in physics – a set of transformations leaving the original equation invariant. If we think of (x, y) above as points on the Cartesian space $\mathbb{R}^2$, then the transformation of the form $x' = ax$ is called **dilatation**, a "stretch" or a **scale transformation**. Expressions which are invariant under such a transformation are called **scale invariant**. For example, the simple equation

$$F(x, y) = \frac{x}{y} \tag{5.25}$$

is scale invariant under the transformation $x' = ax$ and $y' = ay$ for $a \neq 0 \in \mathbb{R}$.

Example: Consider the following 2 parameter transformation on a 1D **target space** $x \in \mathbb{R}$,

$$x' = a_1x + a_2 \ , \ a_1, a_2 \in \mathbb{R} \ , \ a_1 \neq 0 \ , \ x \in \mathbb{R} \tag{5.26}$$

and a second successive transformation yields

$$x'' = b_1x' + b_2 = b_1a_1x + (b_1a_2 + b_2) \equiv c_1x + c_2. \tag{5.27}$$

This transformation is non-abelian, as one can easily check by reversing the order of the transformation via

$$x' = b_1x + b_2 \ , \ x'' = a_1x' + a_2 \tag{5.28}$$

and

$$x'' = a_1b_1x + (a_1b_2 + a_2) \equiv c'_1x + c'_2, \tag{5.29}$$

which clearly $c_2 \neq c'_2$ in general. The identity is $a_1 = 1, a_2 = 0$. We can find the inverse by requiring that $b_1a_1 = 1$ and $b_1a_2 + b_2 = a_1b_2 + a_2 = 0$, so the inverse of (a_1, a_2) is $(a_1^{-1}, a_1^{-1}a_2)$. Notice that it is the requirement of the existence of an inverse which imposes our initial condition that $a_1 \neq 0$ (but *not* $a_2 \neq 0$). Can you construct a function $F(x)$ which is left invariant by this transformation?

[4]Can you show that it is also an Abelian group?

To show that this transformation is analytic (and hence forms a Lie Group), we note that the transformations Eq. (5.26) acts on the target space x

$$x' = f(x, a_1, a_2) = a_1 x + a_2, \tag{5.30}$$

and is analytic in both a_1 and a_2, i.e. it is clear that we can Taylor expand around both a_1 and a_2.

The above example is a type of transformations which you see often, which is that the Lie Group operators map elements of a **target space** T to other elements in the same target space T, i.e.

$$L : T \to T. \tag{5.31}$$

In general, consider an n-parameter Lie Group labeled by $(a_1, a_2, \ldots, a_n)$ acting on an N-th dimensional target space labeled by coordinates $\{x_i\}$ for $i = 1, 2 \ldots N$. A general transformation is given by the set of equations

$$\begin{aligned} x'_1 &= f_1(x_1, x_2, \ldots, a_1, a_2, \ldots) \\ x'_2 &= f_2(x_1, x_2, \ldots, a_1, a_2, \ldots) \\ &\vdots \\ x'_i &= f_i(x_1, x_2, \ldots, a_1, a_2, \ldots) \end{aligned} \tag{5.32}$$

and *as long as all the f_i's are analytic in the a_j's Lie parameters the transformation is analytic.*

As an example, recall the transformation at the start of this section Eq. (5.18) which is a single parameter Lie Group acting on a 2D target space $\{x, y\}$, it is clear that both $x'(a, x, y) = ax$ and $y'(a, x, y) = a^{-1}y$ is analytic in $a \in \mathbb{R}^*$.

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

$$T(a_1, a_2, a_3) = \begin{pmatrix} a_1 & a_2 \\ a_3 & \frac{1+a_2a_3}{a_1} \end{pmatrix} \in SL(2, \mathbb{R}) \ , \ a_i \in \mathbb{R} \tag{5.33}$$

where it is clear that the parameterization is not analytic at $a_1 \to 0$. 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 parameter 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.

5.4 Linear/Matrix Groups as Lie Groups

In Chapter 4, we discuss Matrix Groups as prototypes of continuous groups. We now reintroduce them in this section – it turns out that matrix groups that we have studied are actually also Lie Groups[5]. Recall

[5] Not all Lie Groups are matrix groups, and not all matrix groups are Lie Groups. But almost all the Lie Groups that you will encounter as physicists are isomorphic to some matrix groups.

that the element of an N-th dimension matrix group is an $N \times N$ matrix, and the natural "object" for this operator to act on is an $N \times 1$ column matrix. Consider an $N = 3$ matrix operator M

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \tag{5.34}$$

and it is clear that what the matrix operator does to the *target space* with coordinates $\{x, y, z\}$ is to execute a **linear transformation**

$$x' = a_{11}x + a_{12}y + a_{13}z \tag{5.35}$$

$$y' = a_{21}x + a_{22}y + a_{23}z \tag{5.36}$$

$$z' = a_{31}x + a_{32}y + a_{33}z. \tag{5.37}$$

If the group is the General Linear group over the field of reals $GL(3, \mathbb{R})$, then the only condition we require on the a_{ij}'s is that the determinant must not be zero $\det(M) \neq 0$ so there is no **algebraic constraint** on the values of a_{ij}. Then it is clear that the transformations Eq. (5.35) is analytic and the dimensionality of the Lie Group $GL(3, \mathbb{R})$ is 9.

However, if the group is a more constrained group, like $SL(3, \mathbb{R})$, i.e. where we impose conditions on the determinant and group operators, then we are not free to choose the values of a_{ij}. Indeed, the determinant of $M \in SL(3, \mathbb{R})$ is given by

$$a_{11}a_{22}a_{33} + a_{13}a_{21}a_{32} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32} = 1 \tag{5.38}$$

which represents an algebraic constraint on the possible values of a_{ij}. This means that the dimensionality of the Group is $9 - 1 = 8$, since once we have chosen 8 different values for a_{ij} the ninth is completely determined by Eq. (5.38). For groups with more constraints, such as $SO(3, \mathbb{R})$, which in addition to possess unit determinant, must also obey $R^T R = e$, the parameter space for a_{ij} is even smaller. Recall our discussion of $SO(2)$ – take a look at section 4.1.3 again – we showed there explicitly that the unit determinant constraint and the orthogonality condition implies that such a strong condition that the initial 4 parameters $a_{11}, a_{12}, a_{21}, a_{22}$ are reduced to a single compact parameter $0 \leq \theta < 2\pi$.

5.4.1 Representations of Lie Groups

Now since matrix groups are also Lie Groups, what about their representations? Well, since they are already in matrices, *they are already in their representations!*. For example, the $N \times N$ matrices of $SO(N)$ are the representations of the group. These matrices, where the matrix dimensions are the same as the N of the group, is called the **fundamental representation**, which is an irrep (although we won't prove it). But, as we have seen in the previous chapter, we can also represent the matrix groups using other matrices. For example, we have seen $SO(2)$ can be represented by its fundamental representation

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \quad 0 \leq \theta < 2\pi. \tag{5.39}$$

or a 1 dimensional representation $U_\theta = e^{i\theta}$ (this is the fundamental representation of $U(1)$ which is isoomorphic to $SO(2)$). You can also use direct sums to construct reducible representations of $SO(2)$, e.g.

$$H(\theta) = R_\theta \oplus U_\theta = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & e^{i\theta} \end{pmatrix}. \tag{5.40}$$

Notationally, sometimes we use bold numbers to denote the matrix dimensions of the representations (as opposed to the parameter dimensions), i.e. Eq. (5.40) is sometimes written as $\mathbf{3} = \mathbf{2} \oplus \mathbf{1}$. In this case, the fundamental rep is an irrep, and $\mathbf{1}$ is an irrep, while $\mathbf{3}$ is a reducible rep.

In the previous chapter 4, we spend a lot of time discussing the representations of finite groups, and many of the theorems require the order $|G|$. Obviously, for Lie Groups, $|G|$ is infinite, so those theorems are not directly applicable. On the other hand, Lie Groups are parameterized by a finite number of real parameters i.e. it's group dimensions. As we will see later, the dimensions corresponds to the number of generators for Lie Groups, and we can use these generators construct their representations in section 5.5.

5.4.2 Isometries : Metric Preserving Operations

What is a **metric**? If you are already well advanced in the study of differential geometry or general relativity, you will think of the metric as some $N \times N$ dimensional matrix which determines the *length* between two infinitisimal points. So, roughly speaking, the "metric preserving operations" means "length preserving operations". Such transformations are sometimes called **isometries**. How we define "length" however, depends on the underlying physics. Let us consider the simplest of cases, which is length in usual 3 dimensional **Euclidean Space** $\mathbb{R}^3$ – this is exactly what you think it is, i.e. length L between origin and a point (x, y, z) is given by $L^2 = x^2 + y^2 + z^2$. A linear transformation that transforms $(x, y, z) \rightarrow (x', y', z')$ is called **Metric Preserving** if

$$x^2 + y^2 + z^2 = x'^2 + y'^2 + z'^2. \tag{5.41}$$

In other words, the symmetry of this group operation is to *leave the length invariant.* For simplicity, let us consider metric preserving operations on a 2D Euclidean space $\mathbb{R}^2$. For a matrix R, then

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \tag{5.42}$$

and hence after the tranformation we get

$$x'^2 + y'^2 = (a_{11}^2 + a_{21}^2)x^2 + 2(a_{11}a_{12} + a_{21}a_{22})xy + (a_{12}^2 + a_{22}^2)y^2. \tag{5.43}$$

Imposing the metric preserving condition $x^2 + y^2 = x'^2 + y'^2$ we get the conditions

$$a_{11}^2 + a_{21}^2 = 1 \ , \ a_{12}^2 + a_{22}^2 = 1 \ , \ a_{11}a_{12} + a_{21}a_{22} = 0, \tag{5.44}$$

which are three algebraic constraints on the parameters, which means that the group is really $4 - 3 = 1$ dimensional. Now, multiplying the first two equations of Eq. (5.44), and then subtracting the square of the 3rd equation, we get

$$\begin{aligned} (a_{11}^2 + a_{21}^2)(a_{12}^2 + a_{22}^2) - (a_{11}a_{12} + a_{21}a_{22})^2 &= 1 \\ \text{or} \quad (a_{11}a_{22} - a_{12}a_{21})^2 &= 1. \end{aligned} \tag{5.45}$$

One can recognize that Eq. (5.45) is just the square of the determinant of R

$$(\det R)^2 = (a_{11}a_{22} - a_{12}a_{21})^2 = 1, \tag{5.46}$$

which means that $\det R = \pm 1$. If we now choose $\det R = +1$, then this is the so-called **Special** condition we have discussed in section 4.1.1. In fact, the derivation you have is exactly what we have done when we talked about $SO(2)$ in that section. If we do not impose $\det R = +1$, then this group is known as $O(2)$.

Again from section 4.1.1, we have learned that the *orthogonality condition* is $R^T R = e$, and you can check for yourself that this leads to the derivation above. In fact, this is a general condition for *all*

compact *metric preserving groups over the real field* $\mathbb{R}$. The proof is just a real valued version of the fact that unitary matrix operators perserve the inner product of two complex vectors which we have already proven in section 4.2.6.

Theorem: *Let R be a real valued $N \times N$ square matrix with non-zero determinant. If R obey the* **Orthogonality Condition**, *$R^T R = e$, then it preserves the inner product of two real vectors.*

Proof Let U, V be two real N dimensional vectors, then $V' = RV$ and $U' = RU$, and recalling that the inner product of two real vectors U and V is $U \cdot V = U^T V$, we have

$$U' \cdot V' = U'^T V' = (RU)^T (RV) = U^T R^T R V \tag{5.47}$$

then for $U' \cdot V' = U \cdot V$, we require

$$R^T R = e. \quad \square \tag{5.48}$$

Using this theorem, and setting $V = U$ and $U = (x_1, x_2, x_3, \ldots, x_N)$, then a matrix satisfying the orthogonality condition will transform such that it preserves its "metric" (length)

$$U \cdot U = x_1^2 + x_2^2 + x_3^2 + \cdots + x_N^2 = {x'_1}^2 + {x'_2}^2 + {x'_3}^2 + \cdots + {x'_N}^2 = U' \cdot U'. \tag{5.49}$$

The eerie similarity to the unitary operators studied in section 4.2.6 is of course not a coincidence.

(Definition): Orthogonal groups $O(N)$ are **compact** metric preserving Lie groups over the real fields, while Unitary groups $U(N)$ are **compact** metric preserving Lie groups over the complex fields. If the determinants are furthermore unitary, they are *special* ($SO(N)$ and $SU(N)$ respectively).

The *special* condition restricts the group to simple rotations. If we do not impose specialness, then the group also include the parity symmetry operations – it is clear that if we set $x \to -x$, $y \to -y$, etc and not necessarily all of them at once, the metric is still preserved. This also means that $SO(N)$ are subgroups of $O(N)$, and similarly for Unitary groups.

The additional moniker **compact** means that the Lie Group parameters occupy a compact space, for example for $SO(2)$, the parameter $0 \leq \theta < 2\pi$ is clearly compact. We can classify the groups in the following way

$$\begin{array}{lll} \text{Real Field} & R^T e R = e & O(N) \\ \text{Complex Field} & U^\dagger e U = e & U(N) \\ \text{Quartenion Field} & Q^\dagger e Q = e & Sp(N) \end{array} \tag{5.50}$$

where we have sneaked in the extra identity e into the equation with a view of generalizing this set-up even further later. We have also added an extra Field, called **quartenions**, whose metric preserving Lie Group is called the **Symplectic Group** $Sp(N)$. You have encountered quartenions before, when you worked on the quartenion group Q_8 in your homework set, which is an order 8 group. We will not discuss the details of quartenions or the Symplectic group in this class[6].

In addition to compact metric preserving Lie Groups, there are **non-compact metric preserving Lie Groups** as well. Consider the following metric

$$s^2 = -(ct)^2 + x^2 + y^2 + z^2 \tag{5.51}$$

which you may recognize as the *invariant spacetime length* that you have encountered when you studied special relativity. Let us set $c = 1$ as all high energy physicists do, then we want to find a group whose linear transformations leave the metric Eq. (5.51) invariant, i.e.

$$-t^2 + x^2 + y^2 + z^2 = -t'^2 + x'^2 + y'^2 + z'^2. \tag{5.52}$$

[6]Quartenions are generalizations of complex numbers in the following way. Let a, b be real numbers, then we can construct a complex number by $z = ai + b$ with $i^2 = -1$. Given two complex number y, z, a quartenion can be constructed by writing $q = yj + z$, here $j^2 = -1$, and $ij + ji = 0$.

It turns out that this is not hard to do at all. Consider the Orthogonal group $O(4)$ with the condition $R^T eR = e$; all we need to do is to replace

$$e = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix} \rightarrow I_{3,1} \equiv \begin{pmatrix} -1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix} \tag{5.53}$$

such that non-compact metric preserving group satisfies

$$R^T I_{p,q} R = I_{p,q} \tag{5.54}$$

and we are done. You will be asked to prove this is true in a homework problem. We have introduced the notation $I_{p,q}$ above to denote p number of $+1$ and q number of -1, arranged in a diagonal. The dimensions of the matrix is $N = p + q$, i.e. they are represented by $(p + q) \times (p + q)$ square matrices. Non-compact Orthogonal (Unitary and Symplectic) groups are also called Orthogonal (Unitary and Symplectic) groups, but with the name $O(p, q)$ ($U(p, q)$ and $Sp(p, q)$). Hence, the group which preserves the metric Eq. (5.51) is $O(3, 1)$. If we further impose the condition that the determinant is $+1$, then it is $SO(3, 1)$, also known as the 4D **Lorentz Group**[7] – the underlying symmetry by which Einstein's Special Relativity can be understood.

5.5 Lie Algebras and Exponentiation

In Finite group theory, one of the crucial property of a finite group is that its *order*, i.e. the number of elements in the group, is finite. With that property, many theorems and properties can be studied and proven. In continuous group however, the "order" of the group is infinite, so in general it is harder to study it. However, Lie Groups are continuous groups with a powerful trump card: their analyticity.

Much like the fact that an analytic function allows us to study its *local* properties around a small infinitisimal region, the analyticity of Lie Groups allow us to study the group in some small local region (in parameter space). As you will soon see, this is a significant simplification, as sometimes knowledge of a small local region is sufficient to describe the whole group. This is, of course, not general – there are some groups whose *global* properties cannot be extracted from studying their local properties. Nevertheless, we will get far.

5.5.1 Infinitisimal Generators of Lie Groups

Recall that in finite groups, often we can generate the entire group using only a smaller subset of the group elements which we call generators, by the *repeated application of the generators.* In Lie Groups, the generators are a bit more subtle to define – we can indeed find some (often non-unique) set of generators which allow us to *sometimes* generate the entire group by repeated application ("integration"). If such generators exist, then knowledge of the generators is as good as knowing the whole group, and as we will see in this section these generators are intimately related to the *local* properties of the group.

Let us begin by considering our favorite example, $SO(2)$ in its reducible 2×2 representation from section 4.1.3

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} , \quad 0 \leq \theta < 2\pi. \tag{5.55}$$

[7]Sometimes, it is called the Homogenous Lorentz Group, where the Inhomogenous Lorentz Group (or more popularly the Poincaré Group) includes translations.

Since this is a Lie Group, analyticity allows us to Taylor expand[8] around $\theta = 0$

$$R(\theta) = R(0) + \left.\frac{dR}{d\theta}\right|_{\theta=0} \theta + \frac{1}{2!}\left.\frac{d^2R}{d\theta^2}\right|_{\theta=0} \theta^2 + \dots. \tag{5.57}$$

We now want to calculate $dR(\theta)/d\theta|_{\theta=0}$. To do that, we use the usual composition law

$$R(\theta_1 + \theta_2) = R(\theta_1)R(\theta_2) \tag{5.58}$$

and now taking derivatives of both side with respective to θ_1 and then setting $\theta_1 = 0$

$$\left.\frac{dR(\theta_1+\theta_2)}{d\theta_1}\right|_{\theta_1=0} = \left.\frac{dR(\theta_1)}{d\theta_1}\right|_{\theta_1=0} R(\theta_2). \tag{5.59}$$

The LHS of this equation is, via the chain rule,

$$\begin{aligned} \left.\frac{dR(\theta_1+\theta_2)}{d(\theta_1+\theta_2)}\frac{d(\theta_1+\theta_2)}{d\theta_1}\right|_{\theta_1=0} &= \left.\begin{pmatrix} -\sin(\theta_1+\theta_2) & -\cos(\theta_1+\theta_2) \\ \cos(\theta_1+\theta_2) & -\sin(\theta_1+\theta_2) \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right|_{\theta_1=0} \\ &= \begin{pmatrix} -\sin(\theta_2) & -\cos(\theta_2) \\ \cos(\theta_2) & -\sin(\theta_2) \end{pmatrix} \\ &= \frac{dR(\theta_2)}{d\theta_2} \end{aligned} \tag{5.60}$$

which must equal the RHS of Eq. (5.59)

$$\begin{aligned} \frac{dR(\theta_2)}{d\theta_2} &= \left.\frac{dR(\theta_1)}{d\theta_1}\right|_{\theta_1=0} R(\theta_2) \\ &= \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} \cos(\theta_2) & -\sin(\theta_2) \\ \sin(\theta_2) & \cos(\theta_2) \end{pmatrix} \\ &= X\begin{pmatrix} \cos(\theta_2) & -\sin(\theta_2) \\ \sin(\theta_2) & \cos(\theta_2) \end{pmatrix} \end{aligned} \tag{5.61}$$

where we have defined

$$X \equiv \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{5.62}$$

This means that, dropping the subscript 2, we have

$$\frac{dR(\theta)}{d\theta} = XR(\theta) \tag{5.63}$$

and hence using $R(\theta) = \mathbb{I}$

$$\left.\frac{dR(\theta)}{d\theta}\right|_{\theta=0} = X. \tag{5.64}$$

We will now show that X *generates the entire group*, and hence is a **generator**, using Eq. (5.57). Using very similar calculations from above, we can show that

$$\left.\frac{d^2R(\theta)}{d\theta^2}\right|_{\theta=0} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = X^2. \tag{5.65}$$

[8]The derivative of a matrix is

$$\frac{d}{d\theta}\begin{pmatrix} a_{11}(\theta) & a_{12}(\theta) \\ a_{21}(\theta) & a_{22}(\theta) \end{pmatrix} = \begin{pmatrix} da_{11}/d\theta & da_{12}/d\theta \\ da_{21}/d\theta & da_{22}/d\theta \end{pmatrix}, \tag{5.56}$$

which we are allowed to do because the operator $d/d\theta$ is a linear operator.

Either via bruteforce, or via induction, you can show that

$$\left.\frac{d^m R(\theta)}{d\theta^m}\right|_{\theta=0} = X^m. \tag{5.66}$$

Plugging in Eq. (5.66) into the Taylor expansion Eq. (5.57), we get

$$\begin{aligned} R(\theta) &= R(0) + \left.\frac{dR}{d\theta}\right|_{\theta=0}\theta + \frac{1}{2!}\left.\frac{d^2R}{d\theta^2}\right|_{\theta=0}\theta^2 + \ldots. \\ &= \mathbb{I} + X\theta + \frac{1}{2!}X^2\theta^2 + \ldots \\ &= \sum_{m=1}^{\infty}\frac{1}{m!}(X\theta)^m \\ &= e^{X\theta}. \end{aligned} \tag{5.67}$$

Where in the last line, we have treated the matrices X,X^2 etc as *functions.* This is probably something new that you may not have seen before. Indeed, one can take the last line as a *definition for the exponentiation action of a square matrix*, i.e.

$$\boxed{e^{A\theta} \equiv \sum_{m=1}^{\infty}\frac{1}{m!}(A\theta)^m \qquad \text{for any square matrix } A}. \tag{5.68}$$

From Eq. (5.67), we can see that given X, we can generate the entire group $R(\theta)$ with a single generator X via the **exponentiation** (or **Exponent Map**)

$$\boxed{R(\theta) = e^{X\theta}}. \tag{5.69}$$

While it may sound weird for now, we say that X *is an element of the Lie Algebra* $\mathfrak{so}(2)$ *of the Lie Group* $SO(2)$, with the weird fonts $\mathfrak{so}(2)$ to differentiate it from capital $SO(2)$. Why it forms an *algebra* – a set of elements with some binary rule which closes – is not clear since with only a single generator $\mathfrak{so}(2)$ is rather trivial.

Notice the structure of equation 5.69 : on the RHS, we only need to know X, i.e. it is a term in an ifinite expansion of the Taylor expansion. But it seems like *knowledge of this local point allows us to generate the entire group!* In other words, the exponent map allows us to generate a large chunk of the group elements using only a small amount of knowledge. This is the power of analyticity.

Now, you might protest that we have cheated – after all, it seems like we need to know the actual rotation operator R_θ before we can compute what X is. Certainly, in our exposition in these lectures, we have taken that route. However, as it turns out, we have done things the opposite way round – for example, for $SO(2)$ at least, knowledge of X is as good as knowledge of R_θ as they are related by Eq. (5.69). Indeed, Lie groups are almost (not always) defined by knowledge of their generators. To see this, we will consider a more complicated group, $SO(3)$.

Recall that $SO(3)$ is the group of all rotations in 3 dimensions, so let us consider the 3×3 fundamental representation of $SO(3)$. Since there are 3 axes (x_1, x_2, x_3), we can parameterize the rotations around these three *fixed* axes using the angles $(\theta_1, \theta_2, \theta_3)$. Let arrange the target space as a vector

$$P = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}. \tag{5.70}$$

Consider rotations around the x_3 axis. Since this keeps x_3 fixed, we can write down the representation as

$$R(\theta_3) = \begin{pmatrix} \cos\theta_3 & -\sin\theta_3 & 0 \\ \sin\theta_3 & \cos\theta_3 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{5.71}$$

We can carry out similar calculations as we have done above for $SO(2)$ to find the generator X_3

$$X_3 \equiv \frac{dR(\theta_3)}{d\theta_3}|_{\theta_3=0} = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{5.72}$$

It should be obvious now how we can derive the generators X_1 and X_2 for rotations around the x_1 and x_2 axes, which are

$$R(\theta_1) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_1 & -\sin\theta_1 \\ 0 & \sin\theta_1 & \cos\theta_1 \end{pmatrix} \ , \ X_1 \equiv \frac{dR(\theta_1)}{d\theta_1}|_{\theta_1=0} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}. \tag{5.73}$$

$$R(\theta_2) = \begin{pmatrix} \cos\theta_2 & 0 & \sin\theta_2 \\ 0 & 1 & 0 \\ -\sin\theta_2 & 0 & \cos\theta_2 \end{pmatrix} \ , \ X_2 \equiv \frac{dR(\theta_2)}{d\theta_2}|_{\theta_2=0} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}. \tag{5.74}$$

Now, we perform the following redefinition $X_i \equiv -iL_i$ to make contact with more familiar notation, i.e.

$$\begin{aligned} L_1 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} \\ L_2 &= \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix} \\ L_3 &= \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \end{aligned} \tag{5.75}$$

These operators obey the following **commutator relation**

$$[L_i, L_j] = L_iL_j - L_jL_i = i\epsilon_{ijk}L_k \tag{5.76}$$

where ϵ_{ijk} is the **Levi-Civita symbol** which has the relation $\epsilon_{123} = \epsilon_{312} = \epsilon_{231} = 1$, $\epsilon_{132} = \epsilon_{321} = \epsilon_{213} = -1$ and everything else zero[9]. You have seen these relations before – these are the **angular momentum operators** in your quantum mechanics class, modulo the Planck constant $\hbar$. Notice that we have derived the relations without even talking about quantum mechanics – the non-commutative nature of the rotation operators comes from the fact that *rotations in 3D do not commute even classically*. You can check this fact yourself with a ball.

Now, the operators $\{L_1, L_2, L_3\}$ form a set and the commutator relation Eq. (5.76) *provides a binary relation for this set*, i.e. it tells us how to put two operators from the set to get a third operator which also belongs to the set. In other words, the set of generators for $SO(3)$, $\{L_1, L_2, L_3\}$ is *closed* under the binary relation Eq. (5.76) and hence (recall our definition of an algebra way back in section 2.3.1), this forms an algebra. We call this the **Lie Algebra** of $SO(3)$, and symbolically $\mathfrak{so}(3)$ with the "$\mathfrak{fraktur}$" fonts.

[9]We have used the summation convention over k, i.e. you should read Eq. (5.76) as

$$[L_i, L_j] = i\sum_{k=1}^{k=3} \epsilon_{ijk}L_k. \tag{5.77}$$

Similar to $SO(2)$, if we are given the Lie Algebra $\mathfrak{so}(3)$, we can generate the entire group $SO(3)$ via the exponentiation

$$\boxed{R(\theta_1, \theta_2, \theta_3) = e^{i(\theta_1 L_1 + \theta_2 L_2 + \theta_3 L_3)}} \tag{5.78}$$

where $\theta_1, \theta_2, \theta_3$ defines the *rotation axis* $\vec{\theta} = \theta_1 \hat{x}_1 + \theta_2 \hat{x}_2 + \theta_3 \hat{x}_3$.

(Remark) : We say that "the generators Eq. (5.75) generate the representations of $SO(3)$ in the parameterization given by Eq. (5.73), Eq. (5.74), Eq. (5.71)". In general, *different matrices – or different parameterizations of the Lie Algebras – will generate different representations of* $SO(3)$.

Finally, it should be obvious that *the number of generators must equal the number of dimensions of the Lie Group, since there will be a generator for each parameter.*

5.5.2 Lie Algebras, Structure Constants and the Adjoint Representation

In the previous section 5.5.1 we have derived the Lie Algebras $\mathfrak{so}(2)$ and $\mathfrak{so}(3)$, by *linearizing* around some point in the parameter space of the Lie Group. In particular, for $SO(3)$, we found that its Lie Algebra is the set of generators $\mathfrak{so}(3) = \{L_1, L_2, L_3\}$ with the commutator relation

$$[L_i, L_j] = i\epsilon_{ijk} L_k. \tag{5.79}$$

which serves as the binary relation for the algebra. In general, it is always possible to find the closure relation in terms of commutators for any Lie Algebra, with the *structure* of the algebra determined by something else other than ϵ_{ijk}.

(Definition) Stucture Constants: For any given Lie Group L, we can find its (finite) n-dimensional set of generators $l = \{l_1, l_2, \ldots, l_n\}$ which will close under the commutator relation

$$\boxed{[l_i, l_j] = c_{ij}^k l_k} \tag{5.80}$$

where c_{ij}^k can be read either a 3-tensor, or (more prosaically) a set of $n \times n$ matrices c_{ij} labeled by k. These are called the **Structure Constants** of the Lie Algebra. *The structure constants are invariant under the change of representation*[10].

The commutator relation (or Lie Bracket) $[A, B] \equiv AB - BA$ must obey certain relations. One of them is the ubiquitous **Jacobi Identity**.

(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*

$$\boxed{[[A, B], C] + [[B, C], A] + [[C, A], B] = 0}. \tag{5.81}$$

You will be asked to prove the Identity in a homework problem.

Furthermore, the commutator relation (sometimes known as the **Lie Bracket**) is the binary operator of the Lie Algebra. What about the elements of the algebra? Clearly the set of generators $\{l_i\}$ associated with each parameter of the Lie Group belongs to the Algebra. But then, as we said, linear combinations of $\{l_i\}$, e.g. $a_1 l_1 + a_2 l_2 + a_3 l_3$ is also a generator – exponentiating $\exp(a_1 l_1 + a_2 l_2 + a_3 l_3) \in G$ generates an element of the group as we mentioned before. Thus such linear combinations of group generators also belongs to the Algebra. Indeed, the set $\{l_i\}$ can be thought of as forming the **basis vectors** for a vector space of generators. As we have studied in Section 4.2.2, vector spaces is a structure which are closed under two binary operators : scalar multiplication and linear addition.

[10] Hence "constants", doh

What happens when we plug these non-basis generators into the commutator relationship? Suppose $A = al_i + bl_j$ for $a, b \in \mathbb{R}$,

$$\begin{aligned} [A, l_k] &= [al_i + bl_j, l_k] \\ &\equiv a[l_i, l_k] + b[l_j, l_k] \\ &= aC_{ik}^m l_m + b_{jk}^m l_m \in \mathfrak{g} \ . \end{aligned} \tag{5.82}$$

where we have *defined* the operator from the 1st line to the 2nd line, i.e. $[al_i + bl_j, l_k] \equiv a[l_i, l_k] + b[l_j, l_k]$. Such a property is called **bilinearity**. The last line is true since the structure constants are just numbers.

To summarise:

Properties of Lie Algebras:

- The Lie Algebra $\mathfrak{g}$ is a set of elements which is closed under the binary operator given by the commutator relation $[X, Y] \in \mathfrak{g}\ \forall\ X, Y \in \mathfrak{g}$.
- The commutator $[X, Y]$ is defined by its structure constants, anti-commutes $[X, Y] = -[Y, X]$ and obeys bilinearity and the Jacobi Identity.
- *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

$$aX + bY \in \mathfrak{g}. \tag{5.83}$$

Notice in the above, we have made *no* mention of the Lie Group. This is a rather subtle point – abstractly Lie Algebra can exists on its own without its "parent" Lie Group. In fact, as we will soon see, different Lie Groups can have the same Lie Algebras. Nevertheless, when related to the Lie Group, the two following properties also holds:

- Each element of the algebra $X \in \mathfrak{g}$ generates an element of the group G via exponentiations, i.e. $\exp X = g \in G$.
- The dimensions of the Lie Group is the dimensions of the Lie Algebra.

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.

(** No need to know the Adjoint Representation**.)

**** (Definition) Adjoint Representation**: The set of matrices defined by the structure constants $(C_i)_j^k \equiv -c_{ij}^k$ of a Lie Algebra forms a representation of the parent Lie Group[11].

Let's parse the above definition as follows. We substitute Eq. (5.80) into the Jacobi Identity Eq. (5.81), using $A = l_i$,$B = l_j$ and $C = l_k$, some algebra leads us to the relation

$$c_{ij}^m c_{mk}^n + c_{jk}^m c_{mi}^n + c_{ki}^m c_{mj}^n = 0. \tag{5.84}$$

Now if we consider c_{ij}^k as a set of n $n \times n$ matrices C_i, defined as

$$(C_i)_j^k \equiv -c_{ij}^k \tag{5.85}$$

[11]This representation is faithful if and only if the Lie Group is simply connected and possesses no non-trivial **Center**, where the Center of a group is defined to be the set of elements which commute with every other element in the group. Just to confuse you even further, a group is called **centerless** if the center consists only of the identity e.

where the upper index labels the columns and the lower index labels the rows of the matrix, we can rewrite Eq. (5.84) as

$$(c_{ij}^m C_m)_k^n + (C_j C_i)_k^n - (C_i C_j)_k^n = 0. \tag{5.86}$$

(Note that a single small c_{ij}^m survives the substitutions.) Dropping the subscripts k and superscript n, and replacing the index $m \to k$, we get

$$C_i C_j - C_j C_i = [C_i, C_j] = c_{ij}^k C_k \tag{5.87}$$

which is just the binary operation Eq. (5.80) which defines the algebra! In other words, *the structure constants themselves form a Lie Algebra of the group.*

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

$$\begin{aligned} C_1 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} = L_1 \\ C_2 &= \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix} = L_2 \\ C_3 &= \begin{pmatrix} 0 & i & 0 \\ -i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = L_3. \end{aligned} \tag{5.88}$$

In other words, the generators $\{L_1, L_2, L_3\}$ form the structure constants themselves. However, this is purely a coincident in our case – in general there are *many different paramterizations of the Lie Algebra* $\mathfrak{g}$ *which do not have the same matrices as the structure constants* – as we will soon see in chapter **??**.

5.6 Differential representations of Lie Algebras

We have focussed on matrix representations of Lie Algebras so far. But they don't always appear as matrices – as long as they obey the algebraic relation Eq. (5.80), then they form algebras.

Consider an infinitisimal counter clockwise rotation of $\Delta\theta$ around the z-axis, see figure 5.2. The

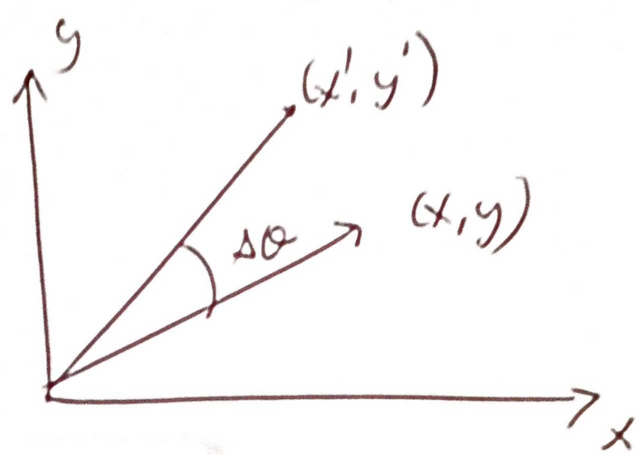


Figure 5.2: Infinitesimial counter clockwise rotation.

transformations are

$$x' = x\cos(\Delta\theta) - y\sin(\Delta\theta)\ ,\ y' = x\sin(\Delta\theta) + y\cos(\Delta\theta)\ ,\ z' = z \tag{5.89}$$

where in the limit $\Delta \to 0$, we can Taylor expand around $\Delta\theta = 0$, (moving to infinitisimal notation $d\theta$),

$$x' = x - yd\theta + \dots\ ,\ y' = xd\theta + y + \dots\ ,\ z' = z \tag{5.90}$$

Now, consider some analytic function $F(x, y, z)$. Under the transform Eq. (5.90), we can again Taylor expand

$$F(x,y,z) \longrightarrow F(x',y',z') = F(x - yd\theta, xd\theta + y, z) = F(x,y,z) + \underbrace{\left(-y\frac{\partial}{\partial x} + x\frac{\partial}{\partial y}\right)}_{I_3} Fd\theta + \dots \tag{5.91}$$

where we have defined the **differential operator**

$$I_3 \equiv \left(-y\frac{\partial}{\partial x} + x\frac{\partial}{\partial y}\right). \tag{5.92}$$

We can do the same calculation for infinitisimal rotation around the x and y axes, to get

$$\begin{aligned} I_1 &\equiv \left(-z\frac{\partial}{\partial y} + y\frac{\partial}{\partial z}\right) \\ I_2 &\equiv \left(-x\frac{\partial}{\partial z} + z\frac{\partial}{\partial x}\right). \end{aligned} \tag{5.93}$$

Defining $L_i \equiv -i\hbar I_i$, we obtain the familiar **Angular Momentum Operators** of quantum mechanics

$$\boxed{L_i = -i\hbar\epsilon_{ijk}x_j\frac{\partial}{\partial x^k}}. \tag{5.94}$$

You can easily check that Eq. (5.94) forms a Lie Algebra of $SO(3)$ by computing the commutator of two angular momentum operators

$$[L_i, L_j] = i\hbar\epsilon_{ijk}L_k. \tag{5.95}$$

5.7 Conclusions and some final Remarks

We have taken rather unsystematic tour of Lie Algebras and Lie Groups, and have left some loose ends hanging which we will now make a few remarks on. Unfortunately, we will not have time to go into the details of the following facts, but you should aware of them. Much of the subtlety stems from the fact that Lie Algebras captures only the *local* behavior of a group. There is a large amount of literature on Lie Groups and Lie Algebras – the diligent student is encouraged to check out the references suggested in the front page!

- *Is there a unique Lie Algebra for every Lie Group?* **Yes.** However, there may not be a Lie Group for every Lie Algebra – this is one reason why mathematicians like to distinguish between Lie Algebras and Lie Groups. They are intimately connected, yet are not the same thing!

- *Are Lie Groups with isomorphic Lie Algebras isomorphic?* **No.** Different Lie Groups may share the same Lie Algebra. For example, the group $SU(2)$ and $SO(3)$ share the same Lie Algebra – both are 3 dimensional and obey the commutator $[L_i, L_j] = i\epsilon_{ijk}L_k$, i.e. they have the same structure constants and hence $\mathfrak{su}(2) \cong \mathfrak{so}(3)$. However $SU(2)$ is not isomorphic[12] to $SO(3)$.

[12] The fact that they share the same Lie Algebra means that they are related in some way. In fact, there exists a 2 to 1 onto homeomorphic map from $SU(2)$ to $SO(3)$, i.e. "$SU(2)$ is twice a big as $SO(3)$". We say that "$SU(2)$ **covers** $SO(3)$ twice". In general, when two algebras are isomorphic but their parent groups are not, it is due to the fact that one of the group covers the other multiple times. The smallest group of this set (and hence is always the one being "covered") is called the **Universal Cover Group**. Here $SO(3)$ is the universal cover for $SU(2)$.

- *Can we always generate the entire group with some set of generators?* **Yes.** If we blindly consider the exponentiating maps of some Lie Algebra $l_i \in \mathfrak{l}, L(\alpha_i) = e^{\alpha_i l_i}$, these in general do *not* generates the entire group. In particular, if the group is *non-compact* then this exponentiating will fail. However, we can generate the entire group if we split the map into a product of exponentiating maps of compact and non-compact generators.

- *Is the map from a Lie Algebra onto a Lie Group unique?* **No.** In other words, while the Lie Algebra for each Lie Group is unique (i.e. they have the same structure constants), there are different parameterizations for the same Lie Algebra. However, they are related to each by the **Baker-Hausdorf-Campbell** formula.

- *Do we know all the Lie Groups and Lie Algebras in existence?* **Yes.** Like Finite Groups, remarkably mathematicians led by Killing and Cartan have managed to classify all the inequivalent Lie Groups in existence. In analogy to Finite groups, which are based upon the classification of Simple groups which cannot be factored (i.e. which have no proper normal subgroup), Lie Groups have a special class of fundamental groups called Simple Lie Groups which has no non-trivial normal subgroup and which one can use to build other non-simple Lie Groups. Killing and Cartan have classified all of the simple Lie Groups.

Appendix A

Noether's Theorem

A.1 Euler-Lagrange Equation

The Lagrangian L of a dynamical system with n number of coordinates labeled by q_i, for $i = 1, 2, 3, \ldots, n$ is in general a functional of the canonical coordinate q_i and its time derivatives $L[q_i, \dot{q}_i, t]$. For every coordinate q_i there is an associated **canonical momentum** P_i given by

$$P_i \equiv \frac{\partial L}{\partial \dot{q}_i} \quad . \tag{A.1}$$

The **action** S of the dynamical system is given by an integral over all possible paths, i.e.

$$S = \int_{t_0}^{t_1} dt \; L[q_i, \dot{q}_i, t] \; , \tag{A.2}$$

The **Action Principle** states that the motion of the dynamical system is path $(q_i(t), \dot{q}_i(t))$ which extremizes the action S.

This path is a solution of the Euler-Lagrange Equations

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right) - \frac{\partial L}{\partial q_i} = 0 \; . \; . \tag{A.3}$$

A.2 Theorem

Emmy Noether proved the following theorem :

For every continuous symmetry in the Lagrangian, there is a conservation law

We begin with a general Lagrangian with a set of coordinates q_i and its associated canonical momenta $P_i = \partial L/\partial \dot{q}$. Let's now shift each of the coordinates q_i by a small amount,

$$q_i \rightarrow q_i' + \delta q_i \tag{A.4}$$

where the shift

$$\delta q_i \equiv f_i(\mathbf{q})\epsilon \; , \tag{A.5}$$

may depend on the other coordinates through the arbitrary functions f_i (i.e. there is a f_i function for each variable q_i). This is a *continuous* transformation – you can make ϵ as small as you want until the transformation is infinitisimally small. The change in the Lagrangian, δL due to these transformation can be calculated easily using the chain rule, and it is (we have dropped the primes from q_i for simplicity)

$$\delta L = \sum_i \left(\frac{\partial L}{\partial \dot{q}_i}\delta \dot{q}_i + \frac{\partial L}{\partial q_i}\delta q_i\right) \quad . \tag{A.6}$$

Now the first term of Eq. (A.6), using our definition of the canonical momentum Eq. (A.1), we get

$$\frac{\partial L}{\partial \dot{q}_i}\delta\dot{q}_i = P_i\delta\dot{q}_i \ . \tag{A.7}$$

For the 2nd term of Eq. (A.6), we use the Euler-Lagrange equation Eq. (A.3) to write $\partial L/\partial q_i = dP_i/dt$ to get

$$\frac{\partial L}{\partial q_i}\delta q_i = \dot{P}_i\delta q_i \ . \tag{A.8}$$

Inserting Eq. (A.7) and Eq. (A.8) back into Eq. (A.6), we then have

$$\begin{aligned}\delta L &= \sum_i P_i\delta\dot{q}_i + \dot{P}_i\delta q_i \\ &= \frac{d}{dt}\sum_i P_i\delta q_i \ .\end{aligned} \tag{A.9}$$

Using Eq. (A.5), we can then write this as

$$\frac{d}{dt}\sum_i P_i f_i(\mathbf{q}) = \delta L \ , \tag{A.10}$$

where we have canceled the ϵ as it is a constant. Now Noether told us that if the Lagrangian is invariant under the transformations, then $\delta L = 0$, and hence we get

$$\frac{d}{dt}\sum_i P_i f_i(\mathbf{q}) = 0 \ , \tag{A.11}$$

but this equation is simply $d/dt(\text{something}) = 0$, which we already learned that this means that the "something" is a conserved quantity! In other words, if the Lagrangian is invariant under the transformations Eq. (A.4), then there is a conservation law that says that the quantity

$$Q = \sum_i P_i f_i(\mathbf{q}) \ , \tag{A.12}$$

is always constant in time!

Example : Consider the motion of a *free particle* with mass m and in position q in 1D, traveling under the potential,

$$L = \frac{1}{2}m\dot{q}^2 \ . \tag{A.13}$$

There is a **translational symmetry** associated with this Lagrangian, i.e. $q \to q+c$ where c is a constant. It's easy to show that the $L \to L$. Then $f_i = (1,1,1)$ and the Noether charge Eq. (A.12) is

$$Q = P \ , \tag{A.14}$$

i.e. the momentum $P = m\dot{q}$ is conserved.

Example : Consider the motion of a particle with mass m with position (r,θ) in polar coordinates, traveling under the potential $V(r)$ that is independent of θ. The Lagrangian is

$$L = \frac{1}{2}m\left(\dot{r}^2 + r^2\dot{\theta}^2\right) - V(r) \ . \tag{A.15}$$

There is a rotational symmetry $\theta \to \theta + c$ and $r \to r$ where c is any constant. Thus, $f_r = 0$ and $f_\theta = 1$, with the associated noether charge

$$Q = P_\theta = mr^2\dot{\theta} \ , \tag{A.16}$$

which is just the angular momentum. Hence, rotational symmetry is associated with conservation of angular momentum. Notice that we have done the problem using polar coordinates. You can do the same problem using cartesian coordinates instead, with the Lagrangian

$$L = \frac{1}{2}m\left[\dot{x}^2 + \dot{y}^2\right] - V(\sqrt{x^2+y^2}) \ . \tag{A.17}$$

Can you find the transformation on x and y that leave this Lagrangian invariant?

A.3 Conservation of Energy

Finally, we will discuss the conservation of energy – what is the symmetry associated with it? As it turns out, the conservation of energy is associated with the **time translation symmetry**. To be precise, the Lagrangian must be invariant under *explicit* time translation symmetry $t \rightarrow t + f(\mathbf{q})\epsilon$. This needs a bit of care to explain. There are two ways the Lagrangian can depend on time – **implicit** and **explicit**. The variables in the Lagrangian are all functions of time – so as time changes, the variables $q_i(t)$ and $\dot{q}_i(t)$ also changes – in this case we say that the Lagrangian is *implicitly* depending on time through the functions $q_i(t)$ and $\dot{q}_i(t)$. In our two examples above, the Lagrangians are implicit functions of time.

On the other hand, the Lagrangian can also depend *explicitly* on time, if the variable t explicitly appear in the Lagrangian. For example, the problem of a coupled pendulum with the Lagrangian

$$L = \frac{1}{2}m\dot{x}_1^2 + \frac{1}{2}m\dot{x}_2^2 - \frac{mg}{l}(x_1^2 - x_2^2) + \frac{1}{2}k(x_1 - x_2)^2 \ , \tag{A.18}$$

where k is the spring constant. If k does not depend on time, then L is not explicitly dependent on time. However, if we heat up the spring during the experiment, and due to the expansion of the spring, k changes as a function of time, i.e. $k \rightarrow k(t)$. Then its Lagrangian Eq. (A.18) would become

$$L = KE - PE = \frac{1}{2}m\dot{x}_1^2 + \frac{1}{2}m\dot{x}_2^2 - \frac{mg}{l}(x_1^2 - x_2^2) + \frac{1}{2}k(t)(x_1 - x_2)^2 \ . \tag{A.19}$$

where k is now a function of time. This means that the Lagrangian now *explicitly* depends on time through $k(t)$ in addition to being *implicitly* dependent on time through its other variables. Noether's theorem now state that the energy of the system is conserved if the Lagrangian is invariant under an *explicit time translation*. Let's see how this works.

In general, the Lagrangian can now depend on its variables and also the explicit time variable t, so we write the functional as

$$L[\mathbf{q}, \dot{\mathbf{q}}, t] \ . \tag{A.20}$$

Taking the *total derivative* of Eq. (A.20), we get

$$\frac{dL}{dt} = \sum_i \left[\frac{\partial L}{\partial q_i}\dot{q}_i + \frac{\partial L}{\partial \dot{q}_i}\ddot{q}_i\right] + \frac{\partial L}{\partial t} \ , \tag{A.21}$$

noting that the last term is a *partial derivative* on L with respect to t *which will vanish unless the Lagrangian has an explicit dependence on time.* Using the Euler-Lagrange Eq. (**??**), we can rewrite each term in the sum as

$$\frac{\partial L}{\partial q_i}\dot{q}_i + \frac{\partial L}{\partial \dot{q}_i}\ddot{q}_i = \dot{P}_i\dot{q}_i + P_i\ddot{q}_i = \frac{d}{dt}\left(P_i\dot{q}_i\right) \ , \tag{A.22}$$

which we can insert back into Eq. (A.21) to get

$$\frac{dL}{dt} = \frac{d}{dt}\sum_i \left(P_i\dot{q}_i\right) + \frac{\partial L}{\partial t} \ . \tag{A.23}$$

Now, we define a quantity, called the **Hamiltonian** H as

$$H \equiv \sum_i \left(P_i\dot{q}_i\right) - L \ , \tag{A.24}$$

then Eq. (A.23) becomes

$$\frac{dH}{dt} = -\frac{\partial L}{\partial t} \ . \tag{A.25}$$

This Eq. (A.25) tells us that, if L has no explicit dependence on time, then $dH/dt = 0$, and hence the Hamiltonian is a conserved quantity. The Hamiltonian is actually an exact and mathematical precise definition of "energy" in a dynamical system, and thus Noether's theorem says that if the Lagrangian is invariant under time translations – i.e. it has no explicit dependence on time – then the energy of the system is conserved.

To check again, consider the Lagrangian of a particle m traveling in the potential $V(q)$, $L = (1/2)m\dot{q}^2 - V(q)$, its Hamiltonian is $H = P\dot{q} - L = (1/2)m\dot{q}^2 + V$ which says that the total energy of the system is given by the sum of the kinetic energy and the potential energy of the particle as we expected.

This symmetry principle is an extremely powerful tool – every conservation law that we know off has an underlying symmetry associated with it. Beyond momentum and energy, conservation of electric charges, neutrons, protons and other more esoteric things like quarks or lepton number. Indeed, just like the how physicists now think about the dynamics of physics in terms of the action principle, they now think of the content of physics in terms of symmetry principles.