PHYSICS THROUGH SYMMETRIES

S G Rajeev

World Scientific

PHYSICS
THROUGH
SYMMETRIES

This page intentionally left blank

PHYSICS THROUGH SYMMETRIES

S G Rajeev
University of Rochester, USA

World Scientific
NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI • TOKYO

Published by

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

PHYSICS THROUGH SYMMETRIES

ISBN 978-981-98-1288-2 (hardcover)
ISBN 978-981-98-1289-9 (ebook for institutions)
ISBN 978-981-98-1290-5 (ebook for individuals)

For any available supplementary material, please visit
https://www.worldscientific.com/worldscibooks/10.1142/14305#t=suppl

Desk Editor: Rhaimie B Wahap

Typeset by Stallion Press
Email: enquiries@stallionpress.com

Printed in Singapore

This book is dedicated to my family: Lynn, Alec and Leela.

This page intentionally left blank

PREFACE

For over a hundred years, group theory has played an essential role in theoretical physics. Even before that it was known that the solution of many mechanical systems (harmonic oscillator, Kepler problem) are possible because of symmetries. E. Noether showed that symmetries imply conservation laws in mechanics. Groups were also useful in classifying periodic structures (crystals) and their X-ray diffraction patterns.

Groups and their representation found its deepest uses in Quantum Theory. Wigner and Weyl were the pioneers. Groups served as a powerful tool to classify and understand particles and their states in nuclear and high energy physics. More fundamentally, interactions among elementary particles (electromagnetic and strong forces) turn out to be a consequence of non-abelian gauge symmetries. This idea has its roots in General Relativity, where Einstein found that general covariance leads to gravitational interactions.

So you cannot understand modern theoretical physics without knowing some group theory. The field is too broad for one person to know everything about every group. Choices must be made.

My choice of topics is admittedly personal. I have tried to emphasize the concepts (Lie theory, Peter-Weyl Theorem etc.) instead of techniques (listing of irreducible representations, calculation of Casimir invariants etc.). There are already excellent books that could train you in techniques. To say nothing of various online resources (Wikipedia, Scholarpedia etc.) that give a quick summary. A good understanding can be obtained by studying a few representative cases in depth. Even if you encounter different cases in your own research, the knowledge so obtained can often be adapted.

There are also indications that future physicists will need symmetries that go beyond groups. The most intriguing of these are "quantum groups". So I have

included them. I have tried to avoid symmetries that theorists love, but which have not found much experimental support.

I wish there was time to do more. Among the omissions I regret most are: Co-adjoint orbits, Virasoro and Kac-Moody algebras and Harish-Chandra's theory of unitary representations of non-compact groups.

The book is not meant to be read linearly, cover to cover. Several sections can be omitted in a first reading; they will make more sense when you return to them afterwards. They are denoted by a star in the section heading. The later chapters are more terse and technically involved. They can also be omitted in a first reading.

The more examples and exercises you work out the better your understanding will be. Some of the exercises are solved in detail. But look at the solution only after a serious attempt.

My own knowledge and appreciation of group theory is from my teacher, A. P. Balachandran; augmented by later interactions with Mark Bowick, Feza Gursey and Susumu Okubo. I owe a debt of gratitude also to Rakesh Tibrewala, who read most of the book with great care. His comments were very helpful. Of course, I own entirely any mistakes or confusions that remain.

Thanks also to Christopher B Davis and Rhaimie B Wahap of World Scientific for their encouragement and patience through the slow process of writing the book.

CONTENTS

Chapter 1

SYMMETRY BEFORE PHYSICS

Symmetries play a central role in modern physics. Very few problems in dynamics can be solved exactly; only those with a high degree of symmetry (such as rotation invariance in the Kepler problem.) At a deeper level, symmetries determine the laws of physics: For example, conservation of energy is a consequence of symmetry under time translations and conservation of momentum is due to translation invariance in space. Further, the standard model of elementary particles is determined to a large extent by gauge invariance; Einstein's theory of gravity by general covariance and the Euler equations of a fluid by the Lie algebra of incompressible vector fields. Modern Physics can only be understood through symmetries.

It is useful to step away for a moment to understand how symmetries appeared originally, even before physics

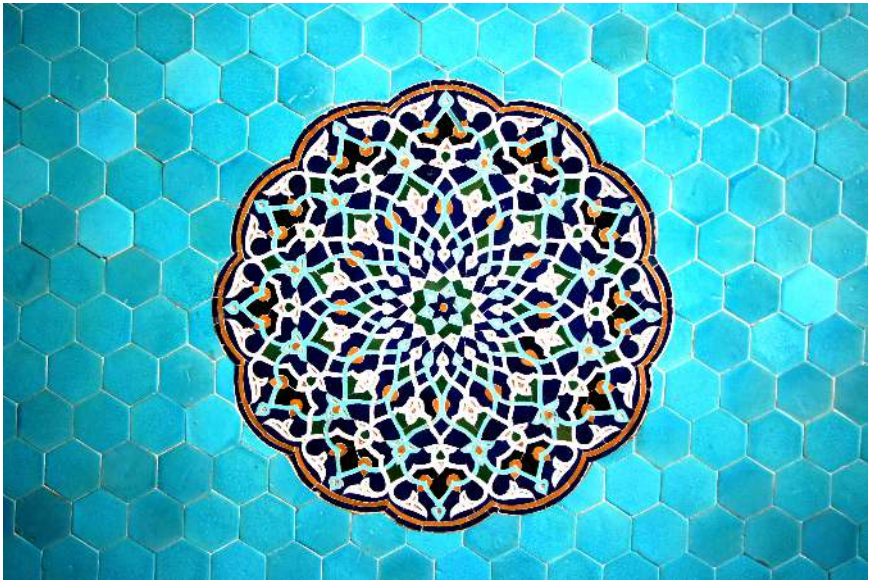
1.1. Symmetries in Art and Architecture

1.1.1. *Cultural references to symmetries can be found long before physics*

Here is an example, from a Buddhist manual on meditation, quoted in [1]:

```
In the glistening surface of each pearl
are reflected all the other pearls
In each reflection, again are reflected
all the infinitely many other pearls,
So that by this process, reflections
of reflections continue without end.
```

Highly symmetric geometric patterns can be found in the mosaics of many temples and mosques, on the Taj Mahal as well as in many Cathedrals of Europe. We seek symmetry everywhere in life, and beyond. There is symmetry even in a cemetery.



1.2. Symmetries in Geometry

Perhaps the first systematic study of symmetry in geometry is by Greek mathematicians; for example, Plato classified regular solids.

Each Platonic solid (regular solid) has a finite set of rotations as a symmetry. The simplest is the tetrahedron, whose symmetry contains the permutation of its four vertices. If you connect the centers of the faces of a Platonic solid, you get

another one, called its dual. The tetrahedron is self-dual; the dual of a cube is an octahedron. The icosahedron and the duodecahedron are dual to each other. Dual solids have the same symmetry. The symmetry group of the icosahedron plays an interesting role in the theory of the quintic equation: It is the group A_5 of even permutations of five elements. There are intriguing connections between algebra and geometry all over this subject.

The vertices of each Platonic solid lie on a sphere; projecting radially from the center, we can associate points on the sphere with the midpoints of the edges and faces. Using the stereographic co-ordinates, we can identify the sphere with the complex projective space $\mathbb{CP}^1$. These devices allow us to make concrete the relation of the geometry of Platonic solids to algebra. To each Platonic solid, there is a rational function $f : \mathbb{CP}^1 \to \mathbb{CP}^1$which vanishes at the vertices, is equal to 1 at the mid-points of the edges and has a pole at the center of the faces. The most sophisticated case, the icosahedron, is worked out in Ref. [2].

1.2.1. *A word of warning*

Like the Sirens who devoured sailors in the Odyssey, beauty can lead us astray. Many physicists and astronomers were led to false theories because of their irresistible beauty.

Kepler originally thought he could explain the orbits of planets as circles centered at the Sun; each planet (there were five known at the time) was associated to a Platonic solid. Such ideas were not unusual among astrologers of that time. Since his model did not agree with Tycho Brahe's data, Kepler persevered and produced another model in which the orbits are ellipses. This discipline in sticking to the less symmetric, but empirically correct, model is what distinguished him from other astrologers; and made Kepler an astronomer. We now know that there are even higher degrees of symmetry underlying mechanics than Kepler could have ever imagined (e.g., symplectic transformations).

Truth has its own fierce beauty which far surpasses that of fancy.

Another example is Kelvin, who thought elements in the periodic table were explained by vortices in ether. The shape of the vortex of hydrogen is a simple knot, that of helium is the trefoil knot and so on. The correct theory (based on Quantum Mechanics and the Coulomb potential) is much weirder and far more beautiful.

The success of the standard model in unifying electromagnetism and weak interactions led physicists to a quest for a Grand Unified Theory in the 1970s. Groups such as $SU(5)$, $SO(10)$ (and the exquisite, irresistible E_8) seemed for a while to be great candidates. They predicted that the proton should decay at a certain rate. Experiments, which did not find these decays, have ruled these theories

out.[1] However it turns out, the right answer will have a fierce beauty of its own, which may not be evident at first.

We resist the temptation to follow the theory of Platonic solids and such further (e.g., Coxeter groups), as they do not lead to currently useful physics. Odysseus listened to the Sirens, but saved himself by being chained to the mast of his ship. He sealed the ears of his crew with wax so they would not be led astray.

1.3. Symmetries in Algebra

1.3.1. *The basic mathematical idea is a group*

We will give a precise axiomatic definition later. For now, a group describes the transformations (symmetries) of some physical or geometrical object. For example an equilateral triangle PQR can be rotated around its center by 120 degrees to get to an equivalent situation, changing only the labelling of the vertices to QRP . If we do it again we get RPQ. A third iteration gets us back to the same triangle, even labelled the same way. The set of these transformations is a group with three elements: The identity 1 (which does nothing), an element $\sigma : PQR \mapsto QRP$ which rotates by 120 degrees, an element σ^2 which rotates by 240 degrees. The condition

$$\sigma^3 = 1$$

expresses the fact that a rotation through 360 degrees is the same as the identity. This group $\{1, \sigma, \sigma^2\}$ is called Z_3, the cyclic group of three elements. This is an example of an abelian group, one in which the product of two elements does not depend on the order of multiplication. Abel was a Norwegian mathematician who, along with Galois, invented the idea of a group.

An equilateral triangle also has a symmetry under reflections along the line orthogonal to a side and passing through the opposite vertex: e.g., $\tau : PQR \mapsto QPR$. If we apply σ followed by τ we end up with $PQR \mapsto QRP \mapsto RQP$; i.e., $\tau\sigma : PQR \mapsto RQP$. If we apply them in the opposite order $PQR \mapsto QPR \mapsto PRQ$ so that $\sigma\tau : PQR \mapsto PRQ$. Thus we see that $\sigma\tau \neq \tau\sigma$; i.e., the group is not abelian.

You can see moreover that all six permutations of the vertices can be obtained by some combination of σ and τ: $PQR, QRP, RPQ, QPR, PRQ, RQP$. In other words, σ and τ are generators of the permutation group S_3 of a set of three objects.

In the next section we will see that the same group S_3 appears in a totally different context: The solution of a cubic equation.

[1] But discovered neutrinos emitted by a supernova, a completely different phenomenon.

Exercise 1. Show that $\sigma\tau = \tau\sigma^{-1}$

1.3.2. *Which polynomial equations can you solve algebraically?*

Everyone knows that the quadratic equation

$$ax^2 + bx + c = 0$$

has the two solutions

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

In the middle ages, it was found that cubic and fourth degree equations can also be solved in a similar way, in terms of cube roots and fourth roots. The formulas are much more complicated though. But no one could find a solution for the general fifth order equation (quintic) in terms of fifth roots, and similarly for higher order equations. But special cases could be solved that way. Galois, continuing ideas of Abel, showed that the general quintic *cannot* be solved this way: There are fifth order equations that cannot be solved even if you could calculate fifth roots.

To understand the connection of symmetries to algebraic equations, let us take a closer look at the solution of the quadratic. Let α, β be the roots of $ax^2 + bx + c = 0$:

$$ax^2 + bx + c = a(x - \alpha)(x - \beta)$$

By comparing the coefficients of the powers of x we get

$$b = -a(\alpha + \beta), \quad c = a\alpha\beta$$

This is the easy part: The hard part is to go in the other direction and determine α, β in terms of a, b, c. Before we do that, notice that the coefficients b, c are symmetric functions of α, β: If we interchange $\alpha \leftrightarrow \beta$ they do not change. Moreover, any symmetric function of the roots can be written in terms of the coefficients. This is the key point. Of particular interest is the discriminant

$$\Delta = (\alpha - \beta)^2$$

This is a function of the roots that vanishes iff the roots coincide. Being symmetric in α, β we can express this in terms of the coefficients. It is not hard to do this explicitly:

$$(\alpha - \beta)^2 = (\alpha + \beta)^2 - 4\alpha\beta = \frac{b^2 - 4ac}{a^2}$$

So,

$$\alpha - \beta = \pm\frac{\sqrt{b^2 - 4ac}}{a}$$

Since we already know $\alpha + \beta = -\frac{b}{a}$, we have reduced the problem of solving a quadratic to solving two linear equations. So, one way to understand the solution of the quadratic is to focus on the symmetries of the equation. We were able to reduce the nonlinear part of the problem to solving for a radical $y^2 = \Delta$.

Let us see if this generalizes to higher orders. The classical idea is to reduce cubics (resp. quartics) to evaluating cube roots (resp. fourth roots) plus some elementary operations of addition, multiplication and division.

1.3.3. *Solving Cubics*

Some cubic equations are easy to solve because they can be reduced to quadratic. For example, $x^3 - 1 = 0$ has an obvious solution 1. Since

$$x^3 - 1 = (x - 1)(x^2 + x + 1)$$

the other two roots are given by the roots of the quadratic $x^2 + x + 1 = 0$. If we define

$$\omega = \frac{-1 + \sqrt{3}i}{2}$$

the three roots of $x^3 - 1 = 0$ are $\omega^0 = 1, \omega, \omega^2$. More generally, any equation of the type

$$x^3 - s = 0$$

has solutions in terms of the cube root of s:

$$x = \sqrt[3]{s}, \omega\sqrt[3]{s}, \omega^2\sqrt[3]{s}$$

What about more general equations?

$$ax^3 + bx^2 + cx + d = 0$$

Clearly $a = 0$ can be excluded, since that would be merely a quadratic. Then we can divide throughout by a and redefine $\frac{b}{a} = -s_1$ etc. to reduce the cubic equation to

$$x^3 - s_1x^2 + s_2x - s_3 = 0.$$

(The signs are chosen to make later formulas simpler.)

If the roots are $\alpha_1, \alpha_2, \alpha_3$ we must have

$$x^3 - s_1x^2 + s_2x - s_3 = (x - \alpha_1)(x - \alpha_2)(x - \alpha_3)$$

Comparing coefficients

$$s_1 = \alpha_1 + \alpha_2 + \alpha_3, \quad s_2 = \alpha_1\alpha_2 + \alpha_2\alpha_3 + \alpha_3\alpha_1, \quad s_3 = \alpha_1\alpha_2\alpha_3.$$

Notice that these are symmetric under permutations of the roots. There are six permutations of three objects labeled by $1, 2, 3$. Of these,

$$123, \quad 231, \quad 312$$

involve an even number of pair-wise interchanges. (0 is an even number; so the identity, which does nothing, is an even permutation.) The remaining are the odd permutations

$$213, \quad 132, \quad 321$$

The composition of two even permutations is again even; so they form a subgroup (i.e., a subset which contains all of the products and inverses of its elements). The composition of two odd permutations is even; so they *do not form a subgroup*. Define

$$\sigma : 123 \mapsto 231, \quad \tau : 123 \mapsto 213$$

Clearly, σ is even and generates the even permutations (i.e., all other even permutations are its powers). τ is odd. Moreover,

$$\sigma^3 = 1, \quad \tau^2 = 1$$

We will now find combinations of roots that transform nicely under σ and τ. The key is a "discrete Fourier transform" (also called the "Lagrange resolvent")

$$A = \alpha_1 + \omega\alpha_2 + \omega^2\alpha_3$$

where

$$\omega = e^{\frac{2\pi i}{3}}$$

is a cube root of unity. Clearly,

$$\sigma : A \mapsto \alpha_2 + \omega\alpha_3 + \omega^2\alpha_1 = \omega^{-1}\left(\alpha_1 + \omega\alpha_2 + \omega^2\alpha_3\right)$$

using $\omega^2 = \omega^{-1}$; i.e.,

$$\sigma : A \to \omega^{-1}A$$

Since $\omega^3 = 1$ we see that A^3 is invariant under σ. But acting with τ on A gives something new:

$$\tau : A \mapsto B \equiv \alpha_2 + \omega\alpha_1 + \omega^2\alpha_3$$

Again B^3 is invariant under σ . And $\tau : B \leftrightarrow A$ since $\tau^2 = 1$.Thus

$$C = A^3 + B^3, \quad D = A^3B^3$$

are invariant under both σ and τ; therefore they are invariant under all permutations of roots. *But that means that C, D can be written in terms of the original coefficients s_1, s_2, s_3.* This is the crucial point.

It is good for your soul to work out C, D in terms of s_1, s_2, s_3 explicitly. If you don't care about your soul, you can make Mathematica do this work (use an operation called Symmetric Reduction).

Exercise 2. Show that

$$C = 2s_1^3 - 9s_1s_2 + 27s_3, \quad D = s_1^6 - 9s_1^4s_2 + 27s_1^2s_2^2 - 27s_2^3$$

We have climbed to a point from which we see the solution to any cubic. From the coefficients of the cubic, we find C and D . Then find A^3 and B^3 because they are solutions of the quadratic

$$(x - A^3)(x - B^3) \equiv x^2 - Cx + D = 0.$$

Taking cube roots of these solutions will give A and B . Once you know A and B we can find $\alpha_1, \alpha_2, \alpha_3$ by solving the linear system (i.e., inverting the discrete Fourier transform)

$$s_1 = \alpha_1 + \alpha_2 + \alpha_3, \quad A = \alpha_1 + \omega\alpha_2 + \omega^2\alpha_3, \quad B = \alpha_1 + \omega^2\alpha_2 + \omega\alpha_3.$$

Exercise 3. Write a Mathematica (or python) program that implements this way of solving a cubic.

Exercise 4. Express the discriminant $\Delta = (\alpha_1 - \alpha_2)^2(\alpha_2 - \alpha_3)^2(\alpha_3 - \alpha_1)^2$ as a polynomial in the coefficients s_1, s_2, s_3 of the cubic.

Answer:

$$\Delta = -4s_3s_1^3 + s_2^2s_1^2 + 18s_2s_3s_1 - 4s_2^3 - 27s_3^2.$$

The lesson is that understanding the symmetries of an equation allows us to develop a strategy to solve it. Or, to show that it cannot be solved by the methods under consideration.

1.3.4. *Higher order equations**

A radical equation like $x^n = s$ has a commutative symmetry group Z_n with n elements (the group of cyclic permutations). The general nth order equation has symmetry under the permutation group S_n (which has $n!$ elements). The reason why cubics can be solved is that S_3, although non-commutative, can be built out of two commutative groups Z_3 and Z_2.

The precise statement is that Z_3 is a normal subgroup of S_3 and that $1 \to Z_3 \to S_3 \to Z_2$ is an exact sequence of group homomorphisms. Or equivalently that S_3 is the extension of Z_3 by Z_2.

The permutation group of four elements S_4 also can be written as an iterated extension of commutative groups. This explains why the quartic can also be solved in general. The formulas are more complicated, but a "resolvent" ("discrete Fourier transform") allows us to reduce the general quartic to a system of linear equations and taking fourth roots.

But the quintic is no longer solvable by such elementary methods. The permutation group of five elements S_5 can be written as an extension of the group of even permutations A_5 by Z_2. But A_5 is a simple group: It cannot be decomposed as the extension of any smaller group. (A better name for such groups would have been "prime groups". But we are stuck with the name.) This was the insight of Abel and Galois, who are founders of this branch of mathematics: The theory of groups and fields.

Of course, nowadays, we would solve polynomial equations numerically by an iterative approximation method such as Newton–Raphson. But the "complexity theory" of algebraic equations continues as a fascinating subject; e.g., work by McMullen *et al.* and various conjectures by Smale. Group theory continues to play an important role.

1.3.5. *The Erlangen Program*

Before they became standard tools in physics, groups became important in geometry through a visionary research program of Felix Klein at Erlangen. It is obvious that a sphere has symmetry under rotations. Also, a square lattice (points on the plane separated by equal steps along each axis), has a symmetry under some translations and rotations. Klein saw a way to extend these ideas to non-Euclidean geometries such as a hyperboloid. These led Lie, Noether, Poincare and others to study deeper connections between groups and physics. It was not until the discovery of relativity and quantum mechanics that the deep role that symmetry plays in physics became clear. A classic text by Hermann Weyl was important in bridging physics and mathematics. The book [1] gives much more detailed account of the mathematical side of this story.

1.3.6. *Galois Theory in other disciplines* (Speculative)*

The essence of Klein's idea is this: Each geometric object has a set of transformations which leave it invariant. For the plane, these are translations and rotations. For the upper half-plane (which can be mapped invertibly to the hyperboloid), they are the fractional transformations $z \to \frac{az+b}{cz+d}$, for real a, b, c, d. Similarly, what we mean by a problem being solvable is that a solution can be found by combining

certain types of operations. For example, the classical definition of a solvable algebraic equation allows for radicals plus the usual arithmetical operations of addition, multiplication and division. Could there be a similar approach in other disciplines?

- Lie's motivation in developing the theory of Lie groups was to find a "Galois theory" of ordinary differential equations. To this day, only a part of his vision has been realized (e.g., Differential Galois Theory). The class of operations allowed to solve a dynamical system is now much broader than in Lie's day: We must include systems than can be solved to reasonable accuracy by numerical methods. In this sense, celestial mechanics is solvable (despite chaos) but not fluid mechanics (because of turbulence). How to make this precise?
- McMullen *et al.*, have extended Galois theory to include solution by iteration (e.g., solvable numerically). Amazingly, the dividing line is at order six: Quintics can also be solved by iteration of a rational function of one variable, but sextics need such an iteration of a function of two variables! Smale has interesting conjectures on further extensions of these ideas.
- Is there a "Galois theory" or "Erlangen program" of solvability of quantum systems? One approach might be to consider as the "Galois group" the fundamental group of the Riemann surface of the energy spectrum, thought of as a function of the coupling constant (analytically continued to complex values). Typically, one expects an infinite number of square root branch points, which might accumulate. Bender and Wu have worked this out in detail for the anharmonic oscillator. But a general theory eludes us.
- The Schrodinger equation of small atoms and molecules can be solved ab initio. But large molecules and materials are beyond the ability of even the largest computers. What is the precise measure of quantum complexity?
- Quantum groups were discovered while solving certain spin chains (Bethe, Yang and Baxter). Do they also appear in more realistic systems?
- Tarski was inspired by the Erlangen program to consider a similar approach to logic: Every logical system has some set of operations that are allowed, allowing us to transform sentences without changing their truth value. This could lead to a theory of Computational Complexity.
- What kind of groups underlie computer science problems of class P (that can be solved in Polynomial time)? Are they different from the groups of class NP?

This page intentionally left blank

Chapter 2

GROUPS AND THEIR REPRESENTATIONS

Once you understand the basic structure of a physical or mathematical theory, it is useful to summarize the basic laws as axioms: Independent facts from which all others can be derived. This was first achieved for plane geometry by Euclid. For mechanics by Newton. There is always a period of experimentation and discovery before a subject become mature enough to be axiomatized. Algebraic concepts we study in this book were developed over the nineteenth century and formalized in the early twentieth century. It is time we gave a mathematically precise definition of a group.

Definition. A *group* is a set G along with a binary operation $G \times G \to G$, which obeys

- Associativity: $(ab)c = a(bc), \quad \forall a, b, c \in G$
- There is an identity element $e \in G$ such that $ea = ae = a$
- For every $a \in G$ there is an inverse such that $aa^{-1} = a^{-1}a = e$.

Usually, you should think of a group element as a transformation on some other set of objects. For example, rotations of an equilateral triangle around its center. Or the permutations of a deck of cards.

Definition. A subset of a group which contains all the products and inverses of its elements is a subgroup.

2.1. Examples

Examples breathe life into an abstract theory. Group theory abounds in many fascinating examples. Much of the following may not make sense in a first reading, but I encourage the reader to look into each example and make as much

sense of them as possible. On a later reading perhaps more of them will become clear.

(1) The smallest group is the singleton set containing just one element. Its product with itself is itself, and it is its own inverse. This is called the trivial group. The empty set cannot be a group. (Why not?)
(2) The only group of two elements is $Z_2 = \{1, \omega | \omega^2 = 1\}$.
(3) The cyclic group of n elements is $Z_n = \{1, \omega, \omega^2, \ldots, \omega^{n-1} | \omega^n = 1\}$. The multiplication law is $\omega^m \omega^k = \omega^{m+k}$. The inverse of ω^m is ω^{n-m}. It describes cyclic permutations of a set of n elements.
(4) The group multiplication law does not have to be commutative. The group of quaternions has 8 elements $1, i, j, k, -1, -i, -j, -k$ satisfying $ij = k = -ji, jk = i = -kj, ki = j = -ik, i^2 = j^2 = k^2 = -1$. **Exercise:** Find 2x2 matrices that satisfy these relations.
(5) The permutations on a set of n elements is a group, also called the symmetric group S_n. It has $n!$ elements. Cyclic permutations are a subgroup. There are volumes dedicated to the study of this group, and its representations. **Exercise:** Show that any finite group is a subgroup of S_n for some n (**Hint:** This is not as impressive as it sounds: Almost a tautology.)
(6) Any permutation can be written as a product of transpositions (pairwise permutations). Even permutations (product of an even number of transpositions) form a subgroup called the alternating group A_n.
(7) The set of integers is a group under addition. But not under multiplication. (Why not?)
(8) The set of rational numbers is a group under addition. Also, the set of non-zero rational numbers is a group under multiplication.
(9) Similarly for real and complex numbers.
(10) Let $\mathbb{Z}/n\mathbb{Z}$ be the set of integers modulo some natural number n. This is an additive group, isomorphic to the cyclic group Z_n. It is not a group under multiplication because for example, 0 does not have an inverse.
(11) The subset $(\mathbb{Z}/n\mathbb{Z})^\times$ of elements of $\mathbb{Z}/n\mathbb{Z}$ which are co-prime to n is a group under multiplication as well. (That is, $ab \equiv 1 \mod n$ has a solution iff a is co-prime to n). The number of elements in this group is the Euler totient function $\tau(n)$, a central fascination of number theorists. **Exercises:** Show that if p is a prime, all of the non-zero elements of $\mathbb{Z}/p\mathbb{Z}$ have inverses. So, $\tau(p) = p - 1$. What is $\tau(10)$?
(12) The set of $n \times n$ complex matrices of non-zero determinant is a group. It is denoted by $GL_n(C)$ or $GL(n, C)$. GL stands for "general linear".

(13) Similarly, $GL_n(R)$ is the group of real matrices of non-zero determinant.
(14) Matrices of determinant one used to be called "special". Thus $SL(n, C) \subset GL(n, C)$ is the subgroup of complex $n \times n$ matrices of determinant one. Similarly for $SL(n, R)$.
(15) The inverse of a matrix with integer entries also has integer entries, as long as the determinant is one. (Prove this.) Of course the product of two matrices with integer entries is integral as well. Thus, $SL(n, Z)$ is a group although $GL(n, Z)$ is not.
(16) For example, $SL(2, Z) = \left\{ \begin{pmatrix} a & b \\ d & d \end{pmatrix} \mid ad - bc = 1 \right\}$ is a group. It is called the modular group and is important in the study of doubly periodic functions (i.e., elliptic functions).
(17) A real matrix is orthogonal if its transpose is also its inverse: $gg^T = 1$. The set of real orthogonal matrices is a group, called $O(n)$.
(18) The determinant of $g \in O(n)$ has to be ± 1. (Why?) The subset of orthogonal matrices of determinant one is called $SO(n)$. We will see that it is the group of rotations in $\mathbb{R}^n$
(19) A complex matrix is said to be unitary if its inverse is its complex conjugate transpose (hermitian conjugate): $gg^\dagger = 1$. The set of unitary matrices $U(n)$ is a group. The subset of unitary matrices of determinant one is the group $SU(n)$. Special cases such as $SU(2)$, $SU(3)$ are at the foundation of particle physics.
(20) There is a close relationship between $SU(2)$ and $SO(3)$. They are the same except for a sign. We will see that this is important in understanding the spin of an electron.
(21) The set of twists you can do to a Rubik's cube is a group. **Exercise:** How many elements does it have? (This is a lot harder than it seems.)
(22) There is a huge literature on applying symmetry groups to molecular and crystal physics. This used to be the main application of groups to physics. We won't do much of that in this book, saving our energy for more modern applications.
(23) Many viruses have symmetric shapes. It is not yet clear if this is important in understanding their biology.
(24) The set of all smooth co-ordinate transformations (of non-zero Jacobian) is a group. This is the invariance group of General Relativity, Einstein's theory of gravity.
(25) The set of smooth functions from space-time to $SU(n)$ is a group. The invariance group (gauge group) of the standard model of elementary particle physics is built out of this.

(26) The set of smooth transformations on a manifold M (with Jacobian equal to one) is a group, which we will call $SDiff(M)$. When the dimension of M is two or three, this group is the configuration space of an incompressible fluid; the Euler equations, which describe the flow of such a fluid, are equations for geodesics on it.

2.1.1. *A map from a group to another that preserves multiplication is called a homomorphism. An isomorphism is a homomorphism that is one-to-one and onto.*

Thus, a homomorphism $f : G \to H$ will satisfy

$$f(g_1 g_2) = f(g_1) f(g_2)$$

The set of elements of G that are mapped to the identity of H is called the *kernel* of this homomorphism. The kernel of any homomorphism is a subgroup. If there is an isomorphism between two groups they have the same structure: At some abstract level they are identical. A moment's reflection will show you that the kernel of an isomorphism is trivial.

2.1.2. *An automorphism is a one-one and onto map $f : G \to G$ that preserves the multiplication*

In a sense, an automorphism is a symmetry of the group itself. The most obvious example is a conjugation. Pick some element $h \in G$ and define

$$f_h(g) = hgh^{-1}.$$

Then

$$f_h(g_1) f_h(g_2) = hg_1h^{-1}hg_2h^{-1} = hg_1g_2h^{-1} = f_h(g_1g_2).$$

Such an automorphism is called an "inner automorphism" . Many groups also have "outer automorphisms" that are not of this type.

2.1.2.1. *Examples*

(1) There is an isomorphism from the cyclic group of n elements, Z_n, to the group of nth roots of unity: $\omega \mapsto e^{\frac{2\pi i}{n}}$.
(2) The determinant is a homomorphism from $O(n) \to \{1, -1\}$. Its kernel is $SO(n)$.

(3) There is a homomorphism $R : SU(2) \to SO(3)$ whose kernel is the subset $\{1, -1\}$ (where 1 stands for the identity matrix). We will study this one in more detail later as it is important in quantum mechanics.
(4) Complex conjugation is an outer automorphism of the group $SL(2, C)$.

2.2. Representations

2.2.1. *A homomorphism $r : G \to GL(n, C)$ is called a representation*

Representations allow us to think of group elements in terms of matrices, which are much more concrete objects, almost as familiar as numbers.

2.2.1.1. *A unitary representation is a homomorphism to $r : G \to U(n)$*

An orthogonal representation is $r : G \to O(n)$ etc.

In quantum mechanics, symmetries are unitary representations. They are therefore the most important representations. In general the representations might be in terms of infinite dimensional matrices. But we will mostly confine ourselves to finite dimensional matrices as the theory is so much simpler, but still useful in physics. Most representations we study will be unitary or orthogonal.

2.2.1.2. *Example*

(1) Any element of the permutation group S_3 can be written as a product of a cyclic permutation $\sigma : PQR \to QRP$ and a reflection $\tau : PQR \to QPR$. We say that S_3 is generated by σ, τ. A representation in $O(3)$ is given by

$$\sigma \mapsto \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \quad \tau \mapsto \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

2.2.1.3. *Recall that the direct sum of two matrices is given by stacking them as a bigger matrix $A \oplus B = \left(\begin{smallmatrix} A & 0 \\ 0 & B \end{smallmatrix}\right)$. This can be extended to the direct sum of two representations $r_1 \oplus r_2(g) = \left(\begin{smallmatrix} r_1(g) & 0 \\ 0 & r_2(g) \end{smallmatrix}\right)$*

Clearly, the direct sum of unitary representations is also unitary. Physically, the direct sum of representations describes two subsets of states of the same system, that are not mixed into each other by the action of the symmetry group. For example, rotations do not mix the p states of hydrogen with the d states. They

belong to two separate direct summands in the representation of the rotation group on the space of states of hydrogen. More on this later.

2.2.1.4. *Conversely, if a representation can be decomposed as the sum of two other representations (if every representation matrix is the direct sum of two others) we say that the representation is "completely reducible"*

2.2.1.5. *Two representations are equivalent if there is an invertible matrix S (independent of g) such that $r_1(g) = Sr_2(g)S^{-1}$, $\forall g \in G$*

If S is unitary we say that they are unitarily equivalent. Equivalent representations only differ by a choice of basis: Not different in an essential way.

2.2.1.6. *The direct product (also called tensor product) of two representations is defined in terms of the direct product of matrices in a similar way*

Physically the direct product describes a system that has two parts. For example, the hydrogen atom has a proton and an electron. The representation of the rotation group is the direct product of the representations on each constituent.

2.2.2. *Group action*

A group G is said to act on a set X if there is a map $G \times X \to X$ which respects the group multiplication

$$g_1 g_2(x) = g_1 \left(g_2(x)\right)$$

Given $x \in X$, the set of all the elements you can get from it by acting with some g is called its "orbit".

The action is said to be *transitive* if there is a single orbit: Any element in X can be taken to any other element by some g. More typically, the set of orbits $G\backslash X$ contains many elements.

(1) Of course, a representation is the particular case of a group action, where G acts on a vector space through matrices.
(2) Rotations around the origin act on $\mathbb{R}^3$. The orbit of the origin is itself. The orbit of every other point is the sphere (centered at the origin) passing through it. Thus, an orbit is determined by its radius. In other words, there is a one-one

correspondence between the set of orbits $SO(3)\backslash\mathbb{R}^3$ and the set of non-negative real numbers.

(3) Another obvious example is the symmetric group S_n acting on some set of n elements by permutations.

(4) A group acts on itself by multiplication on the left. There is a single orbit, which is the whole group.

(5) A subgroup $H \subset G$ acts on G by multiplication on the left. In this case, there can be several distinct orbits. The set of orbits $H\backslash G$ is called the "left coset space" of H in G. (There is a similar story with right multiplication.) For example, rotations around the third axis is a subgroup of $SO(3)$. Since it only affects the first two co-ordinates, this subgroup is isomorphic to $SO(2)$. The coset space $SO(2)\backslash SO(3)$ is $\mathbb{S}^2$, the sphere (Prove this!).

(6) Words like "orbit" are a reminder that this whole theory originates in mechanics. Specifically, the time evolution of a system is an action of the additive group of real numbers on its phase space. The simplest example is the harmonic oscillator with hamiltonian $H = \frac{1}{2}p^2 + \frac{1}{2}\omega^2 q^2$. Its orbits are ellipses:

$$q(t) = A\cos\omega t, \quad p(t) = -\omega A \sin\omega t$$

(7) The orbits of a chaotic dynamical system can be immensely complicated. Your physical intuition can be misled if you spend too much time on exactly solvable systems like the harmonic oscillator. In general, group actions are deep, complicated things that describe all sorts of natural and mathematical phenomena.

2.3. Quotient of Groups

Definition. A subgroup H of G is said to be *normal* if the left and right cosets are equal.

That is, the sets gH and Hg are the same for every $g \in G$. Equivalently, for every $h \in H$ and $g \in G$ there is an $\tilde{h} \in H$ (which may depend on g) such that

$$gh = \tilde{h}(g)g$$

If G is abelian every subgroup is normal. In general this is not the case. We will denote a normal subgroup by $H \lhd G$.

The importance of this idea is that given $H \lhd G$ we can define a multiplication of cosets which turns G/H into a group. If we take a representative of the right

coset gH and another of $g'H$ and multiply them, we can rewrite

$$ghg'h' = gg'\tilde{h}(g)h'$$

which is an element of the coset $gg'H$. This product can be seen to be associative and has an inverse; thus $F = G/H$ is itself a group. The map $p : G \to F$ which sends every $g \in G$ to its right coset gH is a surjective group homomorphism. The kernel is H, the coset of the identity.

Thus we can restate this situation as an "exact sequence" of group homomorphisms:

$$\{1\} \to H \xrightarrow{i} G \xrightarrow{p} F \to \{1\}$$

Here, "exact" means that the image of each map is equal to the kernel of the next one. The first map simply sends the only element in the trivial group to the identity of H. The second, i is simply the inclusion map of H into G as a subset. The last map sends every element of F to the only element of the trivial group.

Since the image of the first map must be the kernel of i, only the identity of H is sent to the identity in G. Equivalently, i is injective. The projection map p is surjective: Every element of F comes from some element of G.

If we have such an exact sequence of groups, H is a normal subgroup and $F = G/H$ .We say that G is an extension of H by F.

2.3.1. *Cardinality of a Group and the Index of a subgroup*

If G is a finite set the number of elements of it $\#(G)$ is called its *order* or *cardinality.* The set of left cosets G/H and right cosets $H\backslash G$ are the same set when H is a normal subgroup of G. Even when H is not normal, there will always be a 1-1 correspondence between G/H and $H\backslash G$. So the left and right cosets have the same cardinality. It is called the *index* of H w.r.t. G. A moments thought will tell you that the cardinality of G is the product of those of H and G/H:

$$\#(G) = \#(H)\ \#(G/H).$$

So, the order of a subgroup H is a divisor of the order of G. In particular, a group of prime order has no proper[1] subgroups!

The idea of cardinality even makes sense for some infinite groups; for example, $\mathbb{Z}$ is a countable set, so has the same cardinality as any other countable set. (Recall that there is a 1–1 correspondence between any two countable sets.) This is usually denoted by the transfinite number $\aleph_0$. The idea continues to make sense

[1] i.e.,The only subgroups are the identity or the whole group.

for uncountably infinite sets, but is more tricky and we will avoid thinking of it this way.

It is possible for a countably infinite group to have a subgroup of finite index. For example $G = \mathbb{Z}$ and $H = 2\mathbb{Z}$. Although each is infinite, G/H is isomorphic to the group of two elements, $\{1, -1\}$. Thus, H is an index two subgroup of G.

Exercise. Give an example of a group with no proper subgroup.

Answer: The group of cyclic permutations of prime order Z_p has no proper subgroup.

Exercise. Show that the subgroup $\{1, \tau\}$ of S_3 is not normal. However the subgroup $A_3 = \left\{1, \sigma, \sigma^2\right\}$ of S_3 is normal. Determine the factor group $F = S_3/A_3$. Construct a homomorphism $p : S_3 \to F$ whose kernel is precisely A_3.

Solution: Recall that S_3 can be described by its generators σ, τ and relations as

$$S_3 = \langle \sigma, \tau \mid \sigma^3 = 1 = \tau^2, \sigma\tau = \tau\sigma^{-1} \rangle$$

Its elements are

$$S_3 = \{1, \sigma, \sigma^2, \tau, \sigma\tau, \sigma^2\tau\}$$

The subgroup $\{1, \tau\}$ is not normal. The left coset of $\sigma \in S_3$ is

$$\sigma\{1, \tau\} = \{\sigma, \sigma\tau\}$$

while its right coset is

$$\{1, \tau\}\sigma = \{\sigma, \tau\sigma\}$$

These are not the same because $\tau\sigma \neq \sigma\tau$, as we saw earlier.

But the subgroup $A_3 = \{1, \sigma, \sigma^2\}$ of S_3 generated by σ is indeed a normal subgroup. The left coset of τ is

$$\tau A_3 = \{\tau, \tau\sigma, \tau\sigma^2\}$$

while the right coset is

$$A_3\tau = \{\tau, \sigma\tau, \sigma^2\tau\}$$

As sets these are the same because, as we saw earlier,

$$\tau\sigma = \sigma^{-1}\tau = \sigma^2\tau$$

and similarly

$$\tau\sigma^2 = \sigma\tau$$

The homomorphism $p : S_3 \to Z_2$ defined by

$$p : \sigma \mapsto 1, \quad p : \tau \to -1$$

has A_3 as its kernel. So,

$$1 \to A_3 \to S_3 \to Z_2 \to 1$$

is an exact sequence: S_3 is an extension of an abelian group A_3 by another abelian group Z_2.

We used this fact to solve the cubic.

Example. Some, but not all, of this can be generalized to arbitrary permutations. Suppose $G = S_n$ is the group of permutations of n objects and H is the subgroup of even permutations A_n. Also, $F = \{1, -1\}$. Then the map p that sends every even permutation to 1 and every odd permutation to -1 is a surjective homomorphism with kernel A_n. You can verify that A_n is indeed a normal subgroup: If g is an even element of G, its left and right cosets are both just A_n. If g is an odd permutation, its left as well as right cosets are both the set of all odd permutations.

The case $n = 4$ is more intricate than $n = 3$ but still understandable. Cyclic permutations are no longer contained in A_4, nor is it any more abelian. Yet, A_4 has an abelian normal subgroup K consisting of two pairwise interchanges

$$(12)(34), \quad (13)(24), \quad (14)(23)$$

as well as the identity. (Here, $(12)(34)$ stands for the interchange of 1and 2 followed by that of 3 and 4.)

In fact $K \lhd A_4$; i.e., K is a normal subgroup of A_4. The quotient A_4/K is isomorphic to the cyclic group of three elements Z_3:

$$1 \to K \to A_4 \to Z_3 \to 1$$

Thus, S_4 can be built as a succession of abelian extensions: First we get A_4 as an extension of K by Z_3, and then S_4 is an extension of A_4 by Z_2. This can be used to devise a method for solving the quartic.

When $n = 5$ such ideas fail. A_5 cannot any more be broken up into smaller groups: It has no normal subgroups at all (other than the identity and itself). This is why the quintic cannot be solved by algebraic methods.

2.3.2. *Abelian Extensions, Solvable groups and Simple Groups*

If $H \lhd G$ and $F = G/H$ is abelian, we say that G is an abelian extension of H. If a group can be obtained by iterating this construction (starting with an abelian group), we say it is solvable. For example, S_3 and S_4 are solvable but not S_5.

A group that has no normal subgroups (other than the identity and itself) is said to be simple. That is, it cannot be obtained as an extension of any smaller group. For example, A_5 is simple [5].

Simple groups are the fundamental building blocks of group theory. So, it is interesting to find all the finite simple groups (i.e., classify them up to isomorphism). After many decades of work this project has been completed. Fascinating as it is, the connections to physics are somewhat tangential. So, we do not pursue this direction.

2.4. The Fundamental Group

An important application of group theory is to another branch of mathematics, algebraic topology. This field has its roots in Poincare's foundational work on dynamical systems. But now topology is a thriving branch of pure mathematics which has turned out to be useful in modern physics in various ways. We will not digress to review the basic ideas of continuous maps between topological spaces. There are several standard textbooks, e.g., [6]. We will be content with summarizing the notion of the fundamental group of a topological space, as it has several uses within the theory of Lie groups.

Once we know that X is a topological space, it makes sense to talk of a continuous map $\gamma : [0, 1] \to X$; it defines a curve in X. A closed curve γ in X, based at a point $x_0 \in X$, is a curve starting and ending at x_0:

$$\gamma(0) = x_0 = \gamma(1).$$

Given two closed curves γ and $\tilde{\gamma}$ based at the same point, we can get a third curve as their composition; that is, we go around γ, (at twice the speed) and once we return to x_0 we go around $\tilde{\gamma}$, again with twice the speed. Translated into a formula, this is

$$\tilde{\gamma} \circ \gamma(t) = \begin{cases} \gamma(2t) & 0 \le t \le \frac{1}{2} \\ \tilde{\gamma}(2t-1) & \frac{1}{2} \le t \le 1 \end{cases}$$

An interesting question is whether this can be turned into a group operation. The identity would be the curve that simply stays at the base point x_0; i.e., the constant map:

$$e(t) = x_0, \quad 0 \le t \le 1$$

A way to define the inverse would be to traverse the curve in reversed time:

$$\gamma^{-1}(t) = \gamma(1-t).$$

The idea does not quite work because the composition $\gamma^{-1}(t) \circ \gamma(t)$ is not the constant map. But we can tweak the idea a bit and get a group out of the composition of curves. There is an equivalence relation among curves (continuous deformation) such that the composition above becomes a group operation on equivalence classes of curves.

To define this precisely, suppose γ and $\tilde{\gamma}$ are curves in a connected space X, that start and end at the same point:

$$\gamma(0) = \tilde{\gamma}(0), \quad \gamma(1) = \tilde{\gamma(1)}$$

We say that γ can be continuously deformed to $\tilde{\gamma}$ if there is a continuous function of two variables

$$\phi : [0,1] \times [0,1] \to X$$

such that

$$\phi(0,t) = \gamma(t), \quad \phi(1,t) = \tilde{\gamma}(t).$$

In other words, there is a continuous function of two variables that interpolates between γ and $\tilde{\gamma}$. It is not hard to see that this is an equivalence relation.

The set of equivalence classes of closed curves based at $x_0 \in X$ is denoted by $\pi_1(X, x_0)$. It is possible to show[6] that the equivalence class $[\tilde{\gamma} \circ \gamma]$ depends only on the equivalence classes $[\tilde{\gamma}]$ and $[\gamma]$. And moreover that $\gamma' \circ (\tilde{\gamma} \circ \gamma)$ is deformable to $(\gamma' \circ \tilde{\gamma}) \circ \gamma$. Furthermore, $\gamma^{-1}(t) \circ \gamma(t)$ is deformable to the constant map $e(t)$. Thus the set of equivalence classes $\pi_1(X, x_0)$ is a group. It is called the fundamental group of X based at x_0.

A *connected* space X (some would say *path connected*) is one where there is a continuous curve starting at any point $x \in X$ to any other point $y \in X$. If X is connected, $\pi_1(X, x_0)$ is isomorphic to $\pi_1(X, y_0)$, the fundamental group based at some other point $y_0 \in X$. (A continuous curve starting at x_0 and ending at y_0 can be used to construct an isomorphism between the two fundamental groups.) So, in this case, we can omit the base point and talk of the fundamental group $\pi_1(X)$.

2.4.1. *Examples*

- Let $\mathbb{C}^\times = \{z \mid z \in \mathbb{C}, z \neq 0\}$ be the space of non-zero complex numbers. A continuous closed curve in this space cannot pass through the origin. We can pick any non-zero complex number as the base point z_0. The space is connected as there is always a continuous curve connecting any two non-zero points. Any curve that does not contain the origin in its interior can be deformed to the constant map; this equivalence class is the identity element in $\pi_1(\mathbb{C}^\times)$. But

curves that surround the origin may not be deformable to each other. We can associate a winding number to each continuously differentiable curve

$$n = \frac{1}{2\pi i}\int_0^1 \frac{d\gamma(t)}{\gamma(t)}$$

For example, the curves

$$\gamma_n(t) = e^{2\pi i n t}, \quad 0 \le t \le 1, \quad n \in \mathbb{Z}$$

wind around the origin n times. This winding number is invariant under smooth deformations: If γ and $\tilde{\gamma}$ are boundary values of a function of two variables ϕ, a simple application of Stokes' theorem of vector calculus shows that

$$\int_0^1 \frac{d\gamma(t)}{\gamma(t)} - \int_0^1 \frac{d\tilde{\gamma}(t)}{\tilde{\gamma}(t)} = \int_0^1\int_0^1 \frac{\partial^2\phi}{\partial s \partial t} ds dt = 0$$

In fact any two curves of the same winding number can be deformed into each other. And thus $\pi_1(\mathbb{C}^\times) = \mathbb{Z}$ as groups; composition of curves corresponds to adding their winding numbers. The class of curves that winds once around the origin in a counter-clockwise direction is the generator of the group. Its inverse is the class that winds once around in the clock-wise direction.

- If we remove two points from the complex plane, we get a space with a non-abelian fundamental group.

$$X = \{z \mid z \in \mathbb{C}, z \neq A, B\}, \quad A, B \in \mathbb{C}, \quad A \neq B$$

This group $\pi_1(X)$ is of some independent interest, so let us look into it in a bit more detail. The continuous curves are curves in the plane that do not pass through either A or B. For example, here is a curve that winds around A once:

$$\alpha(t) = A + \epsilon e^{2\pi i t}, \quad 0 < \epsilon < |A - B|$$

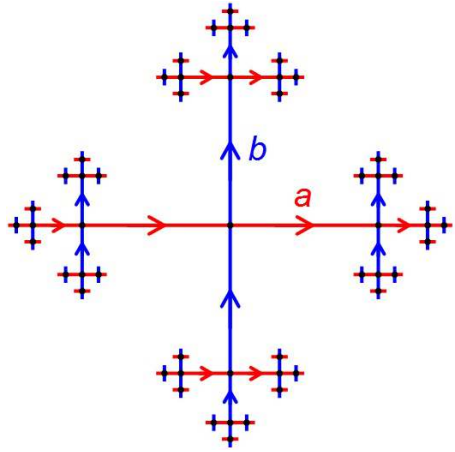
It is a circle centered at Awhose radius is less than the distance from A to B; so the circle does not contain B in its interior. Similarly there is a curve β that winds around B. Let a be the equivalence class of curves deformable to α and similarly for b and β. (It is important that the deformations are functions of two variables $\phi : [0,1] \times [0,1] \to X$ that never take the values A or B). Again, we have classes a^m that wind around A some integer m times (and similarly b^n). But, knowing the winding numbers m, n around A and B no longer determines a curve up to deformation. For example

$$b^{-1}a^{-1}ba$$

(which goes around A, then B, then around A and then B in the opposite direction) is not deformable to a constant curve. In fact there are no relations at all among the generators a and b (except those implied by the group properties, such as associativity.) The fundamental group is F_2, the Free group generated by two elements a and b. This group can be pictured as a tree graph rooted at the identity and branching at every vertex into three copies of itself: Start at the identity and a move to the left (right) by one step is a (rep. a^{-1}) and a move up (down) by one step is b (resp. b^{-1}). In the next generation, we reduce the step size to a half (so that the picture will fit in a plane); starting at a we can get a^2, ba or $b^{-1}a$. (Of course $a^{-1}a$ just returns to the identity.) We can similarly depict $b^2, ab, a^{-1}b$ and $a^{-2}, ba^{-1}, b^{-1}a^{-1}$ and $b^{-2}, ab^{-1}, a^{-1}b^{-1}$. The generation after that, edges have length $\frac{1}{2^2}$ and so on. The resulting tree graph is shown in the figure

The reduction of lengths by a factor two at each generation is not an intrinsic property of the Free Group: It is done so that the picture will fit on the plane. A more natural way to draw this graph would be on a hyperboloid of constant curvature: The edges of the graph will all have the same hyperbolic (Lobachewsky) distance. Indeed, F_2 is isomorphic to a subgroup of the isometry group $SL_2(R)$ of the hyperboloid under the map

$$a \mapsto \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \quad b \mapsto \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$$



2.4.2. *Covering Spaces*

If a group G acts properly on a manifold $\tilde{X}$, the space of orbits $X = G\backslash\tilde{X}$ is again a manifold. (We will call the space of orbits the quotient, even when $\tilde{X}$ is not a group.) [2]

We say that $\tilde{X}$ is a covering space of X. The fundamental groups of X and $\tilde{X}$ are related in an interesting way: There is an exact sequence

$$\{1\} \to \pi_1(\tilde{X}) \to \pi_1(X) \to G \to \{1\}$$

In other words, the fundamental group of X is an extension of that of the covering space by G.

If $\tilde{X}$ is connected and $\pi_1(\tilde{X})$ is the trivial group, we say that $\tilde{X}$ is simply connected. In this case the fundamental group of the quotient $G\backslash\tilde{X}$ is simply G. We say then that $\tilde{X}$ is the universal cover of X. Every manifold admits such a description as the quotient of a simply connected space by a proper action. Let us consider some examples.

- The group of integers $G = \mathbb{Z}$ acts properly on the space of real numbers $\tilde{X} = \mathbb{R}$ by translation $x \mapsto x + n$. The quotient $\mathbb{Z}\backslash\mathbb{R}$ can be identified as the circle $\mathbb{S}^1$; We just have to think of the equivalence class of $x \in X$ as mapped to $e^{2\pi i x} \in \mathbb{S}^1$. It is not hard to see that $\mathbb{R}$ is simply connected. So, $\pi_1(\mathbb{S}^1) = \mathbb{Z}$.
- We can extend this to an action of $G = \mathbb{Z}^d$ on $\tilde{X} = \mathbb{R}^d$ by translation on each co-ordinate component: $x_i \mapsto x_i + n_i, i = 1, \cdots d$. The quotient can be identified as the product of d circles, which is the torus $\mathbb{T}^d$ of dimension d. Since Euclidean space is simply connected, we get $\pi_1\left(\mathbb{T}^d\right) = \mathbb{Z}^d$.
- The group $SU(2)$ can be shown to be, as a manifold the three sphere $\mathbb{S}^3$. More on this later.
- The universal covering space of the complex plane with two points removed can be realized as the upper half plane. It will take us too far afield to describe the proper action of the Free group F_2 on the half plane that gives this covering in detail. The main idea is to represent the Free group by 2×2 real matrices and to use the fractional linear transformation $z \mapsto \frac{az+b}{cz+d}$.

[2] The property of manifolds that we are using is that they are Hausdorff: every pair of unequal points have neighborhoods that do not overlap. A mnemonic is: the points can be "housed off" from each other. Also, "properly" means that every $\tilde{x} \in \tilde{X}$ has a neighborhood $\tilde{U}$ which has no overlap with $g\tilde{U}$, for every $g \in G$ which is not the identity.

In particular, proper group actions cannot have fixed points; also the orbit of a point cannot converge to some point. In such situations, the quotient X may not have the Hausdorff property and so are not manifolds: they are sometimes called "orbifolds".

2.5. Appendix: Vector Spaces

It is useful to have an axiomatic point of view on vectors and matrices as well. In particular it is useful to know about tensors.

2.6. Appendix: Lightning Review of Quantum Mechanics

Good references are [7, 8, 9]. The summary below is not a substitute for a proper course in quantum mechanics: It usually takes a year to learn the material contained in this section. But it might be useful as a review.

2.6.1. *The Postulates*

Quantum theory is still not completely developed. Questions about measurement and interpretation are still being worked out (e.g.,"weak measurement"). Nevertheless we can say, after almost a century of experimental tests, many things for certain about how quantum theory works. It is not too early to summarize them as a set of postulates.

2.6.1.1. *The states of a physical system are represented by vectors in a complex Hilbert space*

This means that we can take linear combinations

$$\alpha \mid \psi\rangle + \beta \mid \phi\rangle$$

of two states $\mid \psi\rangle$ and $\mid \phi\rangle$ to get another state. The quantities α, β are complex numbers. There is a way to take the inner product (scalar product) of two states to get a complex number, denoted by $\langle \psi \mid \phi\rangle$.

This inner product is linear in the second argument

$$\langle \psi \mid \alpha\phi + \beta\chi\rangle = \alpha\langle \psi \mid \phi\rangle + \beta\langle \psi \mid \chi\rangle$$

and anti-linear in the first entry

$$\langle \alpha\psi + \beta\chi \mid \phi\rangle = \alpha^*\langle \psi \mid \phi\rangle + \beta^*\langle \chi \mid \phi\rangle.$$

Remark 5. Be aware that mathematicians use the opposite convention: For them it is the second entry in an inner product that is anti-linear. Mathematics and physics are two neighboring cultures divided by a common language.

Moreover, the inner product of any vector with itself is positive; it is only zero for the zero vector. Thus

$$||\psi||^2 =< \psi|\psi >$$

can be thought of as the square of the length of a vector.

Remark 6. Strictly speaking states are represented by rays (directions) in Hilbert space. It is a fine point though.

A typical situation is that the state is a complex valued function of some real variable (e.g., position), The inner product is then

$$\langle \psi \mid \psi \rangle = \int \psi^*(x)\phi(x)dx.$$

Exercise 7. Verify that this integral has the properties of an inner product.

Or, the states may be represented by a column vector with complex components $\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}$ etc.

Exercise 8. Prove that

$$\frac{|\langle \psi \mid \phi \rangle|^2}{||\psi||^2||\phi||^2} \leq 1$$

for all non-zero states $|\phi >, |\psi >$. This is called the Cauchy–Schwarz inequality.

2.6.1.2. *If a system is in state* $\mid \phi\rangle$*, the probability of finding it in another state* $\mid \psi\rangle$ *is* $\frac{|\langle \psi||\phi\rangle|^2}{||\psi||^2||\phi||^2}$.

This is one of the confusing things about quantum mechanics, until you get used to it. A classical analogue is the polarization of light. About half of a beam of circularly polarized light will pass through a filter that allows only linearly polarized light.

2.6.1.3. *The observables of a physical system are hermitian linear operators on the states.*

A linear operator (or matrix) acting on a state produces another state, such that

$$L\left(\alpha \mid \psi\rangle + \beta \mid \chi\rangle\right) = \alpha L \mid \psi\rangle + \beta L \mid \chi\rangle.$$

hermitian operators satisfy in addition

$$\langle \psi \mid L \mid \chi\rangle = \langle \chi \mid L \mid \psi\rangle^*.$$

That is, the conjugate-transpose of a matrix elements is itself. If

$$L \mid \psi_\lambda\rangle = \lambda \mid \psi_\lambda\rangle$$

for some complex number λ and non-zero vector $\mid \psi_\lambda\rangle$, we say that $\mid \psi_\lambda\rangle$ is an eigenvector of L with eigenvalue λ. The most important property of a hermitian operator is that it has real eigenvalues. Also, the eigenvectors $\mid \psi_\lambda\rangle$ (we ignore degeneracies for simplicity) form a basis.

That comes in handy because of the following:

2.6.1.4. *The possible outcomes of measuring an observable are its eigenvalues*

But here is a word of warning: **The product of two observables is not always an observable**. The point is that $(AB)^\dagger = B^\dagger A^\dagger$. So, if two hermitian operators do not commute, the product may not be hermitian.

Even if we know the state of a system, we may not be able to predict the outcome of measuring an observable. The best we can do is to give probabilities. With $\lambda, |\psi_\lambda >$ defined as above,

2.6.1.5. *If the system is in some state $\mid \phi\rangle$, the probability of getting the outcome λ upon measuring L is*

$$\frac{|\langle\psi_\lambda \mid \phi\rangle|^2}{||\psi_\lambda||^2||\phi||^2}.$$

Recall that this is always less than one. Also, the fact that eigenvectors form a basis implies that the probabilities add up to one.

2.6.1.6. *There is a hermitian operator called the hamiltonian which represents energy; the time dependence of a state is given by*

$$i\hbar\frac{\partial \mid \psi(t)\rangle}{\partial t} = H \mid \psi(t)\rangle.$$

Thus if you know the state at some time, you can in principle predict what it will be at some later time; if you know the exact hamiltonian and if it is simple enough to make the equation solvable. An eigenstate $\mid \psi_E\rangle$ of the hamiltonian satisfies $H \mid \psi_E\rangle = E \mid \psi_E\rangle$; it is a state of energy E. These have a simple time dependence:

$$\mid \psi_E(t)\rangle = e^{-\frac{i}{\hbar}Et} \mid \psi_E(0)\rangle.$$

2.6.2. *Example: Electron in a Magnetic Field*

As an example, think of an electron in a magnetic field. It is bound to an atom (e.g., Sodium) and we ignore the change in its position: only the rotation of its spin. The wave function has two components. The energy of an electron in a magnetic field is proportional to the dot product of the spin and the magnetic field

$$H = \mu \boldsymbol{\sigma} \cdot \boldsymbol{B} \tag{2.6.1}$$

where

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

are the Pauli matrices.

Exercise. Find the eigenvalues and eigenfunctions of the hamiltonian (2.6.1). If the initial state at time $t = 0$ is $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and the magnetic field is along the $x-$axis $\boldsymbol{B} = (B, 0, 0)$ what is the probability that a measurement of σ_3 at a time t will yield the value -1? This illustrates the phenomenon of oscillation of quantum states, also important for neutrinos.

2.6.3. *Symmetries and Conservation Laws*

2.6.3.1. *Symmetries are represented by unitary operators that commute with the hamiltonian*

Recall that the probability of finding a particle in state $\mid \psi\rangle$ in another state ϕ is $|\langle \phi \mid \psi \rangle|^2$ (assuming that the state vectors are of length one.) If the symmetry is represented by a linear transformation L satisfying

$$\langle L\phi \mid L\psi \rangle = \langle \phi \mid \psi \rangle$$

these probabilities are preserved. Recalling the definition of the hermitian conjugate (adjoint)

$$< L\phi|\psi >=< \phi|L^\dagger \psi >$$

this conditions becomes

$$L^\dagger L = 1$$

That is, L is a **unitary transformation**. Most symmetries are of this type. (See below for an exception.)

Recall that a state of energy E is an eigenstate of the hamiltonian.

$$H\psi = E\psi$$

A symmetry must take it to another state of the same energy:

$$H(L\psi) = E(L\psi).$$

This is satisfied if

$$HL = LH.$$

That is, if the hamiltonian commutes with the symmetry operator. Thus, a symmetry is represented by a unitary operator that commutes with the hamiltonian:

$$L^\dagger L = 1, \quad HL - LH = 0.$$

2.6.3.2. *An exceptional case is time reversal, which is an anti-linear operator*

$$\Theta(a \mid \psi\rangle + b \mid \phi\rangle) = a^*\Theta \mid \psi\rangle + b^*\Theta \mid \phi\rangle.$$

We won't have much more to say about this case for now; we will only consider the case of linear operators for now.

2.6.3.3. *An example is Parity*

It reverses the sign of the co-ordinates of a particle

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

Clearly $P^2 = 1$.

The Schrödinger equation for a free particle is invariant under this transformation

$$-\frac{\hbar^2}{2m}\nabla^2\psi = -i\hbar\frac{\partial\psi}{\partial t}.$$

Another way of seeing that this is a symmetry is that the operator P commutes with the hamiltonian

$$H = -\frac{\hbar^2}{2m}\nabla^2, \quad PH = HP.$$

Thus, if ψ is a state with energy E

$$H\psi = E\psi$$

so will be $P\psi$. Even with a potential parity continues to be a symmetry if

$$V(-\boldsymbol{x}) = V(\boldsymbol{x}).$$

For example consider a particle in one dimension with a potential

$$H = -\frac{\hbar^2}{2m}\nabla^2 + V, \quad V(x) = \lambda(x^2 - a^2)^2, \quad \lambda > 0.$$

There are two minima at $x = \pm a$. The eigenstates of energy can also be simultaneously eigenstates of parity because $[H, P] = 0$. It turns out that the ground state is of even parity

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

while the first excited state is of odd parity

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

2.6.3.4. *Translation invariance leads to conservation of momentum*

The translation by a is represented by the operator

$$T(a)\psi(x) = \psi(x + a).$$

A free particle on a line has hamiltonian

$$H = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V$$

with a constant potential. Thus whether we apply the hamiltonian before or after a translation we get the same effect on a wavefunction:

$$HT(a) = T(a)H.$$

For a particle moving in one dimension, an infinitesimal translation is represented by the derivative operator:

$$\psi(x + a) \approx \psi(x) + a\frac{\partial\psi}{\partial x} + \cdots$$

Thus, if a system is invariant under translation, its hamiltonian must satisfy

$$\left[H, \frac{\partial}{\partial x}\right] = 0.$$

The operator $\frac{\partial}{\partial x}$is anti-Hermitian. The corresponding hermitian operator is $-i\frac{\partial}{\partial x}$. If we multiply by $\hbar$ we get the momentum operator

$$p = -i\hbar\frac{\partial}{\partial x}.$$

Thus, translation invariance implies the conservation of the momentum:

$$[H, p] = 0.$$

Similar arguments apply to each component of momentum of a free particle moving in R^3.

2.6.3.5. *Rotation invariance implies conservation of angular momentum*

The infinitesimal generators of rotation are

$$\mathbf{L} = \mathbf{r} \times \mathbf{p}, \quad \mathbf{p} = -i\hbar \frac{\partial}{\partial \mathbf{r}}.$$

They satisfy the relations

$$[L_1, L_2] = i\hbar L_3, \quad [L_2, L_3] = i\hbar L_1, \quad [L_3, L_1] = i\hbar L_2.$$

2.6.3.6. *A particle can have angular momentum even when its momentum is zero*

Total angular momentum $\mathbf{J}$ is the sum of the orbital angular momentum $\mathbf{L} = \mathbf{r} \times \mathbf{p}$ and an intrinsic angular momentum $\mathbf{S}$:

$$\mathbf{J} = \mathbf{L} + \mathbf{S}.$$

The components of $\boldsymbol{S}$ commute with those of $\mathbf{L}$: They are three matrices $\mathbf{S} = (S_1, S_2, S_3)$ satisfying

$$[S_1, S_2] = i\hbar S_3, \quad [S_2, S_3] = i\hbar S_1, \quad [S_3, S_1] = i\hbar S_2.$$

The simplest choice is $\mathbf{S} = 0$. There are several such "spin zero" particles; e.g., the alpha particle. The next simplest choice is

$$S_1 = \frac{\hbar}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad S_2 = \frac{\hbar}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad S_3 = \frac{\hbar}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

These describe 'spin half' particles, since the maximum eigenvalue of a component of spin is half of $\hbar$. An electron, a proton, a neutron are all examples of such particles.

There are many spin one particles, such as the ρ meson. In this case S_1, S_2, S_3 are 3×3 matrices. The photon has spin one, but is a special case because it moves at the speed of light. (We need relativistic quantum mechanics for this.)

There are a set of particles called Δ that have spin $\frac{3}{2}$. Their spin is represented by 4x4 matrices. There are particles with even higher spin but they tend to be unstable.

We will return to rotations and angular momentum repeatedly. It is the basic example around which the whole theory is built.

Chapter 3

LIE THEORY

Sophus Lie was a Norwegian mathematician who worked at the end of the nineteenth century. His audacious quest (still unfinished) was a "Galois theory" for ordinary differential equations. The established physics of the day (mechanics) was based on such equations. The essential insight of mechanics is that complicated dynamics (e.g., the solar system) can be understood as a succession of infinitesimal steps, each of which is given by a simple formula (i.e., vector field on phase space).

Lie showed that a complicated group of transformations can be built from a knowledge of infinitesimals. The group law reduces to something much simpler, a set of commutation relations (Lie algebra). Lie theory bloomed in the twentieth century, as a branch of mathematics with myriad applications to physics. The most fundamental physics of our day (the standard model of elementary particles) is based on the Lie groups $SU(2)$ and $SU(3)$.

3.1. Lie Algebras

3.1.1. *A Lie algebra is a vector space along with a map* $[.,.] : \mathcal{L} \times \mathcal{L} \to \mathcal{L}$ *such that*

$$[\alpha a + \beta b, c] = \alpha[a, c] + \beta[b, c] \quad \text{linear}$$

$$[a, b] = -[b, a] \quad \text{Anti} - \text{symmetry}$$

$$[[a, b], c] + [[b, c], a] + [[c, a], b] = 0 \quad \text{Jacobi identity}$$

Here, α, β are real numbers. (We mostly think of real Lie algebras. But there is a parallel theory where the scalars α, β are complex.) Note that

the anti-symmetry implies that if $[.,.]$ is linear in the first entry, it will also be linear in the second (i.e., it is bilinear). We call $[.,.]$ the *Lie bracket* or *commutator*.

3.1.1.1. *A homomorphism is a linear map among Lie algebras $f : \mathcal{L} \to \mathcal{L}'$ that preserves the commutator*

$$f([a,b]) = [f(a), f(b)], \quad a, b \in L, \quad f(a), f(b) \in \mathcal{L}'$$

3.1.1.2. *An isomorphism is a homomorphism that is invertible*

Often, it is useful to think of this explicitly as a one-one correspondence of basis vectors that preserves the commutation relations.

3.1.1.3. *An homomorphism to a Lie algebra of matrices is called a representation. A representation is faithful if it is an isomorphism*

3.1.2. ***A subspace $\mathcal{L}' \subseteq \mathcal{L}$ which is closed under the Lie bracket is a sub-algebra***

3.1.3. ***The maximum number of linearly independent elements, whose Lie brackets with each other vanish, is called the rank***

3.1.4. ***Examples***

(1) A basic example is the cross product in three-dimensional Euclidean space. Recall that

$$\mathbf{a} \times \mathbf{b} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}$$

The bilinearity and anti-symmetry are obvious; the Jacobi identity can be verified through tedious calculations. Or you can use the fact that any cross product is determined by the cross product of the basis vectors through linearity, and verify the Jacobi identity on the basis vectors using the cross products

$$\mathbf{i} \times \mathbf{j} = \mathbf{k}, \quad \mathbf{j} \times k = \mathbf{i}, \quad \mathbf{k} \times \mathbf{i} = \mathbf{j}$$

Under many different names, this Lie algebra appears everywhere in physics. It is by far the most important example of a Lie algebra. It has dimension three and rank one.

(2) The commutator of matrices is the other basic example:

$$[A, B] = AB - BA$$

Again, bilinearity and anti-symmetry are obvious. The Jacobi identity follows from a simple calculation using the associativity of matrix multiplication. We see that this is the infinitesimal version of the group $GL(n)$. The dimension is n^2: there are n^2 independent matrix elements. The rank is n: the diagonal matrices commute with each other and there are n linearly independent diagonal matrices.

(3) Various sub-algebras of the algebra of matrices provide the other important examples. The product of anti-symmetric matrices need not be either symmetric or anti-symmetric:

$$(AB)^T = B^T A^T = BA.$$

But the commutator of anti-symmetric matrices is always anti-symmetric:

$$[A, B]^T = (AB - BA)^T = BA - AB = -[A, B].$$

This Lie algebra is the infinitesimal version of the orthogonal group $O(n)$: Recall that an orthogonal matrix that is infinitesimally close to the identity is of the form $1 + A$ with $A^T = -A$. We call this Lie algebra $o(n)$. It has dimension $\frac{n(n-1)}{2}$. The rank of $o(n)$ is k if $n = 2k$ or if $n = 2k + 1$.

(4) Similarly, the commutator of anti-Hermitian matrices is anti-Hermitian. This Lie algebra $u(n)$ is the infinitesimal version of the group of unitary matrices $U(n)$. The dimension of $u(n)$ is n^2 and its rank is n; as for $su(n)$, the dimension is $n^2 - 1$ and the rank is $n - 1$.

(5) The trace of a commutator is zero. (Prove this!) Thus, we have the Lie algebra of traceless anti-Hermitian matrices $su(n)$ which is the infinitesimal version of the group $SU(n)$ of unitary matrices of determinant one. Recall that if a matrix is infinitesimally close to one, $\det(1 + A) \approx 1 + \mathrm{tr}A$.

(6) The Lie algebra $o(3)$ is in fact the same as (is isomorphic to) the cross product on three-dimensional vectors. Any anti-symmetric matrix can

be written as

$$A = \begin{pmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{pmatrix} = a_1 S_1 + a_2 S_2 + a_3 S_3$$

for some vector $\boldsymbol{a} \in \mathbb{R}^3$. The matrices S_i form a basis for $o(3)$:

$$S_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad S_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad S_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

The commutation relations

$$[S_1, S_2] = S_3, \quad [S_2, S_3] = S_1 \quad [S_3, S_1] = S_2$$

are isomorphic to those above under the correspondence $\mathbf{i} \mapsto S_1, \mathbf{j} \mapsto S_2, \mathbf{k} \mapsto S_3$. (i.e., the cross product and the commutator relations among the basis vectors are the same under this correspondence.)

(7) Moreover, the Lie algebra $su(2)$ is isomorphic to $o(3)$. Any traceless anti-Hermitian matrix can be written as a linear combination of Pauli matrices

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

The correspondence $S_3 \mapsto -\frac{i}{2}\sigma_3, \quad S_1 \mapsto -\frac{i}{2}\sigma_1, \quad S_2 \mapsto -\frac{i}{2}\sigma_2$ gives an isomorphism. This is fundamental to understanding the spin of an electron.

(8) The Poisson bracket of classical mechanics was the first example of a Lie algebra. Recall that observables of classical mechanics are functions of positions and momenta. For a single degree of freedom (for simplicity), the Poisson bracket is defined as

$$\{A, B\} = \frac{\partial A}{\partial q}\frac{\partial B}{\partial p} - \frac{\partial A}{\partial p}\frac{\partial B}{\partial q}$$

Verifying the Jacobi identity for this is a good way to start an honest day of work. For more than one degree of freedom, we sum over each pair of conjugate variables:

$$\{A, B\} = \sum_i \left(\frac{\partial A}{\partial q_i}\frac{\partial B}{\partial p_i} - \frac{\partial A}{\partial p_i}\frac{\partial B}{\partial q_i} \right)$$

(9) In particular, we have the canonical commutation relations (also called the Heisenberg algebra)

$$\{q, p\} = c = -\{p, q\}, \quad \{p, c\} = \{q, c\} = 0.$$

Here, c is the constant function, equal to 1 everywhere. This is an example of a *nilpotent Lie algebra*: Repeated commutators vanish eventually. In this case, double commutators vanish.

(10) The Poisson brackets of the components of angular momentum provide yet another physically important realization of the Lie algebra $o(3)$

$$\mathbf{L} = \mathbf{r} \times \mathbf{p}$$

$$\{L_1, L_2\} = L_3, \quad \{L_2, L_3\} = L_1, \quad \{L_3, L_1\} = L_2$$

This isomorphism arises because the canonical transformations generated by angular momentum are rotations. We can regard the earlier examples in terms of matrices as representations of this Lie algebra of the angular momentum components.

(11) A Lie algebra that is commutative is trivial: The bracket must vanish. Thus, to be interesting, a Lie algebra must be non-abelian.

(12) The only Lie algebra of dimension one is the trivial algebra.

(13) The only non-abelian Lie algebra of dimension two can be written as

$$[e_0, e_+] = e_+$$

by a choice of basis. (By the way, e_+ is called that because it is analogous to the "raising" or "creation operator" of quantum mechanics.)

(14) **Exercise:** Find a representation for it in terms of 2×2 matrices.
Answer: $e_0 \mapsto \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{pmatrix}$, $e_+ \mapsto \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$

(15) Another three-dimensional Lie algebra, which is not isomorphic to $o(3)$ or $su(2)$, is called $sl(2, R)$:

$$[e_1, e_2] = -e_3, \quad [e_2, e_3] = -e_1, \quad [e_3, e_1] = e_2$$

The sign of the first two commutators is different from $o(3)$.

(16) **Exercise:** Find an isomorphism of $sl(2, R)$ with the space of traceless real 2×2 matrices.

(17) There are more non-trivial three-dimensional Lie algebras, such as $so(3) \approx su(2)$, $sl(2, R)$, and Heisenberg.

(18) Also, find a set of three functions of position and momentum with Poisson brackets isomorphic to $sl(2, R)$.
Hint: Think of $\frac{p^2-q^2}{4}$, $\frac{p^2+q^2}{4}$, $\frac{pq}{2}$.

(19) In addition to matrix algebras such as $su(n), so(n)$, there is also a finite sequence of exceptional Lie algebras. Many physicists have tried hard to explain elementary particles in terms of exceptional Lie algebras, seduced by their mathematical beauty. So far, no luck.

3.2. Lie Groups

3.2.1. *A Lie group is a group on which there is a co-ordinate system such that the multiplication and inverse are differentiable functions*

In other words, a Lie group is a manifold along with a multiplication and inverse which are differentiable functions.

If you don't know what a manifold is, don't worry about this. Lie himself thought of Lie group as transformations which depend on some parameters in a smooth way. Any abstract mathematical theory is best understood by working out physically realized examples. The axiomatization always comes later. Its main purpose is to serve as a firm foundation for the next level of abstraction. By the end of this book, you will know many examples of Lie groups.

3.2.1.1. *A countable group like the set of integers, or the set of of rationals, or a finite group like the permutation groups, are not Lie: There is no way to differentiate group elements*

But the idea of a Lie algebra makes sense even when the scalars underlying the vector space form a finite or countable ring (like integers). There is a way to construct a Lie algebra from a discrete group, using its "central series". The power of a mathematical idea can be measured by its utility in areas far from its origins. Lie's ideas on groups and algebras have applications in practically every branch of mathematics and physics. They are among the most powerful of mathematical ideas.

3.2.1.2. *$GL(n, R)$ is a Lie group*

The matrix elements themselves provide a co-ordinate system. Just stay away from matrices of zero determinant. (The condition $\det g \neq 0$ leaves behind an open neighborhood of R^{n^2}.)

3.2.1.3. *$U(n)$ is a Lie group*

We have to solve the constraints defining the group

$$g^\dagger = g^{-1}$$

The matrix elements themselves are no longer a co-ordinate system: We need to solve these rather complicated nonlinear equations. The substitution

$$g = e^a, \quad a^\dagger = -a$$

allows us to solve them in a neighborhood of the identity.

Aside on Exponential Co-ordinates. The exponential of a matrix is defined by an infinite series in the same way as the exponential of a number

$$e^a = 1 + a + \frac{a^2}{2!} + \frac{a^3}{3!} + \cdots$$

It satisfies the conditions

$$(e^a)^\dagger = e^{a^\dagger}, \quad (e^a)^{-1} = e^{-a}.$$

The tricky part in using these "exponential co-ordinates" is that

$$e^a e^b \neq e^{a+b}$$

unless $[a, b] = 0$. There is a much more complicated formula that replaces this. (We see it soon.)

If $a^\dagger = -a$, then e^a is a unitary matrix. The matrix elements of the anti-Hermitian matrix provide a co-ordinate system on $U(n)$ in the neighborhood of the identity. More precisely, define the norm of an anti-Hermitian matrix by $||a|| = \sqrt{\mathrm{tr} a^\dagger a}$. As long as $||a|| < \pi$, the exponential function is injective (i.e., e^a completely defines a within the disc $||a|| < \pi$). This establishes a co-ordinate system around the origin. Next, we establish a co-ordinate system around the roots of unity by setting $g = e^{\frac{2\pi i}{n}k} e^b$, for $k = 0, 1, \cdots n-1$ again with $||b|| < \pi$. It is possible to show (we omit the details of the construction and the proof) these n co-ordinate systems cover all over $U(n)$:

Every unitary matrix can be written in the form $g = e^{\frac{2\pi i}{n}k} e^b$ for some $k = 0, 1 \cdots n-1$ and $b^\dagger = -b$ with $|b| < \pi$. This formula can be thought of as providing a co-ordinate system in $U(n)$ in a neighborhood of $e^{\frac{2\pi i}{n}k}$.

Of course, most parts of $U(n)$ are covered by more than one of these co-ordinate systems: The change of variables from one system to the other is differentiable. This is similar to the way that a polar co-ordinate system cannot cover all of the plane: The origin and the line $\theta = 0$ have to be excluded. But two polar systems with different centers and axes can cover all of the plane; in regions covered by both systems, we can differentially change variables among them.

If $a^\dagger = -a$ and $\mathrm{tr}a = 0$, then $e^a \in SU(n)$. The point is that $\det e^a = e^{\mathrm{tr}a}$. This identity is obvious for matrices that can be diagonalized. (Prove it!) More generally, it follows by continuity as the determinant, trace, and exponential are all continuous functions; matrices that cannot be diagonalized can be perturbed infinitesimally and made diagonalizable. This makes $SU(n)$ into a Lie group by similar arguments.

If $a^T = -a$, and a has real elements, then $e^a \in SO(n)$. Recall that anti-symmetric matrices have zero trace. Hence, $\det e^a = e^{\mathrm{tr}a} = 1$. It is not possible to express parity as the exponential of an anti-symmetric matrix.

Lie groups don't always have to be thought of in terms of matrices. As long as the multiplication law can be written in terms of some co-ordinates, we can verify associativity and look for an inverse.

Exercise 10. Let $G = \{(a, x) \mid a > 0, x \in \mathbb{R}\}$ be the half-plane. Define the product $(a, x)(b, y) = (ab, ay + x)$. Show that this is a group. (What is the identity? What is $(a, x)^{-1}$?) Find a representation for this group in terms of 2×2 matrices. Calculate the commutator $ghg^{-1}h^{-1}$ for two arbitrary elements of this group.

3.3. From Lie Groups to Lie Algebras

3.3.1. *Every Lie group determines a Lie algebra.*

More than one Lie group might lead to the same Lie algebra. For example, we see later that $SU(2)$ and $SO(3)$ yield the same Lie algebra, although they are not isomorphic as groups.

3.3.2. *The set of elements infinitesimally close to the identity in a Lie group form a Lie algebra.*

For matrix groups like $GL(n, R), SU(n), SO(n)$ above, we put

$$g = e^{\epsilon a}, \quad h = e^{\epsilon b}$$

where ϵ is thought as small. Then, we expand in powers of ϵ:

$$e^{\epsilon a} \approx 1 + \epsilon a + \epsilon^2 \frac{a^2}{2} + O(a^3)$$

$$g^{-1} = e^{-\epsilon a} \approx 1 - \epsilon a + \epsilon^2 \frac{a^2}{2} + O(a^3)$$

$$gh \approx 1 + \epsilon(a+b) + \epsilon^2 \frac{a^2 + 2ab + b^2}{2} + O(a^3, b^3)$$

$$g^{-1}h^{-1} = 1 - \epsilon(a+b) + \epsilon^2 \frac{a^2 + 2ab + b^2}{2} + O(a^3, b^3)$$

$$ghg^{-1}h^{-1} = 1 + \epsilon^2[a,b] + O(a^3, b^3)$$

(Calculate each line out and verify this.) Thus, the lack of commutativity of group multiplication taken to second order determines the commutator. This commutator defines a Lie algebra associated to the Lie group.

Even if the group is not built out of matrices, we can calculate the "group commutator" $ghg^{-1}h^{-1}$ in some co-ordinate system centered at the identity. Then, expand it to the leading (i.e., second) order to extract the Lie bracket.

Exercise 11. This continues Exercise (10) Find $ghg^{-1}h^{-1}$ to second order in ϵ, where $g = (1 + \epsilon\alpha, \epsilon\xi)$ and $h = (1 + \epsilon\beta, \epsilon\eta)$. Use this to find the Lie algebra of the group. Find a representation in terms of 2×2 matrices.

Solution: The identity element is $(1, 0)$. The inverse is $(a, x)^{-1} = (a^{-1}, -a^{-1}x)$ and the commutator $gg'g^{-1}g'^{-1} = (1, ax - a'^2x - x' + a^2a'x')$. A representation is given by

$$(a, x) \mapsto \begin{pmatrix} a & x \\ 0 & 1 \end{pmatrix}$$

since

$$\begin{pmatrix} a & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} b & y \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} ab & ay + x \\ 0 & 1 \end{pmatrix}$$

A calculation shows that

$$ghg^{-1}h^{-1} = (1, 0) + \epsilon^2 (0, \alpha\eta - \beta\xi) + \mathrm{O}(\epsilon^3)$$

This is the Lie algebra

$$[(\alpha, \xi), (\beta, \eta)] = (0, \alpha\eta - \beta\xi)$$

Defining[1]

$$e_0 = (1,0), e_+ = (0,1)$$

this corresponds to the Lie algebra (we saw it as an example in the section on Lie algebras)

$$[e_0, e_+] = e_+.$$

A representation of the Lie algebra is the infinitesimal version of the group representation:

$$e_0 \mapsto \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad e_+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

This Lie group and its Lie algebra do not seem to have a standard name (may be it should).

3.3.3. *The Lie algebra of $U(n)$ is $u(n)$, the set of anti-Hermitian matrices; that of $SU(n)$ is $su(n)$, the traceless anti-Hermitian matrices.*

3.3.4. *The Lie algebra of $SO(n)$ is $so(n)$ the set of anti-symmetric matrices.*

3.3.5. *Structure constants*

It is very useful to think of a Lie algebra in terms of a basis X_i. An element is expanded as a linear combination $u = u^i X_i$. (We sum over repeated indices as often in geometry and algebra). If you know the commutator of every pair of basis vectors, you can calculate the commutator of any element of the Lie algebra. There must be a set of numbers called "structure constants" such that

$$[X_i, X_j] = c_{ij}^k X_k$$

To be more precise, c_{ij}^k are the components of a tensor in this basis. The axioms of a Lie algebra become some identities satisfied by the structure constants:

- $c_{ij}^k = -c_{ji}^k$ anti-symmetry
- $c_{ij}^l c_{lk}^m + c_{jk}^l c_{li}^m + c_{ki}^l c_{lj}^m = 0$ Jacobi identity

[1] Beware that $(1,0)$ as an element of the Lie algebra means something different from $(1,0)$ as an element of the group. Hopefully, the meaning is clear from the context.

Most physicists think of Lie algebras in terms of their commutation relations of basis elements and the structure constants. Beware that "basis elements" are called "generators" in the physics literature. "Generators" have a different meaning in mathematics; remember that σ and τ are generators of the permutation group S_3.

Exercise 12. Find a basis and the structure constants for the Lie algebra in Exercise (11).

Solution: We saw that $[e_0, e_+] = e_+$, so the non-zero structure constants are

$$c^+_{0+} = 1 = -c^+_{+0}$$

3.4. From Lie Algebras to Lie Groups*

This section can be omitted in a first reading. It will make more sense later.

The passage from Lie groups to Lie algebras is a kind of differentiation. The converse is a kind of non-commutative integration: You should expect this to be much harder. The following material is just a guide to those who want to venture further. The book by Hausner and Schwartz [3] *Lie Groups and Lie Algebras* is quite good for this topic. A purely algebraic proof of the BCH lemma is in *Free Lie Algebras* by Reuttenauer [4]. We do not need the fearsome details of the formulas for most purposes: Only the first one or two orders are sufficient usually. The purpose of this section is to show that it is possible to get the group law from the Lie algebra.

3.4.1. *The Lie bracket completely determines the group multiplication*

In the exponential co-ordinate system, the multiplication of group elements follows from taking repeated commutators (in the corresponding Lie algebra) and adding them up in a particular way. The key is a formula that allows us to multiply the exponentials of matrices that do not commute.

3.4.2. *The Baker–Campbell–Hausdorff formula*

$$e^a e^b = e^{a + \int_0^1 dt \psi\left(e^{\hat{a}} e^{t\hat{b}}\right) b} \tag{3.1}$$

Here,

$$\psi(x) = \frac{x \log x}{x-1}$$

Also, $\hat{a}$ is the linear operator defined by the commutator:

$$\hat{a}b \equiv [a, b].$$

This is also called ad$a$ in mathematics books. So, $\hat{a}^2 b \equiv [a, [a, b]]$,$\hat{a}^3 b = [a, [a, [a, b]]]$, etc. Indeed,

$$e^{\hat{a}} b = b + [a, b] + \frac{1}{2!}[a, [a, b]] + \frac{1}{3!}[a, [a, [a, b]]] + \cdots$$

The function ψ is closely related to the generating function of Bernoulli numbers:

$$\begin{aligned}\psi(e^x) = \frac{xe^x}{e^x - 1} &= \sum_{n=0}^{\infty} \frac{B_n}{n!} x^n \\ &= 1 + \frac{x}{2} + \frac{x^2}{12} - \frac{x^4}{720} + \frac{x^6}{30240} + O(x^8).\end{aligned}$$

By expanding ψ and exp in power series, (3.1) becomes an explicit formula for multiplication of exponentials.

The first few terms are

$$e^a e^b = e^{a+b+\frac{1}{2}[a,b]+\frac{1}{12}([a,[a,b]]+[b,[b,a])-\frac{1}{24}[b,[a,[a,b]]]+\cdots}$$

To prove this, we need a series of intermediate results. We start with the following:

Lemma 13.

$$e^a b e^{-a} \equiv e^{\hat{a}} b$$

Proof. Let $b(t) = e^{ta} b e^{-ta}$,$b(0) = b$. Then,

$$\begin{aligned}b(t+\epsilon) &= e^{(t+\epsilon)a} b e^{-(t+\epsilon)a} = e^{\epsilon a} e^{ta} b e^{-ta} e^{-\epsilon a} \\ &\approx (1 + \epsilon a) b(t) (1 - \epsilon a) \approx b(t) + \epsilon [a, b(t)]\end{aligned}$$

up to terms second order in ϵ. Thus,

$$\frac{d}{dt} b(t) = [a, b(t)] = \hat{a} b(t).$$

Regarding $\hat{a}$ as a linear operator on b, the solution is

$$b(t) = e^{t\hat{a}}b.$$

□

We can now get a formula to differentiate the exponential of a function valued in a Lie algebra. This is very useful in many quantum mechanic calculations as well.

Lemma 14. *Let $a(t)$ be a function of a real variable, valued in the Lie algebra. Then, with $\phi(z) = \frac{e^z-1}{z} = 1 + \frac{1}{2!}z + \frac{1}{3!}z^2 + \cdots$,*

$$e^{-a(t)}\frac{d}{dt}e^{a(t)} = \phi(-\hat{a}(t))\frac{da(t)}{dt}.$$

Proof. Define $g(s,t) = e^{sa(t)}$. Then

$$A(s,t) \equiv g^{-1}\frac{\partial g}{\partial s} = a(t)$$

by the definition of the exponential. Define

$$B(s,t) \equiv g^{-1}\frac{\partial g}{\partial t} = e^{-sa(t)}\frac{\partial e^{sa(t)}}{\partial t}$$

We can verify the identity

$$\frac{\partial B}{\partial s} - \frac{\partial A}{\partial t} + [A,B] = 0$$

which now becomes

$$\frac{\partial B}{\partial s} - \frac{\partial a(t)}{\partial t} + [a(t),B] = 0$$

or

$$\frac{\partial B}{\partial s} = -\widehat{a(t)}B + \dot{a}, \quad B(0,t) = 0$$

The dot denotes differentiation w.r.t. t.

We can think of t as a constant and solve this as a power series in s:

$$B(s,t) = s\dot{a} + \frac{s^2}{2!}(-\widehat{a(t)})\dot{a} + \cdots \frac{s^n}{n!}\left(-\widehat{a(t)}\right)^{n-1}\dot{a} + \cdots$$

Putting $s = 1$ in this we get the result we want. □

Lemma 15. *Let $e^a e^{tb} = e^{c(t)}$. Then,*

$$e^{-c(t)} \frac{d}{dt} e^{c(t)} = b$$

Proof. Just calculate

$$\frac{d}{dt} e^{c(t)} = e^a \frac{d}{dt} e^{tb} = e^a e^{tb} b = e^{c(t)} b. \qquad \square$$

Now, we can prove the BCH formula (3.1):

Proof. Using the lemmas above,

$$\phi(-\widehat{c(t)})\dot{c} = b$$

Now, the function $\psi(z) = \frac{z \log z}{z-1}$ satisfies

$$\psi(z)\phi(-\log z) = 1$$

so that

$$\frac{dc}{dt} = \psi(\widehat{e^{c(t)}})b$$

If we integrate this differential equation (recall the boundary condition $c(0) = a$) and evaluate it at $t = 1$, we get the result claimed. $\square$

All this leads up to a fundamental idea of Lie theory: The product of exponentials is determined by a series of repeated commutators. The exponential defines a co-ordinate system in the neighborhood of the identity of the Lie group; the BCH formula gives the product in this co-ordinate system. The co-ordinates may break down far away from the identity, but we can establish additional co-ordinate charts based at a countable number of other points on the group. (To fully understand this, you have to know differential geometry beyond the scope of this book.) The most important physical application is to spin which we discuss in detail later. We explain the topology of $SU(2)$ and $SO(3)$ at that time. We merely state the result.

3.4.3. *A Lie algebra determines a unique connected, simply connected Lie group. Every connected Lie group with this Lie algebra is a quotient of this simply connected Lie group by a countable abelian normal sub-group.*

Exercise 13. Starting with the Lie algebra commutation relations, $[e_0, e_+] = e_+$, reconstruct the group multiplication law for the two-dimensional Lie group.

Chapter 4

ROTATIONS: SO(3) AND SU(2)

Our first lessons in geometry are on the plane. Euclid's treatise on plane geometry remains the model of all later works on geometry. Let us also begin by understanding rotations in the plane. Then we will pass to "solid geometry", the geometry of three-dimensional Euclidean space.

4.1. SO(2)

Recall the relation between polar and cartesian co-ordinates on the plane

$$x = r\cos\theta, \quad y = r\sin\theta$$

If we rotate this point through an angle ϕ around the origin, r remains unchanged. But the angular co-ordinate changes to $\theta + \phi$, giving a new point

$$x' = r\cos(\theta+\phi), \quad y' = r\sin(\theta+\phi)$$

Using the addition formula for sin and cos

$$x' = r\cos\theta\cos\phi - r\sin\theta\sin\phi, \quad y' = r\cos\theta\sin\phi + r\sin\theta\cos\phi$$

We can write this as

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

In other words, the effect of a rotation can be represented as multiplication by a 2×2 matrix. This matrix

$$R(\phi) = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}$$

has determinant one

$$\det \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} = \cos^2\phi + \sin^2\phi = 1$$

and is orthogonal:

$$R(\phi)R(\phi)^T = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix}$$

$$= \begin{pmatrix} \cos^2\phi + \sin^2\phi & \cos\phi\sin\phi - \cos\phi\sin\phi \\ \cos\phi\sin\phi - \cos\phi\sin\phi & \cos^2\phi + \sin^2\phi \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

The trig identities imply that

$$R(\phi')R(\phi) = R(\phi + \phi')$$

That is, the effect of a rotation through ϕ and then another rotation through ϕ' is a rotation through $\phi + \phi'$. We must get used to this idea that rotations are represented by orthogonal matrices of determinant one.

4.1.1. *Meaning of orthogonality*

Let us go in the other direction. Let $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ be an orthogonal matrix. What does it imply for its components?

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} a & c \\ b & d \end{pmatrix} = \begin{pmatrix} a^2 + b^2 & ac + bd \\ ca + db & c^2 + d^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

For this to be equal to one, the columns $\begin{pmatrix} a \\ b \end{pmatrix}$ and $\begin{pmatrix} c \\ d \end{pmatrix}$ must be unit vectors (using the diagonal entries). The off diagonal entries being zero implies that the dot product of the column vectors is zero: They are *orthogonal* to each other. (Hence, the name). Any unit vector is of the form

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \cos\phi \\ \sin\phi \end{pmatrix}$$

for some angle ϕ. The other column will be given by some other angle

$$\begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} \cos\psi \\ \sin\psi \end{pmatrix}$$

To be orthogonal we must have

$$(a\ b) \begin{pmatrix} c \\ d \end{pmatrix} = 0 = \cos\phi\cos\psi + \sin\phi\sin\psi$$

That is

$$\cos(\phi - \psi) = 0$$

there are two solutions to this

$$\phi - \psi = \pm\frac{\pi}{2}$$

Thus, the most general orthogonal matrix is of the form

$$\begin{pmatrix} \cos\phi & \cos\left(\pm\frac{\pi}{2} + \phi\right) \\ \sin\phi & \sin\left(\pm\frac{\pi}{2} + \phi\right) \end{pmatrix} = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}, \quad \text{or} \begin{pmatrix} \cos\phi & \sin\phi \\ \sin\phi & -\cos\phi \end{pmatrix}$$

One of these is the rotation we found earlier. The other solution has determinant minus one:

$$\det \begin{pmatrix} \cos\phi & \sin\phi \\ \sin\phi & -\cos\phi \end{pmatrix} = -\cos^2\phi - \sin^2\phi = -1$$

It is the product of a rotation and a reflection:

$$\begin{pmatrix} \cos\phi & \sin\phi \\ \sin\phi & -\cos\phi \end{pmatrix} = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

The matrix $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ reverses the sign of the $y-$co-ordinate: It is a reflection around the $x-$axis.

In conclusion,

- Orthogonal matrices of determinant one are rotations.
- Orthogonal matrices of determinant minus one are a product of a reflection and a rotation.

Note also that rotations in the plane commute, since $R(\phi')R(\phi)$ depends only on $\phi + \phi'$.

4.2. SO(3)

Much of what we said above generalizes to three dimensions. The main difference is that rotations do not always commute.

The square of the length of a vector $\boldsymbol{r} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ can be written as

$$\boldsymbol{r}^T \boldsymbol{r} = x^2 + y^2 + z^2$$

What linear transformations $\mathbf{r} \mapsto M\mathbf{r}$ will leave the length of a vector unchanged?

$$(M\boldsymbol{r})^T (M\boldsymbol{r}) = \boldsymbol{r}^T M^T M\boldsymbol{r} = \boldsymbol{r}^T \boldsymbol{r}$$

This is satisfied for all x, y, z iff

$$M^T M = 1$$

That is, the matrix must be orthogonal. Again, this means that the columns of M are orthogonal to each other; and that each of them have length one. In terms of indices

$$\sum_j M_{ji} M_{jk} = \delta_{ik}$$

Expanding it out with $i = 1, k = 2$ for example

$$M_{11}M_{12} + M_{21}M_{22} + M_{31}M_{32} = 0$$

This is the condition that $\begin{pmatrix} M_{11} \\ M_{21} \\ M_{31} \end{pmatrix}$ is orthogonal to $\begin{pmatrix} M_{12} \\ M_{22} \\ M_{32} \end{pmatrix}$. Similarly for the other choices of i, k.

Lemma. *The determinant of an orthogonal matrix is either* 1 *or* -1.

Proof. Take the determinant of the condition above and use $\det M = \det M^T$:

$$(\det M)^2 = 1 \implies \det M = \pm 1$$

□

Under a continuous change of the matrix elements, the determinant of an orthogonal matrix (which is a continuous function of the matrix elements, indeed a polynomial) cannot change: It can only jump from one value to the other. We say that the group of orthogonal matrices splits into two "connected components". One with determinant one and the other with determinant -1.

The orthogonal matrices of determinant 1 form a subgroup: The product still has determinant one. It also contains the identity. This is the group $SO(3)$. It is simply the set of all rotations. For, any rotation is determined by an angle and an axis of rotation. If we choose the third axis to lie along the axis, the rotation will only change the first two axes. With this choice of co-ordinate system, a rotation will be

$$R_3(\phi_3) = \begin{pmatrix} \cos\phi_3 & -\sin\phi_3 & 0 \\ \sin\phi_3 & \cos\phi_3 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

This is an orthogonal matrix, as we saw last time; the extra row and column does not change this fact. It also has determinant one. A rotation around the other two axes will look like (by cyclically permuting co-ordinate axes)

$$R_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\phi_1 & -\sin\phi_1 \\ 0 & \sin\phi_1 & \cos\phi_1 \end{pmatrix}, \quad R_2(\phi_2) = \begin{pmatrix} \cos\phi_2 & 0 & \sin\phi_2 \\ 0 & 1 & 0 \\ -\sin\phi_2 & 0 & \cos\phi_2 \end{pmatrix},$$

They also have determinant one.

The most general rotation is determined by three independent angles: Two to determine the axis of rotation (which is a unit vector) and another to determine the angle of rotation. Alternatively, a general rotation is the product of the three elementary rotations above. There are also other ways of parametrizing them. The most popular are the Euler angles, which we will see later on.

The main point for now is that $SO(3)$ is a three-dimensional Lie group; i.e., as a manifold it is of dimension three. Which manifold is it? Turns out to be $\mathbb{RP}^3$. If that means nothing to you yet, that is OK. We will return to this also in more detail later.

4.2.1. *Parity*

It is conventional in particle physics to define the operation of parity as the reflection of all three co-ordinates:

$$P : \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -x \\ -y \\ -z \end{pmatrix}.$$

As a matrix, $P = -1$, the negative of the identity matrix. Clearly, it is orthogonal, $P^T P = 1$ and has determinant minus one. Any orthogonal matrix of determinant minus one is a product of parity and a rotation.

Exercise. Show that a left handed glove can be turned into a right handed one by turning it inside out. (Best done with a latex glove; a good trick at a party of geeks). Explain why.

It is possible to split the group of symmetries $O(3)$ of euclidean space as a product of rotations and the cyclic group Z_2 generated by parity.

Exercise. Let $SO(3) \times Z_2 = \{(R, \epsilon) | R \in SO(3), \epsilon = \pm 1\}$. This is a group under pairwise multiplication. Show that

$$h : O(3) \to SO(3) \times Z_2, \quad h : M \mapsto (M \det M, \det M)$$

is an isomorphism of $O(3)$ with $SO(3) \times Z_2$.

There is ample evidence that rotations are a symmetry of nature even at the fundamental level of elementary particles. A consequence of rotation invariance is conservation of angular momentum, which is verified every day in particle accelerators.

So, everyone thought that Parity is a symmetry as well. A big surprise was when it was discovered (in 1957) that reflections are not always symmetries!.[1] The culprit is an elusive particle known as a neutrino.

We will see that there is a natural, but subtle, way that particles of spin $\frac{1}{2}$ (such as neutrinos) can violate reflection symmetry. That nature takes advantage of this subtle possibility is one of the remarkable things about the standard model of elementary particles. To understand this we will have to dig deep into the structure of the rotation group and its Lie algebra.

That said, parity violation remains a tiny effect. The vast majority of physical systems preserve parity: Electromagnetic interactions (which govern chemistry), strong interactions which govern nuclear reactions and gravity all preserve parity. Only weak interactions (many of which involve the neutrinos) violate parity. So, we will see how to implement parity in quantum systems.

4.2.2. *The Lie Algebra o*(3)

If the angles are infinitesimally small, the elementary rotations around the axes become

$$R_1(\phi_1) \approx 1 + \phi_1 S_1 + \cdots, \quad R_2(\phi_2) \approx 1 + \phi_2 S_2 + \cdots \quad R_3(\phi_3) \approx 1 + \phi_3 S_3 + \cdots$$

where

$$S_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad S_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad S_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

These are anti-symmetric matrices such that

$$R_1(\phi_1) = e^{\phi_1 S_1}, \quad R_2(\phi_2) = e^{\phi_2 S_2}, \quad R_3(\phi_3) = e^{\phi_3 S_3}$$

This might be a good time to learn some basic facts about matrix exponentials.

Proposition. *The exponential of an anti-symmetric matrix is orthogonal*

[1] Even before this discovery, it was known that some chemical reactions in living things are not invariant under reflections. There are some molecules which come in different shapes which are interchanged by reflections: stereo-isomers. (Sucrose, common sugar, is an example). But this is not a violation of reflection symmetry at a fundamental physical level. By accident, early life forms used one of the possible orientations and we have inherited that preference; chemically, they have identical properties.

Just note that $\left[e^A\right]^T = e^{A^T} = e^{-A}$. Since A commutes with $-A$, we have $\left[e^A\right]^T e^A = e^{-A}e^A = e^{-A+A} = 1$

Proposition. $\det e^A = e^{\mathrm{tr}A}$

Proof. This is obvious if the matrix is diagonal:

$$A = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ . & . & . & . \\ 0 & 0 & . & \lambda_n \end{pmatrix} \Longrightarrow e^A = \begin{pmatrix} e^{\lambda_1} & 0 & . & 0 \\ 0 & e^{\lambda_2} & . & 0 \\ . & . & . & . \\ 0 & 0 & . & e^{\lambda_n} \end{pmatrix}$$

$$\det e^A = e^{\lambda_1} e^{\lambda_2} \cdots e^{\lambda_n} = e^{\lambda_1+\lambda_2\cdots\lambda_n} = e^{\mathrm{tr}\ A}$$

The determinant and trace are invariant under equivalence transformations,

$$\det\left\{ S \begin{pmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ . & . & . & . \\ 0 & 0 & . & \lambda_n \end{pmatrix} S^{-1} \right\} = e^{\lambda_1} e^{\lambda_2} \ldots e^{\lambda_n},$$

$$\mathrm{tr}\left\{ S \begin{pmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ . & . & . & . \\ 0 & 0 & . & \lambda_n \end{pmatrix} S^{-1} \right\} = \lambda_1 + \lambda_2 + \cdots + \lambda_n$$

for any invertible matrix S. This extends the proof to any matrix that can be diagonalized. That is, as long as there is an invertible (possibly complex) matrix such that

$$A = S \begin{pmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ . & . & . & . \\ 0 & 0 & . & \lambda_n \end{pmatrix} S^{-1}$$

the identity above is true. But then, the set of matrices that can be diagonalized is dense in the vector space of square matrices: As long as the characteristic equation $\det[A - z1] = 0$ has distinct roots, the matrix is diagonalizable. (The exception is when the discriminant of the characteristic equation vanishes. This is a subset of co-dimension one, over complex numbers.) Since the determinant, trace and exponential are continuous functions, once the identity holds on a dense subset it will hold everywhere. □

Conversely, any rotation $R \in SO(3)$ can be written as

$$R = e^A$$

for some anti-symmetric matrix. (Beware that there may be more than one such matrix. For example, $e^{2\pi S_1} = 1$). It is not difficult to see that there is a dense subset of $SO(3)$ of this form. That in fact *every* rotation can be expressed this way takes some additional work,which does not add much to physical insight. So, we skip the proof.

Proposition. *The matrices S_1, S_2, S_3 above form a basis in the vector space of 3×3 anti-symmetric matrices*

We can simply write any anti-symmetric as a linear combination of the S_i in a unique way:

$$\begin{pmatrix} 0 & A_{12} & A_{13} \\ -A_{12} & 0 & A_{23} \\ -A_{13} & -A_{23} & 0 \end{pmatrix} = -A_{23}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} + A_{13}\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix} - A_{12}\begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Proposition. *The vector space of 3×3 anti-symmetric matrices is a Lie algebra. The commutation relations of the basis elements are*

$$[S_1, S_2] = S_3, \quad [S_2, S_3] = S_1 \quad [S_3, S_1] = S_2 \tag{4.2.1}$$

You can verify the commutators by calculating the matrix products directly.

These relations can also be written as

$$[S_j, S_k] = \epsilon_{jkl} S_l \tag{4.2.2}$$

The quantity ϵ_{jkl} (called the Levi–Civita tensor) is completely anti-symmetric in its indices:

$$\epsilon_{jkl} = -\epsilon_{kjl} = -\epsilon_{klk}$$

So, it vanishes if any pair of indices are equal. It is completely specified by the case where the indices are all different and ordered in the standard way, in which case it is defined to be unity.

$$\epsilon_{123} = 1$$

You can check that (4.2.1) and (4.2.2) say the same thing. Thus, the structure constants of the Lie algebra $o(3)$ are the components of the Levi–Civita tensor.

Exercise. Show by direct calculation that $(\theta_1 S_1 + \theta_2 S_2 + \theta_3 S_3)\, \boldsymbol{r} = \boldsymbol{\theta} \times \boldsymbol{r}$ where $\boldsymbol{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix}$ and $\boldsymbol{r} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$. Also, $\boldsymbol{\theta} \times \boldsymbol{r}$ is the change in $\boldsymbol{r}$ due to an infinitesimal rotation around an axis parallel to $\boldsymbol{\theta}$.

We can summarize the relation of the Lie group $SO(3)$ and Lie algebra $o(3)$:

Proposition. *There is a neighborhood of the identity of $SO(3)$ in which $R = e^{\theta_1 S_1 + \theta_2 S_3 + \theta_3 S_3}$ for some vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$ of length $|\boldsymbol{\theta}| < \pi$. The direction of the vector $\boldsymbol{\theta}$ determines the axis of rotation and its length is the angle of rotation.*

As we approach the boundary of the ball, $|\boldsymbol{\theta}| \to \pi$, the uniqueness of this representation breaks down. A rotation by π around any axis is the same as a rotation through $-\pi$ around the same axis[2]. This means that $SO(3)$ can be identified with the solid disk of radius π in three-dimensional space, with antipodal points at the boundary identified. As a manifold $SO(3)$ is the real projective space $\mathbb{RP}^3$.

Exercise. Show that $S_3^2 = -1 + P_3$ where $P_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$; hence that $P_3 S_3 = 0$, and

$$S_3^3 = -S_3, \quad S_3^5 = S_3, \quad S_3^7 = -S_3, \cdots$$
$$S_3^4 = 1 - P_3, \quad S_3^6 = -1 + P_3, \quad S_3^8 = 1 - P_3$$

This leads to

$$e^{a_3 S_3} = \cos a_3 + \sin a_3 S_3 + (1 - \cos a_3) P_3.$$

More generally,

Exercise. Sow that $(\mathbf{a} \cdot \mathbf{S})^2 = -|\mathbf{a}|^2 + \mathbf{a} \otimes \mathbf{a}$ where $\mathbf{a} \otimes \mathbf{a} = \begin{pmatrix} a_1^2 & a_1 a_2 & a_1 a_3 \\ a_1 a_2 & a_2^2 & a_2 a_3 \\ a_1 a_3 & a_2 a_3 & a_3^2 \end{pmatrix}$ and hence,

$$e^{\mathbf{a} \cdot \mathbf{S}} = \cos |\mathbf{a}| \ + \frac{\sin |\mathbf{a}|}{|\mathbf{a}|} \, \hat{\mathbf{a}} + \frac{1}{|\mathbf{a}|^2} \, (1 - \cos |\mathbf{a}|) \, \mathbf{a} \otimes \mathbf{a} \tag{4.2.3}$$

where $\hat{\mathbf{a}} = \mathbf{a} \cdot \mathbf{S} = \begin{pmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{pmatrix}$ so that $\hat{\mathbf{a}}\mathbf{b} = \mathbf{a} \times \mathbf{b}$. Also, $\mathbf{a} \otimes \mathbf{a} \ \mathbf{b} = \mathbf{a} \cdot \mathbf{b} \ \mathbf{a}$. In particular, if $|\mathbf{a}| = \pi$, we have $e^{\mathbf{a} \cdot \mathbf{S}} = e^{-\mathbf{a} \cdot \mathbf{S}}$.

4.2.3. *The Real Projective Space* $\mathbb{RP}^n$ for $n \geq 2$*

The unit sphere $\mathbb{S}^n$ can be thought of the set of directions in $\mathbb{R}^{n+1}$: We think of two non-zero vectors as the same as long as they point in the same direction, dropping the information in the length. Formally, we say that $u \sim v$ if there is a *positive*

[2] It is important that not all the points at the boundary of the disk of radius π are identified with each other: each such point is identified only with its antipode. Rotations around an angle π around different directions describe distinct elements of $SO(3)$.

number λ such that $u = \lambda v$. The set of equivalence classes of non-zero vectors under this relation is the sphere.

We can go further and allow λ to be negative; the resulting equivalence class describes a ray passing through the origin. The set of such rays is the real projective space $\mathbb{RP}^n$. Another point of view is that $\mathbb{RP}^n$ is $\mathbb{S}^n$ with anti-podal points (which only differ by an overall sign in the co-ordinates) identified. For any $n \geq 2$, the sphere $\mathbb{S}^n$ is the universal cover of $\mathbb{RP}^n$; the fundamental group of $\mathbb{RP}^n$ is Z_2.

It is easier to visualize this if $n = 2$. A ray passing through the origin cuts the sphere $\mathbb{S}^2$ at two points. If the ray cuts the sphere somewhere on the Northern Hemisphere, it must also cut it at the anti-podal point in the Southern Hemisphere. We can uniquely determine the ray knowing just the co-ordinate of the point in the Northern Hemisphere. Points in the Northern Hemisphere are in 1–1 correspondence with the interior of the unit Disk: Given (x, y) with $x^2 + y^2 < 1$ we can determine $z = \sqrt{1 - (x^2 + y^2)}$ uniquely.

If the ray intersects the sphere at the Equator, its anti-podal point is also on the Equator. So, we can think of $\mathbb{RP}^2$ as the Disk with the antipodal points on its boundary (the Equator) identified.

This idea goes over to $n = 3$ as well, even if harder to imagine.

The case of $n = 1$ is different from those with $n \geq 2$. Identifying anti-podal points on a circle will give another circle., of half the circumference.

4.3. $SU(2)$ and its Lie Algebra $su(2)$

This basic example occurs all over physics, especially in the quantum mechanics of spin. We have alluded to it several times already. Still it is worth going over it in some more detail. Let us begin with the Lie algeba.

4.3.1. *Basis in* $su(2)$

$$su(2) = \left\{a \mid a^\dagger = -a, \mathrm{tr}a = 0\right\}$$

is the vector space of traceless anti-Hermitian matrices. A typical element can be written as $a = \begin{pmatrix} a_{11} & a_{12} \\ -a_{12}^* & -a_{11} \end{pmatrix}$ where a_{11} is purely imaginary and a_{12} is complex; so it depends on three independent real parameters. It is not hard to see that

$$s_1 = -\frac{i}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad s_2 = -\frac{i}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad s_3 = -\frac{i}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

form a basis:

$$a = a_1 s_1 + a_2 s_2 + a_3 s_3 = \frac{-i}{2} \begin{pmatrix} a_3 & a_1 - ia_2 \\ a_1 + ia_2 & -a_3 \end{pmatrix} \equiv \mathbf{a} \cdot \mathbf{s}$$

It will be useful to define $\mathbf{a} = (a_1, a_2, a_3)$ and $\mathbf{s} = (s_1, s_2, s_3)$.

They satisfy the commutation relations

$$[s_1, s_2] = s_3, \quad [s_2, s_3] = s_1, \quad [s_3, s_1] = s_2$$

Thus, the correspondence $s_i \mapsto S_i$ is an isomorphism between the Lie algebras $su(2)$ and $so(3)$. We can also say that s_i provide a two dimensional faithful representation of $o(3)$. We saw that $o(3)$ describes the effect of infinitesimal rotations on vectors in $\mathbb{R}^3$.

Some identities (easily verified) satisfied by the spin matrices will come in useful soon:

$$s_1^2 = -\frac{1}{4} = s_2^2 = s_3^2$$

$$s_1 s_2 + s_2 s_1 = 0 = s_2 s_3 + s_3 s_2 = s_3 s_1 + s_1 s_3$$

The quadratic identities above then become (again, easy to verify)

$$(\mathbf{a} \cdot \mathbf{s})^2 = -\frac{1}{4} |\mathbf{a}|^2$$

The scalar on the r.h.s. is to be thought as a multiple of the identity matrix.

Exercise. Show the more general identity

$$\mathbf{a} \cdot \mathbf{s} \;\; \mathbf{b} \cdot \mathbf{s} = -\frac{1}{4} \mathbf{a} \cdot \mathbf{b} + \frac{1}{2} (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{s} \tag{4.3.1}$$

for any pair of vectors $\mathbf{a}, \mathbf{b}$. Therefore (using some vector identities)

$$\mathbf{a} \cdot \mathbf{s} \;\; \mathbf{b} \cdot \mathbf{s} - \mathbf{b} \cdot \mathbf{s} \;\; \mathbf{a} \cdot \mathbf{s} = (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{s} \tag{4.3.2}$$

$$\mathbf{a} \cdot \mathbf{s}\, \mathbf{b} \cdot \mathbf{s}\, \mathbf{a} \cdot \mathbf{s} = -\frac{1}{2} \mathbf{a} \cdot \mathbf{b}\, \mathbf{a} \cdot \mathbf{s} + \frac{1}{4} |\mathbf{a}|^2 \mathbf{b} \cdot \mathbf{s} \tag{4.3.3}$$

There is a sort of uniqueness to the choice of spin matrices.

Exercise. Show that any triple of 2×2 matrices s_1', s_2', s_3' satisfying the relations

$$s_1' s_1' = -\frac{1}{4} = s_2' s_2' = s_3' s_3' \tag{4.3.4}$$

$$s_1' s_2' + s_2' s_1' = 0 = s_2' s_3' + s_3' s_2' = s_3' s_1' + s_1' s_3' \tag{4.3.5}$$

are linear combinations of the above matrices:

$$s'_i = R_{ji} s_j$$

Moreover, the 3×3 matrix R appearing here is an orthogonal matrix : $R^T R = 1$.

Answer To see this, just note that

$$(\mathbf{a}\cdot\mathbf{s}')^2 = -\frac{1}{4}|\mathbf{a}|^2$$

for any vector $\mathbf{a}$. On the other hand,

$$\mathbf{a}\cdot\mathbf{s}' = a_i s'_i = a_i R_{ji} s_j = \mathbf{a}'.\mathbf{s}, \quad \mathbf{a}' = R\mathbf{a}$$

Thus

$$(\mathbf{a}\cdot\mathbf{s}')^2 = (\mathbf{a}'\cdot\mathbf{s})^2 = -\frac{1}{4}|\mathbf{a}'|^2$$

so that

$$|R\mathbf{a}|^2 = |\mathbf{a}|^2$$

for any vector $\mathbf{a}$; this means that R is orthogonal.

4.3.2. *SU(2) is $\mathbb{S}^3$ as a manifold*

$SU(2)$ is the set of all unitary matrices of determinant one:

$$SU(2) = \left\{ g \mid g^\dagger = g^{-1}, \det g = 1 \right\}$$

Let us unpack this. Suppose

$$g = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix}$$

The conditions g can be expressed in terms of the matrix elements

$$\begin{pmatrix} g^*_{11} & g^*_{21} \\ g^*_{12} & g^*_{22} \end{pmatrix} = \frac{1}{g_{11}g_{22} - g_{12}g_{21}} \begin{pmatrix} g_{22} & -g_{12} \\ -g_{21} & g_{11} \end{pmatrix}, \quad g_{11}g_{22} - g_{12}g_{21} = 1$$

We can solve for g_{21} and g_{22} in terms of g_{11} and g_{22}:

$$g_{22} = g^*_{11}, \quad g_{21} = -g^*_{12}$$

The condition on the determinant becomes

$$|g_{11}|^2 + |g_{12}|^2 = 1$$

Thus, elements of $SU(2)$ are in $1-1$ correspondence with pairs of complex numbers $(g_{11}, g_{12}) \in \mathbb{C}^2$ satisfying the condition above. In terms of real parameters, $\mathbb{C}^2 = \mathbb{R}^4$; and the condition that the sum of absolute values squared says that the sum of the squares of the four real components is one. Thus, there is a $1-1$ correspondence between $SU(2)$ and the space of unit vectors in $\mathbb{R}^4$; that is, $SU(2)$ is $\mathbb{S}^3$ as a manifold.

This is very useful as it gives a way of picturing $SU(2)$. And makes its topology easy to understand. (The Lie groups $SU(n)$ for $n \geq 3$ are harder to describe as manifolds: They are not spheres any more. But they are well understood anyway) A co-ordinate system centered at the identity is given by the exponential map

$$g = e^{a_1 s_1 + a_2 s_2 + a_3 s_3}$$

The exponential is defined by an infinite series as usual

$$g = 1 + \mathbf{a}\cdot\mathbf{s} + \frac{1}{2!}(\mathbf{a}\cdot\mathbf{s})^2 + \frac{1}{3!}(\mathbf{a}\cdot\mathbf{s})^3 + \cdots$$

Combine all the even terms and the odd terms:

$$\begin{aligned} g = {} & 1 + \frac{1}{2!}(\mathbf{a}\cdot\mathbf{s})^2 + \frac{1}{4!}(\mathbf{a}\cdot\mathbf{s})^4 + \cdots \\ & + \mathbf{a}\cdot\mathbf{s}\left[1 + \frac{1}{3!}(\mathbf{a}\cdot\mathbf{s})^2 + \frac{1}{5!}(\mathbf{a}\cdot\mathbf{s})^4 \cdots\right] \end{aligned}$$

We can simplify this using the identity $(\mathbf{a}\cdot\mathbf{s})^2 = -\frac{1}{4}|\mathbf{a}|^2$ noted earlier:

$$\begin{aligned} g = {} & 1 - \frac{1}{2!}\left(\frac{|\mathbf{a}|}{2}\right)^2 + \frac{1}{4!}\left(\frac{|\mathbf{a}|}{2}\right)^4 + \cdots \\ & + \mathbf{a}\cdot\mathbf{s}\left[1 - \frac{1}{3!}\left(\frac{|\mathbf{a}|}{2}\right)^2 + \frac{1}{5!}\left(\frac{|\mathbf{a}|}{2}\right)^4 \cdots\right] \end{aligned}$$

We recognize the series:

$$\begin{aligned} \cos x &= 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 + \cdots \\ \frac{\sin x}{x} &= 1 - \frac{1}{3!}x^2 + \frac{1}{5!}x^4 + \cdots \end{aligned}$$

So, we have

$$g = \cos\left(\frac{|\mathbf{a}|}{2}\right) + 2\frac{\mathbf{a}\cdot\mathbf{s}}{|\mathbf{a}|}\sin\left(\frac{|\mathbf{a}|}{2}\right).$$

This gives us another way to understand the identification $SU(2) \approx \mathbb{S}^3$. The elements close to the identity are indeed described by a vector $\mathbf{a}$ of small length. But all the vectors with $|\mathbf{a}| = 2\pi$ describe the same element, $g = -1$: The sin term (which is the only one that knows about the direction of $\mathbf{a}$) vanishes. Thus, $SU(2)$ may be thought of as the disk of radius 2π in $\mathbb{R}^3$, but *with all the points on the boundary of the disk identified*. This is another way of thinking of the sphere.

In particular, we can now see that $SU(2)$ is connected and simply connected. Any point $e^{\mathbf{a}\cdot\mathbf{s}}$ is connected to the identity by the curve $e^{t\mathbf{a}\cdot\mathbf{s}}$: For $t = 0$ it is the identity and at $t = 1$ it is at $e^{\mathbf{a}\cdot\mathbf{s}}$. For example, the curve $e^{2\pi t s_3}$ for $0 \le t \le 1$ connects the identity to -1.

An example of a closed curve is a circle of radius $r < 2\pi$ in the plane with $a_3 = 0$:

$$\mathbf{a} = r\,(\cos\theta, \sin\theta, 0)\,.$$

It is not hard to see that this can be continuously deformed to a point (the identity) by shrinking $r \to 0$.

4.4. The homomorphism $R : SU(2) \to SO(3)$

The natural way to think of $SU(2)$ (its defining representation) is by its action on $\mathbb{C}^2$.

$$u \mapsto gu, \quad g \in SU(2), \quad u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \in \mathbb{C}^2.$$

This leads to an action of $SU(2)$ on the vector space of 2×2 matrices:

$$M \mapsto gMg^\dagger.$$

If a matrix transforms this way under $SU(2)$, then Mu transforms the same way as u:

$$Mu \mapsto gMg^\dagger gu = g(Mu)$$

Under this action, the trace of a matrix is unchanged:

$$\mathrm{tr}\left(gMg^\dagger\right) = \mathrm{tr}\left(g^\dagger gM\right) = \mathrm{tr}M.$$

Moreover,

$$M^\dagger \mapsto gM^\dagger g^\dagger$$

so that the property of being anti-Hermitian is preserved by the action of $SU(2)$. Thus, the space of traceless anti-Hermitian matrices carries a representation of $SU(2)$.

This should not be surprising: This space is just the Lie algebra $su(2)$; and every Lie group has a representation on its Lie algebra (the adjoint representation). But, we know that the Lie algebras of $SU(2)$ and $SO(3)$ are isomorphic. Explicitly, there is a 1–1 correspondence between vectors $\mathbf{b} \in \mathbb{R}^3$ and traceless anti-Hermitian 2×2 matrices:

$$\mathbf{b} \cdot \mathbf{s} = \begin{pmatrix} -\frac{i}{2} b_3 & -\frac{i}{2} b_1 - \frac{1}{2} b_2 \\ -\frac{i}{2} b_1 + \frac{1}{2} b_2 & \frac{i}{2} b_3 \end{pmatrix}$$

Now, recall that $\mathrm{tr} M^\dagger M$ is the sum of the absolute squares of its matrix elements; it is invariant under the $SU(2)$ action:

$$\mathrm{tr} M^\dagger M \mapsto \mathrm{tr}\left[g M^\dagger g^\dagger \, g M g^\dagger \right] = \mathrm{tr} M^\dagger M.$$

There is a simple relation between this "norm" of a traceless anti-Hermitian matrix and the length of the corresponding vector in $\mathbb{R}^3$:

$$\mathrm{tr}\left[(\mathbf{b} \cdot \mathbf{s})^\dagger \, (\mathbf{b} \cdot \mathbf{s}) \right] = \frac{1}{2} |\mathbf{b}|^2.$$

Under the action of $SU(2)$ this matrix transforms to $g\mathbf{b} \cdot \mathbf{s} g^\dagger$. There must be a vector $\mathbf{b}'$ such that

$$g\mathbf{b} \cdot \mathbf{s} g^\dagger = \mathbf{b}' \cdot \mathbf{s}$$

From the relationship of the matrix norm to the length of vectors, $\mathbf{b}'$ must have the same length as $\mathbf{b}$.

Now, this vector $\mathbf{b}'$ must depend linearly on $\mathbf{b}$; so there must be a matrix $R(g)$ such that $\mathbf{b}' = R(g)\mathbf{b}$. That is,

$$g(\mathbf{b} \cdot \mathbf{s}) g^\dagger = (R(g)\mathbf{b}) \cdot \mathbf{s} \tag{4.4.1}$$

Since $R(g)\mathbf{b}$ and $\mathbf{b}$ have the same length, $R(g)$ must be an orthogonal matrix. Recall that the determinant of an orthogonal matrix can only take values ± 1. Now, $\det R(g)$ is a continuous function $\det R : SU(2) \to \mathbb{R}$. Since the only allowed values are ± 1 it has to be a constant. This constant has to be 1 because $R(1) = 1$. Thus $R(g) \in SO(3)$.

By its definition we can see that it is a homomorphism

$$R(g_1 g_2) = R(g_1) R(g_2).$$

But R is not an isomorphism! For example, both g and $-g$ are mapped to the same matrix $R(g)$ in $SO(3)$: The lhs of (4.4.1) is unchanged by replacing g by $-g$. So, at best this is a 2 to 1 map.

We can understand this more explicitly working out the formula for $R(g)$ in exponential co-ordinates.

4.4.1. *R(g) in Exponential Co-ordinates*

Let is calculate:

$$g\,(\mathbf{b}\cdot\mathbf{s})\,g^{\dagger} = \left\{\cos\left(\frac{|\mathbf{a}|}{2}\right) + 2\frac{\mathbf{a}\cdot\mathbf{s}}{|\mathbf{a}|}\sin\left(\frac{|\mathbf{a}|}{2}\right)\right\}\mathbf{b}\cdot\mathbf{s}\left\{\cos\left(\frac{|\mathbf{a}|}{2}\right) - 2\frac{\mathbf{a}\cdot\mathbf{s}}{|\mathbf{a}|}\sin\left(\frac{|\mathbf{a}|}{2}\right)\right\}$$

$$= \cos^2\left(\frac{|\mathbf{a}|}{2}\right)\mathbf{b}\cdot\mathbf{s} - \sin^2\left(\frac{|\mathbf{a}|}{2}\right)2\frac{\mathbf{a}\cdot\mathbf{s}}{|\mathbf{a}|}(\mathbf{b}\cdot\mathbf{s})2\frac{\mathbf{a}\cdot\mathbf{s}}{|\mathbf{a}|}$$

$$+ \cos\left(\frac{|\mathbf{a}|}{2}\right)\sin\left(\frac{|\mathbf{a}|}{2}\right)\left\{2\frac{\mathbf{a}\cdot\mathbf{s}}{|\mathbf{a}|}\mathbf{b}\cdot\mathbf{s} - \mathbf{b}\cdot\mathbf{s}2\frac{\mathbf{a}\cdot\mathbf{s}}{|\mathbf{a}|}\right\}$$

Now, we use the identities (4.3.1–4.3.3)

$$g\,(\mathbf{b}\cdot\mathbf{s})\,g^{\dagger} = \cos^2\left(\frac{|\mathbf{a}|}{2}\right)\mathbf{b}\cdot\mathbf{s} - \frac{4}{|\mathbf{a}|^2}\sin^2\left(\frac{|\mathbf{a}|}{2}\right)\left\{-\frac{1}{2}\mathbf{a}\cdot\mathbf{b}\,\mathbf{a}\cdot\mathbf{s} + \frac{1}{4}|\mathbf{a}|^2\mathbf{b}\cdot\mathbf{s}\right\}$$

$$+ \cos\left(\frac{|\mathbf{a}|}{2}\right)\sin\left(\frac{|\mathbf{a}|}{2}\right)\frac{2}{|\mathbf{a}|}(\mathbf{a}\times\mathbf{b})\cdot\mathbf{s}$$

$$g\,(\mathbf{b}\cdot\mathbf{s})\,g^{\dagger} = \cos|\mathbf{a}|\,\mathbf{b}\cdot\mathbf{s} + \frac{\sin|\mathbf{a}|}{|\mathbf{a}|}(\mathbf{a}\times\mathbf{b})\cdot\mathbf{s} + \frac{2}{|\mathbf{a}|}\sin^2\left(\frac{|\mathbf{a}|}{2}\right)\mathbf{a}\cdot\mathbf{b}\,\mathbf{a}\cdot\mathbf{s}$$

By reading off the coefficient of $\mathbf{s}$ we get

$$R(g)\mathbf{b} = \cos|\mathbf{a}|\,\mathbf{b} + \frac{\sin|\mathbf{a}|}{|\mathbf{a}|}\,\mathbf{a}\times\mathbf{b} + \frac{2}{|\mathbf{a}|}\sin^2\left(\frac{|\mathbf{a}|}{2}\right)\mathbf{a}\cdot\mathbf{b}\,\mathbf{a}$$

By thinking of this as a linear operator acting on $\mathbf{b}$,

$$R(g) = \cos|\mathbf{a}| + \frac{\sin|\mathbf{a}|}{|\mathbf{a}|}\,\hat{\mathbf{a}} + \frac{2}{|\mathbf{a}|}\sin^2\left(\frac{|\mathbf{a}|}{2}\right)\mathbf{a}\otimes\mathbf{a}$$

where $\hat{\mathbf{a}} = \begin{pmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{pmatrix}$ and $\mathbf{a}\otimes\mathbf{a} = \begin{pmatrix} a_1^2 & a_1a_2 & a_1a_3 \\ a_1a_2 & a_2^2 & a_2a_3 \\ a_1a_3 & a_2a_3 & a_3^2 \end{pmatrix}$ as defined earlier. That is, they are matrices such that

$$\hat{\mathbf{a}}\mathbf{b} = \mathbf{a}\times\mathbf{b}, \quad \mathbf{a}\otimes\mathbf{a}\,\mathbf{b} = \mathbf{a}\cdot\mathbf{b}\,\mathbf{a}$$

Using the trigonometric identity for $\sin^2\left(\frac{|\mathbf{a}|}{2}\right)$ we can rewrite this as

$$R(g) = \cos|\mathbf{a}| + \frac{\sin|\mathbf{a}|}{|\mathbf{a}|}\,\hat{\mathbf{a}} + \frac{1}{|\mathbf{a}|^2}\,(1-\cos|\mathbf{a}|)\,\mathbf{a}\otimes\mathbf{a}$$

Using the identities (4.2.3) of the matrices S_1, S_2, S_3 we recognize this as $R(g) = e^{\mathbf{a}\cdot\mathbf{S}}$.

Now, recall that $SO(3)$ corresponds to the region $|\mathbf{a}| \leq \pi$; The points on the boundary with $|\mathbf{a}| = \pi$ are identified with their antipodes: $e^{\mathbf{a}\cdot\mathbf{S}} = e^{-\mathbf{a}\cdot\mathbf{S}}$ when $|\mathbf{a}| = \pi$. This is why $SO(3) \equiv \mathbb{RP}^3$ as a manifold.

On the other hand, $SU(2)$ corresponds to a region $|\mathbf{a}| \leq 2\pi$ which is twice as big. All the points with $|\mathbf{a}| = 2\pi$ correspond to the same element $g = -1$ of $SU(2)$. That is, $SU(2) \equiv \mathbb{S}^3$ as a manifold.

For example, there is no identification of $e^{\mathbf{a}\cdot\mathbf{s}}$ with $e^{-\mathbf{a}\cdot\mathbf{s}}$ when $|\mathbf{a}| = \pi$. Instead

$$e^{\mathbf{a}\cdot\mathbf{s}} = -2\frac{\mathbf{a}\cdot\mathbf{s}}{|\mathbf{a}|} = -e^{-\mathbf{a}\cdot\mathbf{s}}, \quad |\mathbf{a}| = \pi$$

The map $R : SU(2) \to SO(3)$ is insensitive to a change of sign of g :$R(g) = R(-g)$. From the explicit formulas, we can see that this is the only ambiguity: The kernel of R is precisely $Z_2 = \{1, -1\} \subset SU(2)$.

Exercise. Construct a curve which starts at some point $g \in SU(2)$ and ends at the point $-g$; its projection to $SO(3)$ via R should however, be a closed curve.

Solution: Pick a vector $\mathbf{a}$ of length $\pi$and define

$$\gamma(t) = e^{t\mathbf{a}\cdot\mathbf{s}}, \quad -1 \leq t \leq 1.$$

Its image under R is

$$R(\gamma(t)) = \cos[|t\mathbf{a}|] \ + \text{sign}(t)\frac{\sin[|t\mathbf{a}|]}{|\mathbf{a}|}\,\hat{\mathbf{a}} + \frac{|t|}{|\mathbf{a}|}\,(1 - \cos[|t\mathbf{a}|])\,\mathbf{a}\otimes\mathbf{a}$$

Since $|\mathbf{a}| = \pi$, the middle term vanishes at $t = \pm 1$. The remaining terms are insensitive to the sign of t. So $R(\gamma(-1)) = R(\gamma(1))$.

4.5. $SU(2)$ as a Group Extension of $SO(3)$

This can be viewed as part of a larger theory of extensions of groups. We need a bit of terminology from group theory.

Definition. The *kernel* of a group homomorphism $f : H \to G$ is the set of elements of H that are mapped to the identity in G. The *image* of f is the set of elements of G that are of the form $f(g)$ for some $g \in H$.

Thus, the kernel of the map $R : SU(2) \to SO(3)$ of the last section is the subset $\{1, -1\}$. The image of R is the whole of $SO(3)$.

Definition. A sequence of group homomorphisms $G_1 \xrightarrow{f_1} G_2 \xrightarrow{f_2} G_3 \xrightarrow{f_3} \cdots$ is *exact* if the image of each map is equal to the kernel of the next.

For example, this means not only that $f_2(f_1(g)) = 1, \forall g \in G_1$ but also that $f_2(x) = 1 \implies \exists g \in G_1$ such that $x = f_1(g)$.

Definition. $\hat{G}$ is the *extension* of a group G by another group H if there is an exact sequence of group homomorphisms

$$\{1\} \to H \xrightarrow{i} \hat{G} \xrightarrow{\Pi} G \to \{1\}$$

Here, $\{1\}$ denotes the trivial group containing just the identity. So, the image of the first map is just the identity. Exactness means that the only element of H mapped to the identity of $\hat{G}$ by i is the identity. This means that $i(H)$ is a subgroup of $\hat{G}$ which is isomorphic to H.

The last map just takes everything in G to the identity; its kernel is all of G. Exactness means that the image of Π must be the whole group: Every element of G arises as $\pi(g)$ for some $g \in \hat{G}$. But some information may be "lost": $\Pi(g)$ might not determine g itself.

Now we come to the essential part: $i(H)$ must be the kernel of Π. That is, the elements of $\hat{G}$ that are "killed off" by Π are precisely those that come from H. This means that G is a kind of "quotient" of $\hat{G}$ by H. Certainly, if they are finite groups, the number of elements in $\hat{G}$ is the number of elements of G times that of H.

In fact we can define two elements of $\hat{G}$ to be equivalent if they project to the same element of G:

$$\hat{g}_1 \sim \hat{g}_2 \iff \Pi(\hat{g}_1) = \Pi(\hat{g}_2)$$

That is, these two elements of $\hat{G}$ are equivalent if they only differ by multiplication by some $i(h)$ for $h \in H$

$$\hat{g}_1 \sim \hat{g}_2 \iff \exists h \in H \text{such that } \hat{g}_1 = i(h)\hat{g}_2$$

Then Π defines a 1-1 correspondence between such equivalence classes and G: We can think of G as $\hat{G}$ modulo this relation.

A rather obvious example of an extension is the product $G \times H$: The set of pairs (g, h) with $g \in G$ and $h \in H$ with the pairwise product

$$(g, h)(g', h') = (gg', hh').$$

This is the trivial extension. The interesting possibility is an extension that is not a product.

The content of the last section is that $SU(2)$ is a non-trivial extension of $SO(3)$ by Z_2

$$1 \to Z_2 \xrightarrow{i} SU(2) \xrightarrow{R} SO(3) \to 1$$

The map i simply takes -1 to the negative of the identity matrix. These are precisely the elements that are mapped to the identity by R. This is a non-trivial extension: We cannot think of $SU(2)$ as the product $SO(3) \times Z_2$. (The latter would instead be $O(3)$, the group of reflections and rotations.)

An even more basic example of a group extension is

$$1 \to \mathbb{Z} \xrightarrow{i} \mathbb{R} \xrightarrow{\Pi} U(1) \to 1$$

$U(1)$ is the set of complex numbers of unit magnitude (the same as 1×1 unitary matrices); it is an abelian group under multiplication. Geometrically it is the unit circle.

$\mathbb{R}$ and $\mathbb{Z}$ are the additive groups of real numbers and integers respectively. i just embeds integers into real numbers in the usual way. Π is the exponential

$$\Pi(x) = e^{2\pi i x}$$

When x is an integer this is equal to one. It is parametrized by an angle, which is a real number modulo 2π:

$$U(1) \approx \mathbb{R}/2\pi\mathbb{Z}.$$

4.5.1. *The Universal Cover of SO* (3)

Another point of view is topological. $\mathbb{U}(1)$ is not simply connected: A closed curve can wind around the circle many times and so cannot always be deformed to a point. The universal covering space of the circle is $\mathbb{R}$; the covering map is Π as defined above.

In the same spirit, $SU(2) \approx \mathbb{S}^3$ is simply connected. That is, any closed curve in it can be deformed to a point. But $SO(3) \approx \mathbb{RP}^3$ is not: The curve $e^{\theta S_3}$ for θ varying from $-\pi$ to π is closed. It starts and ends at the origin. But it is not deformable to the identity continuously. The corresponding curve in $SU(2)$ is $e^{-\frac{i}{2}\theta\sigma_3}$. It is not closed as it connects the antipodal points on the sphere. Topologically, $\mathbb{S}^3$ is a covering space of $SO(3)$. The map R above is the covering map. The fundamental group of $\mathbb{RP}^3$ (the set of closed curves up to homotopy equivalence) is the group Z_2.

Thus, the relation of $SU(2)$ is a kind of "twisted" product of Z_2 and $SO(3)$ both as a group and as a manifold. More precisely, $SO(3) = SU(2)/Z_2$ where two elements of $SU(2)$which only differ by a sign are considered equivalent. (For any group extension the equivalence classes $\hat{G}/i(H)$ form a group.)

This is typical of other Lie groups. Given a Lie algebra, there is a unique simply connected Lie group corresponding to it. The other Lie groups that share the same Lie algebra differ from the simply connected one by "division" by some discrete

subgroup. (The fundamental group of a Lie group is always abelian, so it is always a discrete abelian subgroup.)

4.5.2. *The gauge group of the standard model**

Physicists tend to be sloppy about this fact that there are several Lie groups with the same Lie algebra. For example, you will hear often that the gauge group of the standard model is $SU(3)\times SU(2)\times U(1)$. This is true at the level of a Lie algebra. As a group it is, to be precise[10], $S\,(U(3)\times U(2))$. It has the same Lie algebra. But differs from $SU(3)\times SU(2)\times U(1)$ by a discrete group Z_6. You may think this is a hair-splitting difference that doesn't matter physically. But it has an important physical consequence: It explains why the electric charge of the electron and the proton are equal in magnitude.

For a while there was hope for a Grand Unified Theory: $S(U(3)\times U(2))$ is a natural subgroup of $SU(5)$. Perhaps $SU(5)$ is the true gauge group, which is spontaneously broken to its subgroup $S(U(3)\times U(2))$. This would mean there are rare decays of the proton into leptons (an electron and some neutrino). Experiments have ruled this out. Hope remains that the true gauge group is some even large symmetry such as $O(10)$ or E_8. But in the absence of experimental hints, such endeavors remain speculative.

Exercise. Let $G \equiv S(U(3)\times U(2)) = \left\{\begin{pmatrix} g_3 & 0 \\ 0 & g_2 \end{pmatrix} \mid g_3 \in U(3),\, g_2 \in U(2),\, \det(g_3 g_2) = 1\right\}$ and $\hat{G} \equiv SU(3)\times SU(2)\times U(1) = \{(\hat{g}_3, \hat{g}_2, \hat{g}_1) \mid \hat{g}_3 \in SU(3), \hat{g}_2 \in SU(2), \hat{g}_1 \in U(1)\}$. Show that these two groups have the same Lie algebra. Find a group extension

$$1 \to Z_6 \xrightarrow{i} \hat{G} \xrightarrow{\Pi} G \to 1.$$

Chapter 5

ANGULAR MOMENTUM

5.1. Angular Momentum in Classical Mechanics

We make a digression into classical mechanics. Let us start with the classical theory. Recall that position and momentum satisfy the Poisson bracket relations

$$\{r_i, r_j\} = 0 = \{p_i, p_j\}, \quad \{p_i, r_j\} = \delta_{ij}$$

Here $\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$ is the Kronecker symbol. These relations are invariant under rotations. For example, if we change $p_i \longrightarrow R_{ik} p_k$, $r_j \mapsto R_{jl} r_l$ by an orthogonal matrix R,

$$\{p_i, r_j\} \mapsto R_{ik} R_{jl} \{p_k, r_l\} = R_{ik} R_{jl} \delta_{kl} = R_{ik} R_{jk} = \delta_{ij}$$

The last step follows from the definition of orthogonality. A basic principle of classical mechanics is that an infinitesimal transformation which leaves the canonical relations unchanged is generated by Poisson brackets with some function (called the generator). For example, under an infinitesimal rotation through an agle ϕ_3 around the third axis the changes in $\boldsymbol{r}$ and $\boldsymbol{p}$ are given by $\phi_3 S_3 \boldsymbol{r}$ and $\phi_3 S_3 \boldsymbol{p}$ respectively. There must be a quantity L_3 such that

$$\{L_3, \boldsymbol{r}\} = S_3 \boldsymbol{r}, \quad \{L_3, \boldsymbol{p}\} = S_3 \boldsymbol{p}$$

That is,

$$\{L_3, x\} = -y, \quad \{L_3, y\} = x, \quad \{L_3, z\} = 0$$

$$\{L_3, p_x\} = -p_y, \quad \{L_3, p_y\} = p_x, \quad \{L_3, p_z\} = 0$$

You can check that

$$L_3 = x p_y - y p_x$$

does the job. Similarly we will have

$$L_1 = yp_z - zp_y \implies \{L_1, \boldsymbol{r}\} = S_1\boldsymbol{r}, \quad \{L_1, \boldsymbol{p}\} = S_1\boldsymbol{p}$$

$$L_2 = zp_x - xp_z \implies \{L_2, \boldsymbol{r}\} = S_2\boldsymbol{r}, \quad \{L_2, \boldsymbol{p}\} = S_2\boldsymbol{p}$$

That is, they realize the $o(3)$ Lie algebra in terms of Poisson brackets. The change of any physical quantity f under a rotation by an infinitesimally small angle $\boldsymbol{\phi}$ is the Poisson bracket $\{\boldsymbol{\phi} \cdot \boldsymbol{L}, f\}$. We can conclude that t*monengt*he three functions are the components of the cross product vector

$$\boldsymbol{L} = \boldsymbol{r} \times \boldsymbol{p}$$

This is angular momentum. The components of angular momentum satisfy

$$\{L_j, L_k\} = -\epsilon_{jkl} L_l$$

Proposition. *Rotations are the canonical transformations generated by angular momentum*

Exercise. Show that $L^2 \equiv L_1^2 + L_2^2 + L_3^2$ has zero Poisson Brackets with the components of angular momentum $\{L^2, L_k\} = 0$

If rotations are the symmetry of a mechanical system, they will leave the hamiltonian unchanged. For example, the hamiltonian of the Kepler problem

$$H = \frac{\boldsymbol{p}^2}{2m} - \frac{k}{|\boldsymbol{r}|}$$

is clearly rotation invariant. Since angular momentum generates rotations, it follows that

$$\{\boldsymbol{L}, H\} = 0$$

But the hamiltonian generates time translations: For any observable quantity,

$$\frac{df}{dt} = \{H, f\}$$

Proposition. *If the hamiltonian is rotation invariant, angular momentum is conserved*: $\{\boldsymbol{L}, H\} = 0 \implies \frac{d\boldsymbol{L}}{dt} = 0$

If $\boldsymbol{L} \neq 0$, this has two important consequences for celestial mechanics:

- The orbit lies in the plane orthogonal to $\boldsymbol{L}$. In this plane, choose a polar coordinate system centered at the star and $\boldsymbol{L} = \begin{pmatrix} 0 \\ 0 \\ L_3 \end{pmatrix}$
- Then $L_3 = r^2 \frac{d\theta}{dt}$ must be a constant. This is Kepler's second law.

5.2. Angular Momentum in Quantum Mechanics

This is part of any standard course on quantum mechanics. My favorites references are [7, 9].

A real function on the phase space (representing a classical observable) goes over to a hermitian operator in quantum mechanics. Poisson brackets of observables become commutators of operators (apart from a factor of $-i\hbar$). Thus, the position and momentum operators satisfy Heisenberg relations

$$\left[r_i, r_j\right] = 0 = \left[p_i, p_j\right], \quad \left[p_i, r_j\right] = -i\hbar\delta_{ij}$$

The usual (i.e., Schrodinger) representation is to think of r_i as the multiplication operator and

$$p_i = -i\hbar\partial_i$$

as the derivative. In vector notation $\boldsymbol{p} = -i\hbar\nabla$.

Angular momentum is also then an operator

$$\boldsymbol{L} = -i\hbar\boldsymbol{r} \times \nabla$$

These operators satisfy the commutation relations

$$\left[L_j, L_k\right] = i\hbar\epsilon_{jkl}L_l$$

Apart from the usual factor of $-i\hbar$ these are just the commutation relations of the matrices S_i we found earlier. If we had defined

$$\boldsymbol{K} = \boldsymbol{r} \times \nabla$$

we would have obtained a representation of the Lie algebra $o(3)$:

$$\left[K_j, K_k\right] = \epsilon_{jkl}K_l.$$

While more natural mathematically, this is a bit awkward for physicists: K_i are now anti-Hermitian (unlike L_i which are hermitian). Their eigenvalues are then purely imaginary.

The $\hbar$ is merely the factor that converts from the intrinsic units of quantum mechanics (in which angular momentum is dimensionless) to those we got used to in classical mechanics (in which angular momentum has units of ML^2T^{-1}, same as $\hbar$). In many quantum calculations, it will be convenient to use "natural units" with $\hbar = 1$ and convert to classical units at the end. This is easily done using dimensional analysis. A little care is always needed to translate from physics to mathematical jargon: A factor of $-i\hbar$ is part of this translation.

Recall that a representation of a Lie algebra is said to be *unitary* if the matrices are anti-Hermitian.

$$\boldsymbol{K}^\dagger = -\boldsymbol{K}$$

The name is justified because the corresponding representation of the group (by matrices such as $e^{\boldsymbol{a}\cdot\boldsymbol{K}}$) is by unitary matrices. Equivalently, the angular momentum operators are hermitian.

In summary, Angular momentum operators form a unitary representation of the $o(3)$ Lie algebra.

It would be useful to have a classification of all such representations. More precisely, we need to find all the *irreducible* unitary representations up to *equivalence*. But, before classifying anything we must work out a few examples.

5.2.1. *Examples of representations*

Every Lie algebra has a representation on itself, the adjoint *representation.*

The commutator is itself a linear operator, so we can define

$$[a, u] = \hat{a}u$$

The Jacobi identity can be written as

$$[a, [b, u]] - [a, [b, u]] = [a, b], u]$$

That is

$$(\hat{a}\hat{b} - \hat{b}\hat{a})u = \widehat{[a, b]}u$$

which means that $a \mapsto \hat{a}$ is a representation. Another way to think of this is in terms of a basis and structure constants. Recall that

$$[X_i, X_j] = c_{ij}^k X_k \implies c_{ij}^l c_{lk}^m + c_{jk}^l c_{li}^m + c_{ki}^l c_{lj}^m = 0$$

Again, the Jacobi identity can be written as

$$C_i C_k - C_k C_i = c_{ik}^l C_l$$

where the C_i are matrices whose components are given by the structure constants themselves

$$[C_i]_j^l = c_{ij}^l.$$

If we work out the matrix elements in terms of the Levi–Civita tensor we simply get back the matrices S used to define $o(3)$ earlier!

So, the adjoint representation of $o(3)$ this representation is three-dimensional. We will find its place in the larger scheme once we classify representations; it turns to have spin one.

There is a representation of $o(3)$ in terms of Pauli Matrices. This is the next most important representation

$$S_j \mapsto -\frac{i}{2}\sigma_j$$

By direct calculation you can can verify that

$$\left[-\frac{i}{2}\sigma_1, -\frac{i}{2}\sigma_2\right] = -\frac{i}{2}\sigma_3$$

(and cyclic permutations of this relation.)

Although this is a representation of the Lie algebra $o(3)$, it does not give a representation of the group $O(3)$. Recall that a rotation through an angle 2π is the identity. For example,

$$e^{2\pi S_3} = 1$$

But there is a crucial sign difference in the representation:

$$e^{2\pi\left(-\frac{i}{2}\sigma_3\right)} = e^{-i\pi\sigma_3} = -1.$$

We will see that a quirk of quantum mechanics still allows this as a representation of the rotations. Quantum mechanics allows a "representation up to a phase" or "projective representation". The electron (the most common elementary particle, responsible for chemical reactions and hence life) exists only because of this apparently arcane technicality!

In the larger scheme below, we will see that this is the spin half representation of the Lie algebra. This is the representation with the smallest dimension, two. (If you exclude the trivial representation where S_i are all mapped to 0). All the other representations can be built out of this one by taking tensor products. So, it is also called the fundamental representation. Not coincidentally, the fundamental building blocks of matter (quarks and leptons) carry this representation.

5.3. Representations of $so(3)$

We now take up the task of classifying all the irreducible unitary representations of the Lie algebra $o(3)$ of rotations.

5.3.1. *Equivalent representations*

Any time you classify things, you need a notion of equivalence. Is a set containing two apples different from a set containing two oranges? It is, if you are making apple sauce. Not if you are merely learning to count.

Suppose K_i and K'_i are matrices satisfying the commutation relations of $o(3)$.

$$\left[K_j, K_k\right] = \epsilon_{jkl} K_l, \quad \left[K'_j, K'_k\right] = \epsilon_{jkl} K'_l.$$

We say they are equivalent representations if there is a matrix T such that

$$K'_j = T K_j T^{-1}$$

This is fair, because T is just a change of basis.

5.3.2. *Reducible representations*

The direct sum of two matrices (they could have different dimensions) is

$$A \oplus B = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$$

That is, write the matrices as blocks along the diagonal and fill in the rest with zeros. The rule for matrix multiplication gives

$$\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \begin{pmatrix} C & 0 \\ 0 & D \end{pmatrix} = \begin{pmatrix} AC & 0 \\ 0 & BD \end{pmatrix}$$

It is now clear that if $K_i$and K'_j are two representations (may be of different dimension) the direct sum $M_j = \left(\begin{smallmatrix} K_j & 0 \\ 0 & K'_j \end{smallmatrix}\right)$ is also a representation. For,

$$\left[\begin{pmatrix} K_j & 0 \\ 0 & K'_j \end{pmatrix}, \begin{pmatrix} K_k & 0 \\ 0 & K'_k \end{pmatrix}\right] = \begin{pmatrix} \left[K_j, K_k\right] & 0 \\ 0 & \left[K'_j, K'_k\right] \end{pmatrix}$$

Such representations should be considered reducible: They can be split into smaller pieces. Conversely, once we know the irreducible representations, we can build all representations by taking direct sums. The irreducible representations are the basic building blocks, the elementary objects, of representation theory.

A more general situation (which does not arise for $so(3)$ but does for some other Lie algebras) is that the representation matrices might of the triangular form

$$\begin{pmatrix} A & C \\ 0 & B \end{pmatrix}$$

where C is a rectangular matrix. Then

$$\left[\begin{pmatrix} A & C \\ 0 & B \end{pmatrix}, \begin{pmatrix} A' & C' \\ 0 & B' \end{pmatrix}\right] = \begin{pmatrix} [A, A'] & AC' + CB' - A'C - C'B \\ 0 & BB' \end{pmatrix}$$

These are also to considered as reducible representations, since there are smaller pieces A, B that provide sub-representations. But in this case $\begin{pmatrix} A & C \\ 0 & B \end{pmatrix}$ is not the direct sum $A \oplus B$. If we need to distinguish between the two situations, we will say that the representations which are direct sums are *completely reducible.*

If a representation is unitary (i.e., the representation matrices are anti-Hermitian) it is not hard to see any reducible representation is completely reducible: The hermitian conjugate of $\begin{pmatrix} A & C \\ 0 & B \end{pmatrix}$ is $\begin{pmatrix} A^\dagger & 0 \\ C^\dagger & B^\dagger \end{pmatrix}$; so if it is anti-Hermitian, $C = 0$. For physical reasons we are mostly interested in unitary representations. In any case, all the representations of $so(3)$ are unitary; so any reducible representation is completely reducible. So, the reducible representations we will encounter are mostly completely reducible.

But here is a tricky point: A change of basis can obscure the fact that a representation is reducible. The matrices may not look triangular (or block diagonal) in all bases. So, we could say a representation is reducible, if *there exists* a basis in which the representation matrices are of the upper triangular form (or even the block diagonal form).

There is a basis-independent way of saying this. If a representation is reducible (if such a basis exists), vectors of the form $\begin{pmatrix} u \\ 0 \end{pmatrix}$ (with zero components on the part acted upon by K'_j) are mapped to $\begin{pmatrix} K_j u \\ 0 \end{pmatrix}$: This subspace is invariant. (Each vector changes, but the space of all such vectors is unchanged.) So, here is the official

Definition. A representation is *reducible* if there is a proper (i.e., containing more than the 0 vector but not as big as the whole space) invariant subspace. And if there is no proper invariant subspace, the representation is *irreducible*.

So, our aim is to classify irreducible representations up to equivalence. A useful tool in identifying a reducible representation is[1]

[1] Later we will study Schur's Lemma as part of a deeper theory. We develop only what we need for now.

Lemma. *(Schur) A representation $S_i \mapsto M_i$ is irreducible iff the only matrices that commute with all the M_i are multiples of the identity*

The point is that the projection to an invariant subspace $\begin{pmatrix} u \\ 0 \end{pmatrix}$ is a matrix that commutes with $\begin{pmatrix} K_j & 0 \\ 0 & K'_j \end{pmatrix}$:

ot

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} K_j & 0 \\ 0 & K'_j \end{pmatrix} = \begin{pmatrix} K_j & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} K_j & 0 \\ 0 & K'_j \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

If there is a matrix C that commutes with M_i, its eigenspaces are invariant subspaces. (Unless C is a multiple of the identity, the only eigenspace is the whole space; so there are no proper invariant subspaces.)

In particular this means for $so(3)$ that

$$M_1^2 + M_2^2 + M_3^2$$

must be a multiple of the identity in an irreducible representation: The Lie algebra commutation relations imply that

$$\left[M_1^2 + M_2^2 + M_3^2, M_k\right] = 0.$$

5.3.3. *Unitary equivalence*

We can put a finer point on this. Recall that angular momentum must be a hermitian matrix: It must be an observable of quantum mechanics. The representation matrices K_j are related to angular momentum by

$$\boldsymbol{J} = -i\hbar \boldsymbol{K}$$

So the representation matrices must be anti-Hermitian, giving a unitary representation.

When we say two unitary representations are equivalent,

$$K'_j = T K_j T^{-1}$$

the matrix T must be required to be *unitary*.

$$T^{-1} = T^\dagger.$$

That way, observables (hermitian matrices) are mapped to observables by unitary transformations:

$$\boldsymbol{J}' = T \boldsymbol{J} T^\dagger.$$

Thus we want to classify irreducible unitary representations of $so(3)$ up to unitary equivalence.

5.3.4. *Roots and Weights*

It will be convenient to choose units such that $\hbar = 1$. The commutation relations of angular momentum

$$[J_3, J_1] = iJ_2, \quad [J_3, J_2] = -iJ_1, \quad [J_1, J_2] = iJ_3$$

are more conveniently written as

$$\left[J_3, \frac{J_1 + iJ_2}{\sqrt{2}}\right] = \frac{J_1 + iJ_2}{\sqrt{2}}, \quad \left[J_3, \frac{J_1 - iJ_2}{\sqrt{2}}\right] = -\frac{J_1 - iJ_2}{\sqrt{2}}.$$
$$\left[\frac{J_1 + iJ_2}{\sqrt{2}}, \frac{J_1 - iJ_2}{\sqrt{2}}\right] = J_3$$

The factor of $\frac{1}{\sqrt{2}}$ etc. are chosen so that we have the neat relations

$$[J_3, J_+] = J_+, \quad [J_3, J_-] = -J_-.$$
$$[J_+, J_-] = J_3$$

for

$$J_+ = \frac{J_1 + iJ_2}{\sqrt{2}}, \quad J_- = \frac{J_1 - iJ_2}{\sqrt{2}}.$$

Complex linear combinations as $J_\pm$ satisfying such "ladder" relations are called *roots* : Cartan's terminology in his classic work on Lie theory. The eigenvectors of J_3 (see below) are called *weights*.

For a unitary representation of $o(3)$,

$$J_3^\dagger = J_3$$

and

$$J_+^\dagger = J_-.$$

Now, suppose we have a eigenvector of J_3

$$J_3|m >= m|m >$$

The eigenvalue will be a real number m. The eigenvectors $|m >$ and $|m' >$ will be orthogonal when $m \neq m'$. We will also choose them to be have length one. That is

$$< m|m' >= \delta_{mm'}$$

This still leaves an ambiguity of a phase: We can multiply each $|m >$ by a complex number of modulus one.

Lemma.

$$J_3 J_+ |m> = (m+1) J_+ |m>, \quad J_3 J_- |m> = (m-1) J_+ |m>$$

Proof. For,

$$\begin{aligned} J_3 J_+ |m> &= [J_3, J_+]\,|m> + J_+ J_3 |m> \\ &= J_+ |m> + m J_+ |m> = (m+1) |J_+ |m> \end{aligned}$$

and similarly for J_-. □

The point is that J_+ is a "raising operator" is a "positive root": As long $J_+|m> \neq 0$, it is an eigen-vector of J_3 with eigenvalue $m+1$. (Similarly J_- is a "lowering operator".) We just showed that eigenvalues of J_3 are equally separated by one. We will see next that this is a finite sequence: There is a largest and a smallest value for m.

The key is the quantity ("Casimir operator")

$$J^2 = J_1^2 + J_2^2 + J_3^2$$

Being the sum of squares of hermitian matrices, this is a positive matrix. That is, its expectation value in any state is positive:

$$< u|J^2|u > \geq 0.$$

Lemma. *In an irreducible representation, J^2 is a multiple of the identity.*

Proof. The commutation relations of $o(3)$ require that $\left[J^2, J_k\right] = 0$. By Schur's Lemma this must be a multiple of the identity in an irreducible representation. □

It is useful to write this in terms of raising and lowering operators:

Lemma. $J_3(J_3+1) + 2J_-J_+ = J^2 = J_3(J_3-1) + 2J_+J_-$

Proof. Let us calculate

$$J_-J_+ = \frac{J_1 - iJ_2}{\sqrt{2}} \frac{J_1 + iJ_2}{\sqrt{2}} = \frac{J_1^2 + J_2^2 + i[J_1, J_2]}{2} = \frac{J_1^2 + J_2^2 - J_3}{2}$$

so that

$$J_3^2 + J_3 + 2J_-J_+ = J_3^2 + J_3 + J_1^2 + J_2^2 - J_3 = J^2$$

The other identity follows in a similar way by considering J_+J_- instead. □

Now, J_-J_+ is by itself a positive operator:

$$< u|J_-J_+|u >=| J_+|u >|^2 \geq 0.$$

So, we have[2]

$$J_3(J_3+1) \leq J^2$$

or

$$\left(J_3+\frac{1}{2}\right)^2 \leq J^2+\frac{1}{4}$$

So, J_3 cannot have eigenvalues that are too large in magnitude; there must be a smallest and a largest eigenvalue for J_3. Let the largest eigenvalue ("highest weight") be j:

$$J_3|j >= j|j >$$

It should be impossible to raise this by J_+ since it is already the largest:

$$J_+|j >= 0$$

It follows that

$$J^2|j >= j(j+1)|j >$$

By Schur's lemma, J^2 must be a multiple of the identity and so we now know its value on all the states:

$$J^2|m >= j(j+1)|m >$$

There is a state with the lowest weight:

$$J_-|m_{\min} >= 0$$

Now,

$$J^2|m_{\min} >= [J_3(J_3-1)+2J_+J_-]\,|m_{\min} >= m_{\min}(m_{\min}-1)|m_{\min} >$$

gives

$$j(j+1) = m_{\min}(m_{\min}-1)$$

[2]The meaning of an inequality $A < B$ relating hermitian operators is that the expectation value in *any* state ψ satisfies $\langle \psi, A\psi \rangle \leq \langle \psi, B\psi \rangle$.

The solutions of this quadratic are $m_{\min} = -j$, and $m_{\min} = j + 1$. The second solution can't be the right one, because $m_{\min}$ is supposed to be the smallest eigenvalue and $j + 1$ is even bigger than the maximum j.

So, m ranges between j and $-j$, in steps of one:

$$m = j, j - 1, \ldots, -j$$

There are $2j + 1$ such distinct values. Since $2j + 1$ must a positive integer, we conclude that the allowed values of j are

$$j = 0, \frac{1}{2}, 1, \ldots$$

For each choice j there is one irreducible representation.

To complete the story, we need the matrix elements of $J_{\pm}$. We already know that J_3 is the diagonal matrix

$$< m'|J_3|m >= m\delta_{m'm}.$$

And that

$$< m'|J_+|m > = 0, \quad \text{unless } m' = m + 1$$
$$< m'|J_-|m > = 0, \quad \text{unless } m' = m - 1$$

So, we just need the numbers

$$< m + 1|J_+|m >= \alpha_m$$

Since $J_- = J_+^{\dagger}$,

$$< m|J_-|m + 1 >= \alpha_m^* \implies < m - 1|J_-|m >= \alpha_{m-1}^*$$

Now,

$$< m|J^2|m >= j(j + 1)$$
$$j(j + 1) =< m|J_3(J_3 + 1) + 2J_-J_+|m >$$
$$= m(m + 1) + 2 < m|J_-J_+|m >$$
$$= m(m + 1) + 2 < m|J_-|m + 1 >< m + 1|J_+|m >$$
$$= m(m + 1) + 2|\alpha_m|^2$$

We have determined the magnitudes of the complex numbers

$$|\alpha_m| = \frac{1}{\sqrt{2}}\sqrt{j(j + 1) - m(m + 1)}$$

The phases can be chosen to be anything: These phases only affect the choice of basis. We choose

$$\alpha_m = \frac{1}{\sqrt{2}}\sqrt{j(j+1) - m(m+1)}$$

(The representation you get by another choice of phase is unitarily equivalent).

To conclude,

Theorem. *Up to unitary equivalence there is exactly one irreducible representation of* $o(3)$ *for each dimension* $2j+1 = 1, 2, 3, \ldots$ *An orthonormal basis is* $|m>$ *for* $m = j, j+1, \ldots, -j$.

The representation matrices are

$$< m|J_3|m >= m\delta_{m'm},$$

$$< m'|J_+|m >= \frac{1}{\sqrt{2}}\sqrt{j(j+1) - m(m+1)}\delta(m' = m+1)$$

$$< m'|J_-|m >= \frac{1}{\sqrt{2}}\sqrt{j(j+1) - m(m-1)}\delta(m' = m-1)$$

Exercise. Verify by direct calculation that these matrices satisfy the commutation relations of $o(3)$

We will denote this representation by D^j.

5.3.5. *Examples*

The Scalar

The smallest representation D^0 is one dimensional with $2j+1 = 1$. This is the trivial representation where all the representation matrices are zero. This is the scalar representation: Rotations are all represented by the identity. There are elementary particles (such as the Higgs boson) which belong to this representation.

The spinor. The next smallest representation $D^{\frac{1}{2}}$ is two dimensional. $2j+1 = 2 \implies j = \frac{1}{2}$. The representation matrices are

$$J_3 = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{pmatrix}, \quad J_- = \begin{pmatrix} 0 & 0 \\ \frac{1}{\sqrt{2}} & 0 \end{pmatrix}, \quad J_+ = \begin{pmatrix} 0 & \frac{1}{\sqrt{2}} \\ 0 & 0 \end{pmatrix}$$

or, using $J_1 = \frac{J_+ + J_-}{\sqrt{2}}$, $J_2 = -i\frac{J_+ - J_-}{\sqrt{2}}$

$$J_1 = \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix}, \quad J_2 = \begin{pmatrix} 0 & -\frac{i}{2} \\ \frac{i}{2} & 0 \end{pmatrix}, \quad J_3 = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{pmatrix}$$

We recognize this as the representation in terms of Pauli matrices we found earlier. The 2 dimensional complex vectors on which such matrices are called spinors. Remember that this is a representation of the Lie algebra $o(3)$ but not of the group $SO(3)$.

The vector. When $2j + 1 = 3$ and $j = 1$ we get for the representation D^1,

$$J_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad J_- = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad J_+ = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

or

$$J_1 = \begin{pmatrix} 0 & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & 0 \end{pmatrix}, \quad J_2 = \begin{pmatrix} 0 & -\frac{i}{\sqrt{2}} & 0 \\ \frac{i}{\sqrt{2}} & 0 & -\frac{i}{\sqrt{2}} \\ 0 & \frac{i}{\sqrt{2}} & 0 \end{pmatrix}, \quad J_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

Up to a unitary transformation these are proportional to the matrices representing infinitesimal rotations:

$$iTS_kT^{-1} = J_k$$

where

$$T = \begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{i}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \\ \frac{1}{\sqrt{2}} & \frac{i}{\sqrt{2}} & 0 \end{pmatrix}$$

Thus the $j = 1$ representation is unitarily equivalent to the adjoint representation.

The $D^{\frac{3}{2}}$ representation. Just for fun, we list the four dimensional representation of $o(3)$ with $j = \frac{3}{2}$:

$$J_1 = \begin{pmatrix} 0 & \frac{\sqrt{3}}{2} & 0 & 0 \\ \frac{\sqrt{3}}{2} & 0 & 1 & 0 \\ 0 & 1 & 0 & \frac{\sqrt{3}}{2} \\ 0 & 0 & \frac{\sqrt{3}}{2} & 0 \end{pmatrix}, \quad J_2 = \begin{pmatrix} 0 & -\frac{i\sqrt{3}}{2} & 0 & 0 \\ \frac{i\sqrt{3}}{2} & 0 & -i & 0 \\ 0 & i & 0 & -\frac{i\sqrt{3}}{2} \\ 0 & 0 & \frac{i\sqrt{3}}{2} & 0 \end{pmatrix},$$

$$J_3 = \begin{pmatrix} \frac{3}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & -\frac{1}{2} & 0 \\ 0 & 0 & 0 & -\frac{3}{2} \end{pmatrix}$$

Again, this is a representation of the Lie algebra $o(3)$ but not of the group $SO(3)$. This is true of any representation with half-integer j.

5.4. Irreducible Representations of $SU(2)$

To summarize, for every $j = 0, \frac{1}{2}, 1, \ldots$ there is an irreducible representation D^j of the group $SU(2)$. It has dimension

$$\dim D^j = 2j + 1.$$

If j is an integer, D^j has kernel Z_2, so is really a representation of the smaller group $SO(3) = SU(2)/Z_2$. For half-integer j it is a faithful (i.e., with trivial kernel) representation of $SU(2)$. The Casimir J^2 takes value $j(j+1)$ in D^j.

5.5. Spherical Harmonics

Armed with a knowledge of the irreducible representations of $o(3)$ let us take a second look at the angular momentum operator

$$\boldsymbol{L} = -i\boldsymbol{r} \times \nabla$$

You can see that this is unchanged under the scale transformations$\boldsymbol{r} \to \lambda \boldsymbol{r}$. This suggests that $\boldsymbol{L}$ only depends on the direction of $\boldsymbol{r}$ and not on its magnitude. Indeed, transforming to spherical polar co-ordinates

$$x = r\sin\theta\cos\phi, \quad y = r\sin\theta\sin\phi, \quad z = r\cos\theta, \quad 0 < \theta < \pi,$$
$$0 \le \phi < 2\pi, \quad 0 < r$$

we can calculate that [7, 9]

$$L_3\psi = -i\frac{\partial}{\partial\phi}$$
$$L_1\psi = i\left(\sin\phi\frac{\partial\psi}{\partial\theta} + \cos\phi\cot\theta\frac{\partial\psi}{\partial\phi}\right), \quad L_2\psi = i\left(-\cos\phi\frac{\partial\psi}{\partial\theta} + \sin\phi\cot\theta\frac{\partial\psi}{\partial\phi}\right)$$

Here ψ is a complex-valued function on the sphere $\mathbb{S}^2$. Let $C(\mathbb{S}^2)$ be the vector space of such functions. This representation of $o(3)$ on $C(\mathbb{S}^2)$ is reducible. After all,

$$L^2 = L_1^2 + L_2^2 + L_3^2$$

will commute with L_k. In fact, L^2, L_3 provide a complete set of commuting operators on $C(\mathbb{S}^2)$: Any function can be expressed as a linear combination of their eigenstates.

The simultaneous eigenvalue problem becomes

$$L^2 Y_{lm} = l(l+1)Y_{lm}$$

$$L_3 Y_{lm} = mY_{lm}$$

These partial differential equations appear in other disciplines (e.g., accoustics, electromagnetism) as well. The solutions are the *spherical harmonics.* Explicit solutions can be found by using recursion relations [7, 9]:

$$Y_{lm} = c_{lm} P_l^m(\cos\theta) e^{im\phi}$$

c_{lm} are constants usually chosen so that the integral of $|Y_{lm}|^2$ over $\mathbb{S}^2$ is one. P_l^m are certain polynomials ("Associated Legendre polynomials").

Even without the explicit formulas, we can see that $C(\mathbb{S}^2)$ contains one copy each of the representations with integer angular momentum (i.e., odd dimension)

$$l = 0, 1, \ldots$$

These representations are unitarily equivalent to the odd dimensional representations found above.

It cannot contain the half-integral representations because a function of the sphere must take the same value at $\phi = 0$ and $\phi = 2\pi$; so m must be an integer.

Exercise. Obtain the formula for L^2, $L_\pm$ in terms of polar co-ordinates [7, 9]. Use it to get a recursion relation for Y_{lm} allowing you to determine it from $Y_{ll}(\theta, \phi)$. Determine Y_{ll} by solving the equation for a highest weight vector, $L_+ Y_{ll} = 0$.

5.6. The Hydrogen Atom

Just as mechanics began with the solution of the Kepler problem, quantum mechanics began with the solution of the hydrogen atom. Spherical symmetry plays a crucial role in both problems. The hamiltonian is the sum of the kinetic energy of a particle of mass m and a potential energy which varies inversely with distance:

$$H = \frac{1}{2\mu}\boldsymbol{p}^2 - \frac{k}{|\boldsymbol{r}|}$$

μ is the mass of the electron, k is (up to a constant that has to do with the system of units you use) the product of the charges of the nucleus and the electron.

Representing the canonical commutation relations by

$$\boldsymbol{p} \mapsto -i\hbar\nabla$$

the hamiltonian becomes a differential operator

$$\hat{H} = -\frac{\hbar^2}{2\mu}\nabla^2 - \frac{k}{|\boldsymbol{r}|}$$

The problem is to find the eigenvalues and eigenstates of this operator. It also has scattering states, corresponding to the continuous spectrum, which we ignore for now. We are also ignoring several other small effects: The finite mass of the nucleus, relativistic corrections, the spin of the electron, the hyperfine splitting due to the magnetic moments of the electron and the nucleus, Quantum Electrodynamics effects (such as Lamb Shift). Indeed, a history of quantum theory can be based on the improving understanding of the hydrogen atom. The story is more or less complete by now. We will stick to the simplest case in this section.

The first step towards the solution is to note that the angular momentum

$$\boldsymbol{L} = \boldsymbol{r} \times \boldsymbol{p}$$

commutes with $\hat{H}$. More precisely, $\{\hat{H}, L^2, L_3\}$ form a set of commuting observables. From our earlier discussion, the simultaneous eigenstates of L^2 and L_3 are the spherical harmonics, which determine how the states depend on the angles. So, the eigenvalue problem for $\hat{H}$ must reduce to solving some differential equation in the radial variable r. This reduction is typical of the use of symmetry in physics.

We must express $\hat{H}$ in terms of a radial operator and L^2. Since the potential energy is already known to be $-\frac{k}{r}$, we just need a formula for the Laplacian∇^2 in spherical polar co-ordinates [7, 9]

$$\nabla^2 = \frac{1}{r^2}\frac{\partial}{\partial r}r^2\frac{\partial}{\partial r} + \frac{1}{r^2\sin^2\theta}\left\{\left(\sin\theta\frac{\partial}{\partial\theta}\right)^2 + \frac{\partial^2}{\partial\phi^2}\right\}$$

leading to (in units with $\hbar = 1$)

$$\hat{H}\psi = \frac{1}{2\mu}\left[-\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right) + \frac{1}{r^2}\left(L^2\psi\right)\right] - \frac{k}{r}$$

The angular derivatives occur only in the combination L^2. We can assume that the wave function is of the form

$$\psi(r,\theta,\phi) = R(r)Y_{lm}(\theta,\phi)$$

where Y_{lm} are the simultaneous eigenstates of L^2 and L_3 we saw earlier. Then we get the radial eigenvalue equation

$$\frac{1}{2\mu}\left[-\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right) + \frac{l(l+1)}{r^2}R\right] - \frac{k}{r} = ER$$

The angular factors pull out of the radial derivatives and cancel out. This ordinary differential equation is solved in every book on quantum mechanics [7, 9]. Square integrable solutions occur when

$$E = -\frac{\mu k^2}{2n^2}, \quad n = 1, 2, 3 \cdots$$

The label n is called the "principal quantum number" in the jargon of atomic physics.

The solutions are

$$R_{nl}(r) = C_{nl} r^l e^{-\frac{r}{na}} L^{2l+1}_{n-l-1}\left(2\frac{r}{na}\right)$$

where $L^b_a(r)$ are the "associated Laguerre polynomials" [7, 9].

Here C_{nl} is a constant, which is usually chosen so that eigenvectors have length one. Also,

$$a = \frac{\hbar^2}{\mu k^2}$$

is a quantity with the dimensions of length ("Bohr radius"). $L^{2l+1}_{n-l-1}(x)$ are some polynomials of order $n - l - 1$ associated with LaGuerre. So, we have

$$n - l - 1 \geq 0$$

which leads to the range of allowed values of angular momentum:

$$l = 0,\ 1, \ldots,\ n - 1.$$

Of course, we already know that

$$m = l,\ l - 1,\ \ldots, -l$$

5.6.1. *Numerical estimates*

For completeness, let us recall the numerical values of the parameters involved[7, 9]. A convenient unit of energy in atomic physics is an electron volt eV. If we choose $c = 1$ in addition to $\hbar = 1$, the mass of an electron is, in energy units

$$\mu \approx 0.511 \times 10^6\ eV.$$

$k = e^2$ is best thought of in terms of the fine structure constant, because it is dimensionless:

$$\alpha \equiv \frac{e^2}{\hbar c} \approx \frac{1}{137}$$

So, in energy units

$$E_n = -\frac{\mu\alpha^2}{2n^2} \approx -\frac{13.6}{n^2} eV$$

In particular, the smallest value of energy (ground state energy) is $E_1 \approx -13.6$eV. This is in excellent agreement with experiment. Agreement can be made perfect by adding in various small effects we ignored.

5.7. Spin and $SU(2)$

Originally, angular momentum arose as $\boldsymbol{L} = \boldsymbol{r} \times \boldsymbol{p}$ i.e., *orbital angular momentum.* It is zero when momentum is zero. Now we know that a system can have angular momentum even when its momentum is zero. This is *intrinsic angular momentum* or *spin.*

If we take the non-relativistic theory of the hydrogen atom above literally, the ground state is unique and has zero angular momentum: $n = 1 \implies l = 0$. But this was proven to be wrong experimentally by the Stern–Gerlach experiment. (Originally done with Silver atoms; later, the effects were reproduced in hydrogen as well.) The electron in the ground state of hydrogen has two possible states, which respond differently to a magnetic field. Roughly half the atoms are found to be in the state with angular momentum pointed along the magnetic field and the other half have it pointed opposite.

This can be explained if the electron has an intrinsic angular momentum $\frac{1}{2}\hbar$. The wave function is then a function $\psi : \mathbb{R}^3 \to \mathbb{C}^2$. The two components of the wave function allow the Pauli matrices to act on it. The total angular momentum is the sum of the orbital and intrinsic angular momenta:

$$\boldsymbol{J} = \boldsymbol{r} \times \boldsymbol{p} + \frac{1}{2}\boldsymbol{\sigma}.$$

These provide a representation of the group $SU(2)$ rather than $SO(3)$.

A complete explanation of this needs the relativistic theory (the Dirac equation). To the first approximation the effect of the spin is a doubling of the number of allowed states of the electron. If there is a magnetic field there is an additional contribution to the energy

$$H_{\text{mag}} = g\boldsymbol{\sigma} \cdot \boldsymbol{B}$$

This shifts the ground energy by $\pm g|\boldsymbol{B}|$, removing the degeneracy. By passing a beam of atoms through a magnetic field, we can separate them into two groups depending on the eigenvalue of $\boldsymbol{\sigma} \cdot \boldsymbol{B}$.

Even in the absence of an external magnetic field, there is a magnetic field created by the orbital motion of the electron around the atom ("spin-orbit coupling"). This adds a term

$$H_{SO} = \epsilon \boldsymbol{L} \cdot \boldsymbol{\sigma}$$

with a small coefficient ϵ that can be calculated in terms of the electron mass and charge. This means that the orbital and spin angular momenta are not separately conserved: Only their sum $\boldsymbol{J}$ commutes with the Hamiltonian.

We saw that L_3 has to be an integer multiple of $\hbar$ because the wave function has to return to its value after a rotation through 2π. That is not necessary for spin; it can be integer or half integer. The electron takes advantage of this: $\frac{1}{2}\hbar\sigma_3$ takes half integer values. For these particles, we get a faithful representation of $SU(2)$ instead of $SO(3)$.

There are other elementary particles (the photon, W boson) for which the spin is an integer multiple of $\hbar$; for these we get a representation of $SO(3)$.

The total angular momentum is the sum of various contributions: The spin of the electrons, the orbital angular momentum of the electrons and the spin of the nucleus. We will need to develop a theory of how to combine different sources of angular momentum.

Chapter 6

ADDITION OF ANGULAR MOMENTUM

6.1. Direct Products

The *direct product* (also called the *Tensor Product* or *Kronecker product*) of two matrices is defined as the matrix obtained by taking every possible product of their components. Best to understand this by examples. The direct product of a vector with two components with a vector of 3 components can be written as a 2×3 array with six components:

$$\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \otimes \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} u_1v_1 & u_1v_2 & u_1v_3 \\ u_2v_1 & u_2v_2 & u_2v_3 \end{pmatrix}$$

The direct product $A \otimes B$ of a 2×2 matrix A with a 3×3 matrix B, acting on $u \otimes v$ is defined to be

$$(A \otimes B)(u \otimes v) = (Au) \otimes (Bu)$$

You can verify that this is a left multiplication of the array $u \otimes v$ above by A and a right multiplication by B^T.

$$(A \otimes B)\,(u \otimes v) = A \begin{pmatrix} u_1v_1 & u_1v_2 & u_1v_3 \\ u_2v_1 & u_2v_2 & u_2v_3 \end{pmatrix} B^T \tag{6.1.1}$$

The transpose on the right is needed for this to work out right. You can check that with this definition,

$$(A \otimes B)(C \otimes D) = AC \otimes BD$$

holds.

For symbolic calculations (using Mathematica for example) it is more convenient to rearrange ("flatten") the array $u \otimes v$ into a vector with six components:

$$\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \otimes \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} u_1 v_1 \\ u_1 v_2 \\ u_1 v_3 \\ u_2 v_1 \\ u_2 v_2 \\ u_2 v_3 \end{pmatrix}$$

Then $A \otimes B$ turns into a 6×6 matrix:

$$A \otimes B \equiv \begin{pmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{pmatrix} \otimes \begin{pmatrix} B_{1,1} & B_{1,2} & B_{1,3} \\ B_{2,1} & B_{2,2} & B_{2,3} \\ B_{3,1} & B_{3,2} & B_{3,3} \end{pmatrix}$$

$$= \begin{pmatrix} A_{1,1}B_{1,1} & A_{1,1}B_{1,2} & A_{1,1}B_{1,3} & A_{1,2}B_{1,1} & A_{1,2}B_{1,2} & A_{1,2}B_{1,3} \\ A_{1,1}B_{2,1} & A_{1,1}B_{2,2} & A_{1,1}B_{2,3} & A_{1,2}B_{2,1} & A_{1,2}B_{2,2} & A_{1,2}B_{2,3} \\ A_{1,1}B_{3,1} & A_{1,1}B_{3,2} & A_{1,1}B_{3,3} & A_{1,2}B_{3,1} & A_{1,2}B_{3,2} & A_{1,2}B_{3,3} \\ A_{2,1}B_{1,1} & A_{2,1}B_{1,2} & A_{2,1}B_{1,3} & A_{2,2}B_{1,1} & A_{2,2}B_{1,2} & A_{2,2}B_{1,3} \\ A_{2,1}B_{2,1} & A_{2,1}B_{2,2} & A_{2,1}B_{2,3} & A_{2,2}B_{2,1} & A_{2,2}B_{2,2} & A_{2,2}B_{2,3} \\ A_{2,1}B_{3,1} & A_{2,1}B_{3,2} & A_{2,1}B_{3,3} & A_{2,2}B_{3,1} & A_{2,2}B_{3,2} & A_{2,2}B_{3,3} \end{pmatrix} \tag{6.1.2}$$

Exercise. Verify that two ways (6.1.1,6.1.2)of thinking about $A \otimes B$ above are equivalent.

We will mostly use the "flattened" description of the Direct Product.

Exercise. Show that $\mathrm{tr}(A \otimes B) = \mathrm{tr}A\mathrm{tr}B$. Contrast with the formula $\mathrm{tr}(A \oplus B) = \mathrm{tr}A + \mathrm{tr}B$ which you should also prove.

It follows that the direct product of two representations R_1, R_2 of a group is another representation:

$$R(g) = R_1(g) \otimes R_2(g)$$

$$R(gg') = R_1(gg') \otimes R_2(gg') = R_1(g)R_1(g') \otimes R_2(g)R_2(g') = R(g) \otimes R(g')$$

Infinitesimally, we can also get a representation of a Lie algebra:

$$g = 1 + \epsilon \underline{g} + \mathrm{O}(\epsilon)$$

$$R(g) = 1 + \epsilon \underline{R}(\underline{g}) + \mathrm{O}(\epsilon)$$

$$\underline{R}(\underline{g}) = \underline{R}_1(\underline{g}) \otimes 1 + 1 \otimes \underline{R}_2(\underline{g})$$

This is still called the direct product of Lie algebra representations. The physical application is when each factor represents an independent degree of freedom of the system, transforming under the group separately from each other. So physicists would call it the additional of angular momenta.

Usually such direct products are reducible. Decomposing the direct product as a sum of irreducible representations is a useful way of solving many physics problems.

6.1.1. *Hyperfine splitting of hydrogen*

Let us consider a physical example [11, 12]. In the ground state of the hydrogen atom, the orbital angular momentum of the electron is zero. Both the electron and the proton carry spin half. So, the total angular momentum of the atom is the sum of these spins. Let us ignore all other degrees of freedom for simplicity. The proton and the electron each have two independent states; so the combined space of states is four dimensional. We can describe them by four complex numbers

$$\psi_{ab} \quad a = 1, 2, \quad b = 1, 2$$

where the first index labels the electron spin state and the second the proton state. Thus the electron and proton spin matrices are $\boldsymbol{\sigma} \otimes 1$ and $1 \otimes \boldsymbol{\sigma}$ respectively. The total angular momentum is the sum

$$\boldsymbol{J} = \frac{1}{2}\boldsymbol{\sigma} \otimes 1 + 1 \otimes \frac{1}{2}\boldsymbol{\sigma}$$

It is possible "flatten" the array ψ_{ab} and think of it as a vector $\begin{pmatrix} \psi_{11} \\ \psi_{12} \\ \psi_{21} \\ \psi_{22} \end{pmatrix}$. Operators acting on it are then 4×4 matrices.

For example,

$$\sigma_1 \otimes 1 \begin{pmatrix} \psi_{11} \\ \psi_{12} \\ \psi_{21} \\ \psi_{22} \end{pmatrix} = \begin{pmatrix} \psi_{21} \\ \psi_{22} \\ \psi_{11} \\ \psi_{12} \end{pmatrix}, \quad 1 \otimes \sigma_1 \begin{pmatrix} \psi_{11} \\ \psi_{12} \\ \psi_{21} \\ \psi_{22} \end{pmatrix} = \begin{pmatrix} \psi_{12} \\ \psi_{11} \\ \psi_{22} \\ \psi_{21} \end{pmatrix}$$

so that

$$\sigma_1 \otimes 1 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \quad 1 \otimes \sigma_1 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

etc. leading to

$$J_1 = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}, \quad J_2 = \begin{pmatrix} 0 & -\frac{i}{2} & -\frac{i}{2} & 0 \\ \frac{i}{2} & 0 & 0 & -\frac{i}{2} \\ \frac{i}{2} & 0 & 0 & -\frac{i}{2} \\ 0 & \frac{i}{2} & \frac{i}{2} & 0 \end{pmatrix}, \quad J_3 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

Using this you can check that $\boldsymbol{J}$ above is a representation of angular momentum, in terms of 4×4 matrices. But it is a reducible representation.

As with any representation, $\boldsymbol{J}^2 = J_1^2 + J_2^2 + J_3^2$ will commute with all the components J_k. By direct calculation we will get

$$J^2 = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$

Since this is not a multiple of the identity, Schur's lemma tells us that the representation is reducible. The eigenvalues of J^2 are 2 (repeated three times) and 0. So there is a matrix T such that

$$TJ^2T^{-1} = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

By solving the eigenvalaue problem we can find

$$T = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 0 & 1 \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{pmatrix}$$

We now see that the representation provided by the matrices J_1, J_2, J_3 is equivalent to the direct sum of a representation with spin one and another with spin

zero:

$$TJ_1T^{-1} = \begin{pmatrix} 0 & \frac{1}{\sqrt{2}} & 0 & 0 \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad TJ_2T^{-1} = \begin{pmatrix} 0 & -\frac{i}{\sqrt{2}} & 0 & 0 \\ \frac{i}{\sqrt{2}} & 0 & -\frac{i}{\sqrt{2}} & 0 \\ 0 & \frac{i}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$TJ_3T^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

What does this imply for the hydrogen energy levels? All four of these states have the same energy in the leading approximation. But both the electron and the proton have magnetic moments, which are proportional to their spins. Classically, the energy of a pair of magnets is the dot product of their moments. Quantum mechanically, the components of the magnetic moments are matrices. Still the energy is giving by taking the product of each of the three components and summing over them:

$$H_1 = \lambda \sigma_k \otimes \sigma_k$$

The quantity λ is proportional to the product of the magnetic moments [12]. The matrix T above will diagonalize this hamiltonian:

$$TH_1T^{-1} = \lambda \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -3 \end{pmatrix}$$

The four originally degenerate states are now split into two groups: Three spin one states with energy λ and a spin zero state with energy -3λ. It turns out that the sign of λ is positive (taking into account the subtle effect of the overlap of atomic wavefunction with the position of the proton [12]). Thus the spin zero state has lower energy.

Transitions between these states will occur with emission or absorption of a photon. Angular momentum would be conserved because the photon also has spin one. The energy of the photon will be 4λ, which works out [12] to $5.87 \times 10^{-6} eV$; i.e., a frequency of 1420MHz or a wavelength of 21.1cm. Experiments with microwaves (masers) can measure this with great accuracy, providing precision tests of quantum theory [11].

This is also a very important signal of atomic hydrogen in radio astronomy.

6.1.2. *Symmetric and anti-symmetric states*

What is the meaning of the unitary transformation T that diagonalizes J^2? Let us see its effect on a general vector

$$T\begin{pmatrix}\psi_{11}\\ \psi_{12}\\ \psi_{21}\\ \psi_{22}\end{pmatrix} = \begin{pmatrix}\psi_{11}\\ \frac{\psi_{12}+\psi_{21}}{\sqrt{2}}\\ \psi_{22}\\ \frac{\psi_{12}-\psi_{21}}{\sqrt{2}}\end{pmatrix}$$

The first three components on the r.h.s. are symmetric under the interchange of indices; the last component is anti-symmetric. So, the symmetric combinations have total spin 1 and the anti-symmetric state has total spin zero. The point is that the subspace of symmetric (or anti-symmetric) states is invariant under rotations; and that they cannot be reduced further.

6.1.3. *Highest weight states*

Since an irreducible representation of $o(3)$ is uniquely determined by its dimension (not all Lie algebras have this property) we can write our result as

$$\mathbf{2} \otimes \mathbf{2} = \mathbf{3} \oplus \mathbf{1}$$

More often physicists like to label an irreducible representation in terms of the maximal value of J_3. Then we say that the product of two spin $\frac{1}{2}$ representations is the direct sum of a spin one and a spin zero representation. If we denote the irreducible representation with highest weight j by D^j, we would write this as

$$D^{\frac{1}{2}} \otimes D^{\frac{1}{2}} = D^0 \oplus D^1$$

Thus, the maximum magnitude for angular momentum is the sum and the minimum is the difference.

This is the simplest case of a more general result on the addition of angular momentum.

6.2. General Case of Addition of Angular Momentum

Let us begin with the classical limit. Suppose we have a vector of magnitude j_1 added to another vector of magnitude j_2. What is the possible range of the lengths of the sum of two such vectors?

If they are parallel, the length will be $j_1 + j_2$; if anti-parallel, it will be $|j_2 - j_1|$. All values in between can also occur, depending on the angle between the two vectors. For angular momentum operators, not all lengths are allowed; in an irreducible representation, $\boldsymbol{J}^2$ can only take values $j(j+1)$ for positive integer or half-integer j. So, a first guess might be that the direct product of a spin j_1 representation with a spin j_2 representation will be a direct sum of spin j representations with $j = j_1 + j_2, j_1 + j_2 - 1, \ldots, |j_2 - j_1|$. There is a way to check if that can be true: is the sum of the dimensions of all such irreducible representations equal to $(2j_1 + 1)(2j_2 + 1)$?

We can calculate (choose $j_2 > j_1$ to be definite)

$$\begin{aligned}
\sum_{j=|j_2-j_1|}^{j_2+j_1} (2j+1) &= \sum_{k=0}^{2j_1} (2[j_2 - j_1 + k] + 1) \\
&= (2[j_2 - j_1] + 1)\,(2j_1 + 1) + 2\sum_{k=0}^{2j_1} k \\
&= (2[j_2 - j_1] + 1)\,(2j_1 + 1) + 2\frac{(2j_1+1)2j_1}{2} \\
&= (2j_1 + 1)\,\{2[j_2 - j_1] + 1 + 2j_1\} = (2j_1 + 1)(2j_2 + 1)
\end{aligned}$$

Thus, it is quite possible that $D^{j_1} \otimes D^{j_2}$ and $\bigoplus_{j=|j_2-j_1|}^{j_1+j_2} D^j$ are equivalent representations. This does not prove that these are equivalent: The dimension does not uniquely specify a representation of $su(2)$, unless it is irreducible.

To really prove tequivalence, we need a more powerful tool, the character function.

6.2.1. *The character of an irreducible representation*

Given a representation (reducible or not) R of $SU(2)$ we can define its character to be the function

$$\chi_R(g) = \mathrm{tr}R(g)$$

Note that, being a trace, the character is invariant under equivalence transformations; therefore equivalent representations have the same character.

A particular case is when $g = e^{\frac{i}{2}\sigma_3\phi}$; indeed any $g$can be brought to this form by a conjugation in the group, which leave the character invariant. The matrix representing $\frac{\sigma_3}{2}$ in the representation R is J_3. Therefore,

$$\chi_R(\phi) = \mathrm{tr}e^{iJ_3\phi} = \sum_m d(m)e^{im\phi}$$

where $d(m)$ are the number of eigenstates of J_3 with eigenvalue m. We can think of this as a generating function that keeps count of the degeneracies $d(m)$. Clearly the dimension of the representation is the particular case when $\phi = 0$.

$$\lim_{\phi \to 0} \chi_R(\phi) = \dim R$$

For example, in the spin half representation, J_3 is just $\frac{\sigma_3}{2}$ and

$$\chi_{\frac{1}{2}}(\phi) = e^{i\frac{\phi}{2}} + e^{-\frac{i}{2}\phi} = 2\cos\frac{\phi}{2}$$

Let us find the character of the irreducible representation D^j. The eigenvalues of J_3 are $m = j, j-1, \ldots, -j$, each occurring with degeneracy one:

$$\chi_j(\phi) \equiv \mathrm{tr} D^j \left(e^{\frac{i}{2}\sigma_3\phi}\right) = \sum_{m=-j}^{j} e^{im\phi}$$

Remark 17. A word on notation for clarity: $D^j(g)$ is the matrix representing $g \in SU(2)$ in the representation of spin j . So, g is a 2×2 matrix while $D^j(g)$ is a $(2j+1) \times (2j+1)$ matrix. When g is diagonal, it is of the form $e^{\frac{i}{2}\sigma_3\phi}$ for some angle ϕ. The matrix $D^j \left(e^{\frac{i}{2}\sigma_3\phi}\right) = e^{iJ_3\phi}$ representing it can be found by replacing $\frac{i\sigma_3}{2}$ by J_3, which is also diagonal, with eigenvalues $-j, -j+1, \ldots, j-1, j$. Its trace is the rhs above.

It is a little exercise in algebra and trigonometry to evaluate this geometric series:

$$\chi_j(\phi) = \frac{\sin\left([2j+1]\,\frac{\phi}{2}\right)}{\sin\frac{\phi}{2}}$$

This can be written in terms of the Chebyshev polynomial of the second kind.

$$\chi_j(\phi) = \frac{\sin\left([2j+1]\,\frac{\phi}{2}\right)}{\sin\frac{\phi}{2}} = U_{2j}\left(\cos\left[\frac{\phi}{2}\right]\right)$$

As a check, L'Hospital rule gives

$$\lim_{\phi \to 0} \chi_j(\phi) = 2j+1$$

which is the dimension of the representation (the trace of the identity).

We can also recover the special case

$$\chi_{\frac{1}{2}}(\phi) = \frac{\sin\phi}{\sin\frac{\phi}{2}} = 2\cos\frac{\phi}{2}.$$

Exercise. Show that $\sum_{m=-j}^{j} e^{im\phi} = \frac{\sin\left([2j+1]\frac{\phi}{2}\right)}{\sin\frac{\phi}{2}}$ for j half integer.

Solution Set $z = e^{i\phi}$.L. H. S. is the geometric series

$$\begin{aligned} z^j + z^{j-1} + \cdots z^{-j} &= z^{-j}\left[1 + z + \cdots z^{2j}\right] \\ &= z^{-j}\frac{1 - z^{2j+1}}{1 - z} = \frac{z^{-j} - z^{j+1}}{1 - z} \\ &= \frac{z^{-(j+\frac{1}{2})} - z^{j+\frac{1}{2}}}{z^{-\frac{1}{2}} - z^{\frac{1}{2}}} = \frac{e^{-(j+\frac{1}{2})i\phi} - e^{(j+\frac{1}{2})i\phi}}{e^{-\frac{1}{2}i\phi} - e^{\frac{1}{2}i\phi}} \\ &= \frac{\sin\left(\left[j+\frac{1}{2}\right]\phi\right)}{\sin\frac{\phi}{2}} \end{aligned}$$

which is the claimed answer.

6.2.2. *Character of a reducible representation*

The trace of a direct sum of matrices is the sum of traces. So, the character of a reducible representation is the sum of the characters of its irreducible components, weighted by the multiplicity. A reducible representation can be decomposed as a direct sum over irreducible representations:

$$R = \bigoplus_{j=0,\frac{1}{2},1,\frac{3}{2},\ldots} N_j D^j$$

N_j is zero if D^j does not occur at all in the decomposition. It can be greater than one if the same representation appears multiple times in the decomposition. We will mostly consider finite dimensional representations, so only a finite number of N_j will be non-zero and they are all finite.[1]

[1]Occasionally we will consider an infinite dimensional representation such as the space of functions on S^2. Its decomposition is

$$C\left(\mathbb{S}^2\right) = \bigoplus_{l=0,1,2,\ldots}^{\infty} D^l$$

That is, each integer value of l occurs exactly once; half integer values do not occur at all. (This is a "multiplicity free" representation.)

For example, if $N_1 = 2, N_{\frac{5}{2}} = 3$ the matrix $R(g)$ would be 2 blocks of $D^1(g)$, followed by three copies of $D^{\frac{5}{2}}(g)$ along the diagonal, with zeroes filling out the rest. Altogether it will have dimension $2 \times (2 \times 1 + 1) + 3 \times (2 \times \frac{5}{2} + 1) = 24$:

$$R(g) = \begin{pmatrix} D^1(g) & 0 & 0 & 0 & 0 \\ 0 & D^1(g) & 0 & 0 & 0 \\ 0 & 0 & D^{\frac{5}{2}}(g) & 0 & 0 \\ 0 & 0 & 0 & D^{\frac{5}{2}}(g) & 0 \\ 0 & 0 & 0 & 0 & D^{\frac{5}{2}}(g) \end{pmatrix}$$

The character of such a representation is the sum of the traces of the matrices along the block diagonals:

$$\chi_R(\phi) = \sum_{j=0,\frac{1}{2},1,\frac{3}{2},\ldots} N_j \chi_j(\phi)$$

Since $\chi_j(\phi)$ is a ratio of sines, this is very much like a Fourier series. Indeed the functions χ_j satisfy orthogonality relations

$$\int_0^{4\pi} \chi_j(\phi)\chi_k(\phi) \sin^2 \frac{\phi}{2} d\phi = 2\pi\delta_{jk}$$

Since the $\sin^2 \frac{\phi}{2}$ cancel with the $\sin \frac{\phi}{2}$ factors in the denominator of $\chi_j\chi_k$, this follows from the familiar orthogonality of the Fourier sine series. The meaning is that inequivalent irreducible representations have orthogonal characters.

Remark 18. The factor $\sin^2 \frac{\phi}{2}$ in the measure has a simple geometrical meaning: It is proportional to the area of the set of all elements in $SU(2)$ that can be brought to the form $e^{\frac{i}{2}\sigma_3\phi}$ by a rotation of the axis itself. This is a 2-sphere, embedded inside $SU(2) \approx \mathbb{S}^3$. Its area can be calculated using the non-Euclidean metric of $\mathbb{S}^3$. Our proof of orthogonality does not use this fact,however.

The main point is that the character function uniquely determines the multiplicities, through its "Fourier" series:

$$N_j = \int \chi_R(\phi)\chi_j(\phi) \sin^2 \frac{\phi}{2} \frac{d\phi}{2\pi}.$$

So, to show that two finite dimensional representations are isomorphic, it is enough to show that they have the same character. This also justifies the name "character" for χ_R: It does characterize the representation.

6.2.3. *Decomposition of the direct product*

Now, consider the character of the direct product representation

$$\chi_{j_1\otimes j_2}(\phi) = \mathrm{tr} D^{j_1}\left(e^{\frac{i}{2}\sigma_3\phi}\right)\otimes D^{j_2}\left(e^{\frac{i}{2}\sigma_3\phi}\right)$$

The trace of a direct product is the product of its traces. So

$$\chi_{j_1\otimes j_2}(\phi) = \frac{\sin\left([2j_1+1]\,\frac{\phi}{2}\right)}{\sin\frac{\phi}{2}}\frac{\sin\left([2j_2+1]\,\frac{\phi}{2}\right)}{\sin\frac{\phi}{2}}$$

On the other hand, the trace of a direct sum of matrices is the sum of the traces. So, the character of $\bigoplus_{j=|j_2-j_1|}^{j_1+j_2} D^j$ is

$$\sum_{j=|j_2-j_1|}^{j_1+j_2}\frac{\sin\left([2j+1]\,\frac{\phi}{2}\right)}{\sin\frac{\phi}{2}}$$

Some trigonometry (or a line of Mathematica code) will allow you to prove the identity

$$\frac{\sin\left([2j_1+1]\,\frac{\phi}{2}\right)}{\sin\frac{\phi}{2}}\frac{\sin\left([2j_2+1]\,\frac{\phi}{2}\right)}{\sin\frac{\phi}{2}} = \sum_{j=|j_2-j_1|}^{j_1+j_2}\frac{\sin\left([2j+1]\,\frac{\phi}{2}\right)}{\sin\frac{\phi}{2}}$$

showing the equality of characters of $D^{j_1}\otimes D^{j_2}$ and $\bigoplus_{j=|j_2-j_1|}^{j_1+j_2} D^j$. We restate this result:

Theorem. $D^{j_1}\otimes D^{j_2} \approx \bigoplus_{j=|j_2-j_1|}^{j_1+j_2} D^j$

It is possible to go further and explicitly construct the unitary transformation that relates the basis of $D^{j_1}\otimes D^{j_2}$

$$|m_1 m_2>, \quad m_1 = j_1,\ldots,-j_1, \quad m_2 = j_2,\ldots,-j_2$$

to the basis of D^j:

$$|jm>, \quad j = j_1+j_2, j_1+j_2-1,\ldots,|j_2-j_1|, \quad m = j, j-1,\ldots,-j$$

The matrix elements of this unitary transformation are called Clebsch–Gordon coefficients [9]. They are useful in many detailed calculations involving atomic and nuclear transitions. We worked this out for $j_1 = \frac{1}{2} = j_2$. The general case is (as always) in the book by Landau and Lifshitz [9]. We won't go down this particular rabbit hole.

6.3. The Power of Spinors*

6.3.1. *The symmetric powers of the $D^{\frac{1}{2}}$ representation gives all other representations*

In this sense the $D^{\frac{1}{2}}$ or spin half representation is the "fundamental" representation. This can be understood by repeatedly using the above reduction procedure:

$$D^{\frac{1}{2}} \otimes \left(D^{\frac{1}{2}} \otimes D^{\frac{1}{2}} \right) = D^{\frac{1}{2}} \otimes D^0 \oplus D^{\frac{1}{2}} \otimes D^1 = 2D^{\frac{1}{2}} \oplus D^0 \oplus D^{\frac{3}{2}}$$

etc. But this cumbersome. There is a more direct approach, using polynomials.

6.3.2. *A spinor is a pair of complex numbers carrying a representation of $SU(2)$*

This is supposed to rhyme with vector.

$$z = \begin{pmatrix} z^1 \\ z^2 \end{pmatrix}$$

Under $SU(2)$ a spinor transforms linearly:

$$z \mapsto gz, \quad g \in SU(2)$$

6.3.3. *The space of polynomials in a spinor carry a representation of $SU(2)$*

A polynomial of degree two in z is

$$\psi(z) = \psi_{11}(z^1)^2 + 2\psi_{12}z^1z^2 + \psi_{22}(z^2)^2$$

It can also be written as

$$\psi(z) = \psi_{ab}z^a z^b$$

Since z_a are complex numbers, $z^a z^b = z^b z^a$ and so $\psi_{ab} = \psi_{ba}$ is a symmetric matrix. Under an SU(2) transformation

$$z^a \mapsto g^a_c z^c$$

$$\psi_{ab} \mapsto \psi_{cd} g^c_a g^d_b$$

It is clear that the degree of a polynomial is unchanged under the $SU(2)$ action: The space of homogenous polynomials of degree two in z_a carries a representation of $SU(2)$. (Homogenous means that all the terms in the polynomial have the same degree.)

More generally, the space of homogenous polynomials of order n carries a representation as well. It is useful to get an inner product on this space. The idea is to multiply polynomials and integrate over z , with a measure (weight) chosen to make the integral converge:

$$\langle \psi, \chi \rangle = \int \psi^*(z)\chi(z)e^{-z^\dagger z}\frac{dzd\bar{z}}{\pi}$$

Clearly $z^\dagger z$ is invariant under g since $g^\dagger g = 1$. The factor of π is stuck in so that the constant function equal to one has norm one. The constant function gives, of course, the trivial representation.

Theorem. *The inner product above is invariant under the action*

$$M(g)\psi(z) = \psi(g^{-1}z)$$

of $SU(2)$. *That is, we have a unitary representation.*

Proof. To verify that it is a representation, we calculate

$$\begin{aligned} M(g_1)[M(g_2)\psi](z) &= [M(g_2)\psi](g_1^{-1}z) = \psi(g_2^{-1}g_1^{-1}z) \\ &= \psi([g_1g_2]^{-1}z) = [M(g_1)M(g_2)]\psi(z) \end{aligned}$$

Now, you see why we needed to put g^{-1} in the definition of $M(g)$. To prove that the integral is invariant

$$\langle M(g)\psi, M(g)\chi \rangle = \int \psi^*(g^{-1}z)\chi(g^{-1}z)e^{-z^\dagger z}\frac{dzd\bar{z}}{\pi}$$

we make a change of variables[2] $z \mapsto gz$ (and using the Jacobian for the transformation of the volume element dz)

$$= \int \psi^*(z)\chi(z)e^{-z^\dagger g^\dagger g z}\det g \det g^\dagger \frac{dzd\bar{z}}{\pi}$$

and use $g^\dagger g = 1$ and $\det g = 1$. □

Exercise 4. Show that an orthonormal basis is given by the functions

$$|n_1, n_2 >= \frac{z_1^{n_1}z_2^{n_2}}{\sqrt{n_1!n_2!}}, \quad n_1 + n_2 = n$$

Hint The integral factorizes into separate gaussian integrals in z_1 and z_2. Each integral then is of the kind found in the theory of coherent states of the harmonic oscillator [18].

[2]Remember that we are thinking of $g \in SU(2)$ as a 2×2 matrix.

We can identify the representation by calculating its character. When $g = e^{-\frac{i}{2}\sigma_3\phi}$

$$g\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} e^{-\frac{i}{2}\phi} z_1 \\ e^{\frac{i}{2}\phi} z_2 \end{pmatrix}$$

and

$$M(g)|n_1, n_2 >= e^{-\frac{i}{2}(n_1 - n_2)\phi}|n_1, n_2 >$$

The character is the sum

$$\sum_{n_1=0}^{n} < n_1, n - n_1|M(g)|n_1, n - n_1 >= \sum_{n_1=0}^{n} e^{-\frac{i}{2}(2n_1 - n)\phi}$$

Summing this geometric series,

$$= e^{\frac{i}{2}n\phi}\frac{1 - e^{-i(n+1)\phi}}{1 - e^{-i\phi}} = \frac{e^{\frac{i}{2}(n+1)\phi} - e^{-\frac{i}{2}(n+1)\phi}}{e^{\frac{i}{2}\phi} - e^{-\frac{i}{2}\phi}}$$

$$= \frac{\sin\left[(n+1)\frac{\phi}{2}\right]}{\sin\frac{\phi}{2}}$$

which we recognize as the character of the spin $\frac{n}{2}$ representation.

Thus we conclude:

6.3.4. ***The homogenous polynomials of spinors degree n form the irreducible representation D^j with spin $j = \frac{n}{2}$.***

Chapter 7

ISOSPIN AND STRANGENESS

7.1. The Atomic Nucleus

7.1.1. *The positive charge of an atom is concentrated in a much smaller region than its negative charge*

The typical size of an atom (the range of its negative charge cloud) is about a Bohr radius, $\sim 10^{-10}\, m$. The positive charge (as well as most of the mass) is contained in a much smaller region (the nucleus) with a typical size of $\sim 10^{-15}\, m$. The latter size is a femtometer or a Fermi, abbreviated to fm. This observation is the result of the classical experiment of Rutherford, who scattered α particles (nuclei of the Helium atom) off Gold. Occasionally an alpha particle would get scattered through a wide angle. This can only happen if some charge (as well as mass) is concentrated in a very small region.

The Rutherford experiment disproved the popular atomic model of the time, the "plum pudding model": electrons embedded in a diffuse positive cloud. Scattering experiments have been central to nuclear/particle physics ever since: Throw particles against each other and see what comes out.

7.1.2. *The atomic nucleus contains a positively charged particle (the proton) as well as an electrically neutral particle (the neutron)*

The number of electrons in an atom is equal to the number of protons in its nucleus. Since the electron is negatively charged, with the same magnitude of charge as the proton, the atom is electrically neutral.

7.1.3. *The chemical properties of an atom depend only on the number of electrons (the atomic number Z)*

Chemical reactions involve exchange of electrons: They have energies of about an eV, which is roughly the binding energy of an electron in an atom. Nuclear reactions need an energy of about an MeV: About a million times more. This is why chemical reactions cannot turn lead into gold, the original aim of alchemists. Nuclear reactions can do this. But any dreams of getting rich this way are doomed to failure: The cost of such a transmutation is prohibitive.

The atomic number is equal to the number of protons in the nucleus (since the number of electrons and protons are equal). So, two nuclei with the same number of protons but different numbers of neutrons will have identical chemical properties. They are called *isotopes* . Neutrons and protons are particular cases of a class of particles known as *Baryons*.

7.1.4. *The sum of the number of protons plus the number of neutrons is called the Baryon number (or atomic mass number)*

So $B = N + Z$ where N is the number of neutrons in the nucleus. The mass of an atomic nucleus is approximately B in units of $GeV = 10^3 MeV$. This is because to first approximation the masses of the proton and neutron are equal to about a GeV. (We will be more precise soon).

Examples

- Hydrogen has atomic number one. The most abundant isotope has a nucleus consisting of just one proton. The next most abundant (0.01%) is deuterium, which has baryon number 2. There is also an isotope of baryon number 3 (tritium) which is unstable, with a half-life of 12.5 years.
- The abundant isotope of Helium has $Z = 2, B = 4$. Its nucleus is the alpha particle. Another stable isotope is ^3He which is a product of tritium decay.
- Oxygen which has several (about 15) isotopes. But only three are stable. Almost all the Oxygen in nature has $B = 16$.

	N	Z	Lifetime	Natural Abundance
H	1	1	∞	1
$D = ^2 H$	1	1	∞	10^{-4}
$T = ^3 H$	2	1	12.5 years	10^{-13}
^{16}O	8	8	∞	0.998
^{17}O	9	8	∞	10^{-4}
^{18}O	10	8	∞	10^{-3}
^{19}O	11	8	26 s	0

7.1.4.1. *The proton has a mass of $m_p = 938$ MeV and the neutron has mass $m_n = 939.5MeV$*

Thus, the neutron is just a little bit heavier than the proton. Their masses are too close to be mere coincidence.

7.1.5. *The particles inside a nucleus are held together by the strong interaction*

Without this force, the nucleus would disintegrate due to electric repulsion among the protons. Neutrons are essential for this strong binding to happen. If there are too few neutrons, the nucleus will fission or split up into smaller nuclei. The strong interaction has a binding energy of a few MeV (large compared to atomic energies). It has a range of 1 fm or 10^{-15}m (small compared to an atom). That is, the force decreases exponentially, with a decay constant of about $1\ fm$. This is why all the neutrons and protons don't clump together and form one gigantic nucleus. The size of a nucleus is roughly the same as the range of the strong interaction.

But if there are too many neutrons, some of then will decay by the beta decay (See below). This fine balance between the neutron and proton is behind many coincidences, which make our natural world (including life) possible. Some people even try to explain the values of elementary particle masses based on this (the anthropic principle.)

7.1.5.1. *During beta decay, a neutron converts itself to a proton and an electron*

$$n \rightarrow p + e + \bar{\nu}$$

These decays were among the first nuclear reactions to be discovered. It was a big step forward when J. J. Thompson discovered that "β radiation" consists of negatively charged particles. We call these particles electrons now. Also produced is an anti-neutrino $\bar{\nu}$ which is often hard to detect because it is electrically neutral. More on the neutrino later.

The condition for stability of a nucleus against beta decay is that $M(N-1, Z+1) - M(N, Z) < m_n + m_e \approx 938.58\,\text{Mev} \approx 1.01$ u. Beta decay increases atomic number by one unit and decreases the number of neutrons by one. But it leaves the baryon number unchanged.

7.1.5.2. *The binding energy of the deuteron is* 2.2 *MeV*

The binding energy is the energy needed to break up a nucleus into its constituents. In other words, it is the sum of the masses of the constituents minus its mass. So,

the deuteron has mass 2.2 MeV smaller than the sum of the masses of a proton and a neutron.

7.2. Isospin

Heisenberg introduced a new idea in the 1930s:

7.2.1. *The neutron and proton are different states of the same particle, the nucleon, with different values of a new quantum number called isospin*

Electromagnetism (charge and magnetic moment) and weak interactions responsible for beta decay are small effects in comparison to the nuclear force. The $n - p$ mass difference, is only about .2%. If we ignore these, the neutron and proton really do look like different states of the same particle.

7.2.1.1. *Since there are only two possible values for this new quantum number labeling the neutron and the proton, it is analogous to the spin of an electron.*

Isospin means 'like spin' in pidgin greek.

7.2.2. *Isospin is an approximate $SU(2)$ symmetry of nature*

This is an "internal" symmetry. Although analogous to rotations, it does not describe transformations in space-time. Some "internal space" carries this symmetry.

7.2.3. *The nucleon has spin half and isospin half*

That is under the $SU(2)\times SU(2)$ group of isospin and spin it transforms as $D^{\frac{1}{2}}\otimes D^{\frac{1}{2}}$. The state of a nucleon at rest is a four component complex vector.

7.2.4. *The nucleon is a fermion: A wavefunction of a collection of them must be antisymmetric under pairwise exchange*

So, for a pair of nucleons, the state is described by a 4×4 matrix. This matrix must be anti-symmetric because of the exclusion principle, having thus 6 independent states (if other degrees of freedom, like position, can be ignored). These can be grouped into sets of three states each that are of spin 1 and isospin zero or isospin one and spin zero.

7.2.5. *When a neutron and a proton combines into a deuteron, they form an isospin 0 state*

It turns out that the isospin 0 (hence spin one) state has lower energy. The details of the nuclear force is a complicated subject, even harder than molecular physics or chemistry. If it were not for important applications (nuclear energy, nuclear explosions) physicists would not spend so much time studying them. The larger the nucleus the more complicated its internal dynamics.

7.2.6. *The α particle is a spin zero and isospin zero state; it can be thought of as a bound state of four nucleons.*

The α particle is the nucleus of the abundant isotope of helium.

7.2.6.1. *It has a large binding energy: 28.3 MeV, so is very stable*

Whenever there are the right number of neutrons and protons to form an isospin zero state, the binding energy is unusually large: these are called the 'magic nuclei' and they are usually the stable end products of fission and fusion reactions.

7.2.7. *The electromagnetic interactions do not respect isospin symmetry*

In fact, for nucleons, $Q = I_3 + \frac{1}{2}$ where I_3 is isopin.

7.2.8. *The weak interactions also do not respect isospin symmetry. Nuclear beta decay treats the neutron and the proton differently*

7.2.9. *Thus isospin is a symmetry only of strong interactions, which are responsible for the binding of nucleons into nuclei*

A phenemenological formula of Weizsäcker for binding energy is found to be surprisingly accurate:

$$E(N,Z) = a_{vol}B - a_{surface}B^{\frac{2}{3}} - a_{Coul}\frac{Z(Z-1)}{B^{1/3}} - a_{sym}\frac{(N-Z)^2}{B}.$$

for some constants a. Here, $B = N + Z$ is the baryon number; i.e., the total number of nucleons.

The first term is proportional to the number of nucleons; the second to the surface area, as the density of nuclear matter is roughly constant. The third is the Coulomb repulsion and depends on the number of pairs of protons as well as the inverse of the average distance between them.

The last term is zero if you have equal numbers of neutrons and protons so that we can form an isospin zero combination. It can be explained by the postulate that the nuclear force is independent of isospin and spin states of the nucleons. This is related to the $SU(4)$ model of Wigner.

7.2.9.1. *Iron with $Z = 26, N = 30$ is one of the most tightly bound nuclei.*

7.3. The Pi Meson

7.3.1. *Yukawa suggested that the attractive force among nucleons is due to exchange of a massive particle, of mass $\mu \sim \frac{\hbar}{ac} \sim 100$ MeV*

7.3.1.1. *It is useful for conversions to note that $\hbar \approx 197 MeV\, fm$*

For simplicity we will for now ignore the fact that there are two kinds of particles (n and p) inside the nucleus. In the next section we will restore this doubling, using isospin symmetry.

7.3.1.2. *The Klein–Gordon equation with a point source has an exponential decreasing static solution $\phi = g\frac{e^{-\mu r}}{4\pi r}$.*

Here g is a constant (Yukawa coupling constant) that measures the strength of the field, analogous to electric charge for the Coulomb field. The Klein–Gordon equation is a modification of the wave equation

$$\left[\frac{\partial^2}{\partial t^2} - \nabla^2\right]\phi + \mu^2\phi = 0$$

We are using units with $\hbar = c = 1$. So μ has dimensions of mass, which is the same as length^{-1}.

The plane wave solutions $e^{i[\omega t - \boldsymbol{k}\cdot\boldsymbol{x}]}$ describe particles of frequency $\omega = \sqrt{\mu^2 + \boldsymbol{k}^2}$. By de Broglie formula, $\hbar\boldsymbol{k}$ is the momentum and $\hbar\omega$ is the energy. Then we recognize this as the energy of a particle of mass μ in relativity. (The wave equation is the special case $\mu = 0$ which describes massless particles.)

There are also a static (time independent) spherically symmetric solution. In polar co-ordinates, for a spherically symmetric function

$$\nabla^2\phi = \frac{1}{r}\frac{\partial^2}{\partial r^2}(r\phi)$$

so the KG equation becomes, with

$$\phi = \frac{1}{r}R$$
$$R'' + \mu^2 R = 0$$

The solution that vanishes at infinity is the "Yukawa potential"

$$R(r) = ke^{-\mu r} \implies \phi = \frac{k}{r}e^{-\mu r}$$

Thus, a massive particle yields a potential that decays exponentially. Its range is related inversely to the mass. Using the value $\hbar \approx 197$MeV fm and the size of the nucleus of about 1fm Yukawa predicted that there must be a particle of spin zero and mass ∼100 MeV.

7.3.1.3. *Similar to the photon which mediates the electromagnetic interactions, except the photon is massless and the Coulomb force has infinite range*

The special case $\mu = 0$ of the Yukawa potential is the Coulomb potential.

7.3.1.4. *Yukawa suggested that the particle mediating the strong interactions between nucleons is of spin zero*

At that time no massive spin zero particle was known. Because its mass should be in between that of the electron and the nucleon, it was called a meson (from the Greek for "particle in between'). The modern scientific definition of the meson is not related to its mass, but other properties. (See below).

7.3.2. *This particle has since been discovered and is called the π meson*

It has a mass of about 140 MeV. There was some confusion about its discovery. In fact another particle with a very close mass (105 MeV) was discovered first in cosmic rays, called the muon. But the muon did not get absorbed by nuclei. It was Robert Marshak who resolved the confusion: The muon is a lepton, a copy of the electron only with a higher mass. It has no strong interactions with the nuclei. But pions (which are created by cosmic ray collisions in the upper atmosphere) decay quickly into the muons, which are detected at lower altitudes.

There is no fundamental reason why the π and the μ have masses that are so close. Coincidences happen sometimes. Occam's razor is not always sharp.

7.3.3. *The pi meson has isospin one*

Thus, there are three possible isospin states: There are actually three pi mesons, with almost equal masses and electric charges $\pm 1, 0$.

$$\boldsymbol{\phi} = \begin{pmatrix} \pi^+ \\ \pi^0 \\ \pi^- \end{pmatrix}$$

For them the formula for electric charge is

$$Q = I_3.$$

There is no shift, unlike for the nucleons.

7.3.3.1. *To be precise, the mass of the charged pions are a few percent different from that of the neutral pion*

But we ignore that for now. The strong interactions are caused by exchanges of pions:

$$n \to p + \pi^-, \quad p \to n + \pi^+.$$

Because there may not be enough energy to create a free pion in a nucleus, the pions are often virtual: They exist only for a time of order $\frac{1}{\mu}$.

7.3.4. *We do understand the origin of isospin symmetry: quarks*

It turns out that the nucleons and pi mesons are both bound states of a more fundamental unit of matter (quarks and anti-quarks). There happen to be two quarks (called up and down) which have almost equal mass. The neutron is uud and the proton is udd . This explains why they are so similar, yet slightly different. The pi mesons are combinations of quarks and antiquarks; for example π^+ is $\bar{d}u$. More on the quark model later.

Why the u, d quarks have masses that are so close is still unknown: The next layer of the onion needs to be peeled to understand that.

7.4. Hadrons

In the 1930s it looked as though we were on the verge of a simple description of the fundamental constituents of matter: The proton, the neutron, the pion, the electron (and possibly the neutrino) along with the photon would make up all matter. In the

1940s the muon was identified. (I. I. Rabi famously asked "Who ordered that?") Throughout 1940s, 1950s and 1960s experimentalists discovered a whole zoo of strongly interacting particles.

7.4.1. *Strongly interacting particles are collectively known as hadrons*

Electrons, neutrinos etc. are leptons, not hadrons. Neutrons, protons and pions are hadrons.

7.4.2. *Hadrons of half integer spin are called baryons; those of integer spin are called mesons*

7.4.2.1. *The baryon number B is defined to be equal to 1 for the half integer spin hadrons (baryons) and equal to zero for mesons*

Anti-baryons have baryon number minus one. The baryon number used to be called the "atomic mass number " in nuclear physics: The Deuteron has $B = 2$, the α particle has $B = 4$ and so on:

7.4.3. *The nucleon* $N = \binom{p}{n}$ *is the lightest baryon with a mass of about* $940MeV$

7.4.4. *The pion* $\pi = \begin{pmatrix}\pi^+\\ \pi^0\\ \pi^-\end{pmatrix}$ *is the lightest meson at about* $140MeV$

Now, we enter the zoo of hadrons.

7.4.5. *There is a set of four baryons* $\Delta = \begin{pmatrix}\Delta^{++}\\ \Delta^{+}\\ \Delta^{0}\\ \Delta^{-}\end{pmatrix}$ *of spin and isospin both equal to* $\frac{3}{2}$

The mass of the Δ is about 1230 MeV. (Again, the mass depends on the charge by a few percent which we ignore for now.)These decay into a nucleon and a pion. The nucleon has $I = \frac{1}{2}, J = \frac{1}{2}$ and the pion $I = 1, J = 0$. The strong decay respects both spin and isospin conservation. This restricts the possible decay probabilities.

7.4.6. *There is a set of three spin one mesons* $\rho = \begin{pmatrix}\rho^+\\ \rho^0\\ \rho^-\end{pmatrix}$ *of isospin 1*

Their mass is ≈ 770 MeV: About two thirds of the mass of a nucleon. They decay by strong interactions into pions.

7.4.7. ***The charge of any hadrons is related to isospin by the relation***

$$Q = I_3 + \frac{B}{2}.$$

The shift of charge by a constant for the case of baryons (but not mesons) is explained by the quark model (see below).

7.4.8. ***There are hadrons of spins*** $J = 0, \frac{1}{2}, 1, \frac{3}{2}, 2, \ldots$

7.4.8.1. *If the spin is a half integer, so is the isospin*

Again explained by the quark model.

7.4.8.2. *As the spin grows the masses grow approximately proportionately*

The high mass hadrons are more and more unstable to decay to lower mass ones. Such very unstable particles are called resonances. As the numbers of hadrons grew into the hundreds, physicists accepted that there must be in principle an infinite number of them.

7.4.9. ***String theory arose as an explanation for the infinitely rising spectrum of hadrons***

7.4.9.1. *A string is a curved surface in space time whose action is proportional to its area*

A constant time cross-section of a string is an curve in space (which is the origin of the name). The action being proportional to the area translates to an energy of this cross-section proportional to the length: A kind of "rubber band" model.

7.4.9.2. *Nambu and Goto showed that this implies that the masses of its excited states are proportional to the angular momentum*

The Nambu–Goto model only allowed integer spins. Theirs was a "bosonic string theory". But it turned out to be consistent only in 26 dimensions.

7.4.9.3. *Supersymmetry was invented by Ramond to include fermions*

The idea did not quite work: Superstring theory is consistent only in ten dimensions. No one has yet found a string theory that works in four space time dimensions.

Finding the correct string theory of hadrons remains an important theoretical challenge.

7.4.9.4. *The 10 dimensional version of superstring theory is a candidate for a quantum theory of gravity*

Sometimes a recipe that didn't work for one dish is perfect for a completely different one. Physics is a supremely rational subject. But the creative processes of discovery in physics remains intuitive, irrational and circuitous.

7.5. Quarks

7.5.1. *All of the hadrons are bound states of more elementary particles known as quarks*

This gives a very simple explanation for the proliferation of hadrons.

7.5.1.1. *Quarks have spin* $\frac{1}{2}$

This is reasonable for an elementary particle: It is the smallest non-zero value of spin allowed by quantum mechanics.

7.5.2. *Each quark has an anti-quark of the same mass, spin and isospin, but opposite charge*

7.5.2.1. *Mesons are bound states of quarks and anti-quarks*

Which explains why they are bosons. Any bound state of an even number of fermions is a boson.

7.5.2.2. *Baryons contain three quarks*

It must be an odd number since baryons are fermions. Since there are baryons of spin $\frac{3}{2}$ and even parity (so that the orbital angular momentum is even) we need there to be three quarks in a baryon. It follows that

7.5.2.3. *The baryon number of a quark is* $\frac{1}{3}$

Anti-quarks have $B = -\frac{1}{3}$.

7.5.3. *There are a pair of quarks* $\begin{pmatrix} u \\ d \end{pmatrix}$ *forming an isospin* $\frac{1}{2}$ *system.*

This explains why spin and isospin are equal for the lowest lying baryons.

7.5.3.1. *Their charges are given by* $Q = I_3 + \frac{B}{2}$

$$Q_u = \frac{2}{3}, \quad Q_d = -\frac{1}{3}.$$

This then explains why this formula holds for all hadrons. Note that the piece proportional to B cancels out in mesons since anti-quarks have $B = -\frac{1}{3}$.

7.5.3.2. *Some group theory will allow us to get the spins and isospins of the hadrons out of those of the quarks.*

7.5.4. *But there is a surprise: color*

We will explain what color is soon.

We always push ideas to their limit. But always be prepared to test everything experimentally, because great ideas are often wrong. And there are digressions that were never anticipated (e.g., muon).

7.6. The Static Quark Model

7.6.1. *A first approximation is to treat the hadrons as non-relativistic bound states of quarks*

So the different spin states have the same energy (spin-orbit coupling is a relativistic correction). This gives an $SU(4)$ symmetry: The four states of the quarks (spin up and down, isospin up and down) all have the same energy in this approximation.

7.6.2. *Quarks are fermions*

Spin half particles should satisfy the exclusion principle.

7.6.3. *But then how do we explain the* Δ^{++}*?*

The Δ^{++} must be made of three u quarks, to explain its electric charge. But it also has spin $\frac{3}{2}$: The three u quarks must be in a state symmetric under interchanges. For example, in the state where Δ^{++} has $J_3 = \frac{3}{2}$, all the $u$quarks in it must has spin pointing in the same direction.

But this violates the exclusion principle: Quarks are fermions. One idea to resolve this contradiction was that quarks obeyed some exotic statistics that violates the Pauli exclusion principle. That turned out to be the wrong direction. Another possibility is that quarks have an extra degree of freedom, in addition to spin and isospin.

7.6.4. *Each quark comes in three colors*

Thus there are three states for the up quark (not counting the spin states) and three for the down quark.The word color is used in a figurative way here: This quantum number has nothing at all to do with visible light: Nothing to do with electromagnetism.

7.6.5. *There is an SU(3) symmetry corresponding to rotations among the color states*

Quarks of different colors (but identical in every other way) have the same masses, isospin, charges etc.

7.6.6. *Baryons and Mesons are color neutral*

Nucleons and mesons do not have this extra degree of freedom: We would have seen this in nuclear physics. These states are invariant under the color SU(3) symmetry. This means that color cannot be directly measured: It can be inferred indirectly from properties of mesons and baryons.

7.6.7. *The ground state of a three quark system must be a symmetric combination of three fundamental representations of SU* **(4)**

The wave function of quarks in a baryon is completely antisymmetric in color: That is the way to make it invariant under $SU(3)$ symmetry of color (more on $SU(3)$ later). The wave function of fermions is anti-symmetric overall. Thus, in spin and isospin it must be symmetric. (We are assuming that the orbital angular momentum is zero for the lowest lying states). Recall the in quantum mechanics, symmetric combination usually arise when we combine bosons.

7.6.8. *It is useful to consider the states of a meson and a baryon in the static quark model with N_c colors and N_f flavors*

Quark model that originated in Nuclear Physics had just two kinds of quarks (u, d to explain isospin). Later it had to be enlarged to allow for other kinds of quarks. The

final number appears to be six u, d, s, c, t, b . The first three u, d, s have relatively small masses so are treated together (see below). The jargon of particle physics is to call this flavor (to contrast with color?).

The wavefunction of a $\bar{q}q$ system would depend on the color and flavor quantum numbers of the quark and the antiquark. We can denote it by $\psi^{\beta B}_{\alpha A}$. Here α, β are the color indices and A, B all the remaining quantum numbers (spin, isospin etc.). So $\alpha = 1, 2, \cdots N_c$ while $A = 1, 2, \ldots, 2N_f$, because each flavor of quark has two possible spin states.

In nature the color ranges over three possible values; $\alpha = 1, 2, 3$. Under a color $SU(3)$ transformation the wave function of a quark transforms as $q^{\alpha B} \mapsto g^{\alpha}_{\beta} q^{\beta B}$ and that of an anti-quark as $\bar{q}_{\alpha A} \mapsto g^{*\gamma}_{\alpha} \bar{q}_{\gamma A}$. Then the $\bar{q}q$ state transforms as

$$\psi^{\beta B}_{\alpha A} \mapsto g^{*\mu}_{\alpha} g^{\beta}_{\nu} \psi^{\nu B}_{\mu A}$$

But $g^{*\mu}_{\alpha} g^{\alpha}_{\nu} = \delta^{\mu}_{\nu}$ because g is unitary. So, $\psi^{\alpha B}_{\alpha A}$ is color invariant:

$$\psi^{\alpha B}_{\alpha A} \mapsto \delta^{\mu}_{\nu} \psi^{\nu B}_{\mu A} = \psi^{\mu B}_{\mu A}$$

This works even if the number of colors were some other number N_c and the color symmetry $SU(N_c)$. It is a useful exercise to consider what the world would have been like with N_c colors (According to G. 't Hooft, even the limit $N_c \to \infty$ gives us much insight into strong interactions.) Baryons are symmetric under the interchange of spin and flavor indices, even in a model with $SU(N_c)$ symmetry.

Exercise. Show that the condition $\det g = 1$ becomes, when written in terms of indices,

$$\epsilon_{\alpha_1 \alpha_2 \cdots \alpha_{N_c}} g^{\alpha_1}_{\beta_1} g^{\alpha_2}_{\beta_2} \cdots g^{\alpha_{N_c}}_{\beta_{N_c}} = \epsilon_{\beta_1 \beta_2 \cdots \beta_n}$$

where $\epsilon_{\alpha_1 \alpha_2 \cdots \alpha_{N_c}}$ is the completely anti-symmetric tensor with $\epsilon_{12 \cdots N_c} = 1$.

Solution It is enough to consider the case where $\beta_1 = 1, \beta_2 = 2 \ldots$. We can regard α as a permutation of the set $1, 2, \ldots, N_c$ and $\epsilon_{\alpha_1 \alpha_2 \cdots \alpha_{N_c}} = \pm 1$ according to whether α is an even or and off permutation. The LHS is the sum over all such permutations. This is one of the ways of calculating the determinant; so the above condition is equivalent to $\det g = 1$.

The wavefunction of a system of N_c quarks will depend on spin and flavor :$\Psi^{\alpha_1 A_1 \alpha_2 A_2 \cdots \alpha_{N_c} A_{N_c}}$. Quarks being fermions this has to be completely anti-symmetric and the interchange of color, spin and flavor:

$$\Psi^{\alpha_1 A_1 \alpha_2 A_2 \cdots \alpha_{N_c} A_{N_c}} = -\Psi^{\alpha_2 A_2 \alpha_1 A_1 \cdots \alpha_{N_c} A_{N_c}}$$

etc. To make a color invariant combination we can contract the color indices with the anti-symmetric tensor

$$\epsilon_{\alpha_1\alpha_2\cdots\alpha_{N_c}}\Psi^{\alpha_1 A_1\alpha_2 A_2\cdots\alpha_{N_c}A_{N_c}} = S^{A_1A_2\cdots A_{N_c}}$$

Thus the wavefunction of a baryon will be *completely symmetric* in the spin and flavor indices.

7.6.8.1. *If we ignore color, the baryon wavefunction behaves as though the quarks are bosons*

This is a cheap trick; by no means a fundamental theory of strong interactions. Such a theory has been found: Quantum Chromo Dynamics (QCD); we are just not ready to talk about it yet. The calculation of baryon properties from QCD is a major undertaking that uses the most advanced computers available. What we study here is an approximate picture of quark bound states, called the *static quark model*. It ignores all the effects of interactions between quarks. It is still able to explain some static properties of baryons.

Let us digress to recall some facts about symmetric states.

7.6.8.2. *The number of independent states of a system of N bosons, each with M states, is* $\frac{M(M+1)\cdots(M+N-1)}{N!} = \binom{N+M-1}{M-1}$

There are many ways to establish this combinatorial formula.

To begin with, it holds for $M = 1$. There is exactly one independent way of occupying a single state with N bosons : We put them all into that one available state. More generally, suppose we put N_1 bosons in state 1, N_2 in state 2, and so on. The number of independent states $b_N(M)$ we seek is just the number of solutions to the equation

$$N = N_1 + N_2 + \cdots + N_M$$

where each $N_i = 0, 1, \ldots$. Thus the generating function is

$$\sum_{N=0}^{\infty} b_N(M)x^N = \sum_{N_i=0}^{\infty} x^{N_1+N_2+\cdots+N_n}$$

But

$$\sum_{N_i=0}^{\infty} x^{N_1+N_2+\cdots+N_n} = \sum_{N_1=0}^{\infty} x^{N_1}\sum_{N_2=0}^{\infty} x^{N_2}\cdots\sum_{N_M=0}^{\infty} x^{N_M} = \frac{1}{(1-x)^M}$$

since each factor is a geometric series. Expanding the r.h.s. as a binomial series

$$(1-x)^{-M} = \sum_{N=0}^{\infty} \frac{M(M+1)\cdots(M+N-1)}{N!} x^N = \sum_{N=0}^{\infty} \binom{N+M-1}{M-1} x^N$$

we see the result.

7.6.9. *The special case of two flavors and three colors is most useful*

Since each flavor of quark has two spin states, $M = 4$; and the number of quarks in a baryon $N = 3$, the number of colors.

So there are $\frac{4(4+1)(4+2)}{3!} = 20$ states for a baryon. These can be split into $I = \frac{3}{2}, J = \frac{3}{2}$ and $I = \frac{1}{2}, J = \frac{1}{2}$ states. The first are the Δ and the second set are the nucleons. There are $4 \times 4 = 16$ states for the Δ and $2 \times 2 = 4$ states for the nucleon which do add up to twenty.

7.6.9.1. * *With two flavors and an odd number N_c of colors there would have been baryons with $I = J$ for $I = \frac{1}{2}, \frac{3}{2}, \ldots, \frac{N_c}{2}$*

Exercise. Show that

$$\sum_{I=\frac{1}{2},\frac{3}{2},\ldots,\frac{N_c}{2}} (2I+1)^2 = \frac{(N_c+1)(N_c+2)(N_c+3)}{3!} = \binom{N_c+3}{3}$$

7.6.10. *The states of a bosonic system can be represented as polynomials in complex variables*

Given a symmetric tenor $S^{A_1 A_2 \cdots A_N}$ we can construct a polynomial of degree N in complex variables $z_A, A = 1, \ldots, M$:

$$S^{A_1 A_2 \cdots A_N} z_{A_1} \cdots z_{A_N}$$

Conversely, the coefficients of a homogenous degree N polynomial can be arranged into a symmetric tensor. The symmetry follows from the fact that the components of the z^A commute.

The degree N of the polynomial is the number of bosons; the number of independent variables M is the number of states available to each boson. Thus 1 is the empty state (vacuum) z_A $A = 1, \ldots, M$ are the one particles, and so on. There are $\frac{M(M+1)}{2}$ independent components in a symmetric tensor of rand two. Another way of understanding this is that there are $\frac{M(M+1)}{2}$ independent quadratic homogenous polynomials in M variables. More generally suppose the first variable

z_1 appears n_1 times, the second one appears $n_2$times and so on. Then a basis for polynomials of order N is

$$\frac{z_1^{n_1}}{\sqrt{n_1!}}\frac{z_2^{n_2}}{\sqrt{n_2!}}\cdots\frac{z_M^{n_M}}{\sqrt{n_M!}}, \quad n_1+n_2\cdots n_M = N \tag{7.6.1}$$

The factorials in the denominator are put in to make sure that this is a state of length one.

This is the coherent state description of a harmonic oscillator; the connection to bosonic states is important also in quantum optics (Klauder and Sudarshan). It is part of a general technique known as second quantization.See the next chapter for a more detailed discussion.

A more convenient description often is:

7.6.11. *The inner product on the space of polynomials can be expressed as a gaussian integral*

$$||\psi||^2 = \int |\psi(z)|^2 e^{-\bar{z}^A z_A}\frac{d^M z d^M \bar{z}}{\pi^M}$$

For the basis above, the integral will split into separate integrals over each variable $z_1, z_2, \ldots$. Each of them can then be calculated by transforming to polar co-ordinates.

Exercise 19. Show that the states (7.6.1) is orthonormal in this inner product

7.6.12. *For a fundamental spin half particle the magnetic moment is the Bohr magneton* $\frac{e\hbar}{2m}$

This is a consequence of the Dirac equation (which we will study in more detail later) minimally coupled to the electromagnetic field.

Here e is the electric charge of the particle and m is its mass. For example the magnetic moment of the electron is measured to be

$$\mu_e = -9.2847646917(29)\times 10^{-24} JT^{-1}.$$

For comparison, its Bohr magneton is determined by knowing the mass:

$$\frac{e\hbar}{2m_e} = 9.2740100657(29)\times 10^{-24} JT^{-1}$$

The negative sign of μ_e is because the electron is negatively charged. Similarly for the muon

$$\mu_\mu = -4.49044830(10) \times 10^{-26} JT^{-1}$$

and its magneton is $\left(\text{recalling that the mass ratio is } \frac{m_e}{m_\mu} = \frac{1}{206.768}\right)$

$$\frac{e\hbar}{2m_\mu} = \frac{m_e}{m_\mu}\frac{e\hbar}{2m_e} = 4.48523 \times 10^{-26} JT^{-1}$$

The tiny discrepancy can be explained as due to radiative corrections in Quantum Electrodynamics.

But this does not work for the proton or neutron.

7.6.13. *The proton and the neutron have anomalous magnetic moments*

The neutron is electrically neutral, so if it were an elementary particle, it would not have any magnetic moment at all! Instead it has a magnetic moment oriented opposite to its spin (i.e., negative).

For example, $\mu_p = 1.41060679545(60) \times 10^{-26} JT^{-1}$ and $\mu_n = -9.6623653(23) \times 10^{-27} JT^{-1}$. The nucleon's Bohr magneton $\frac{e\hbar}{2m_p} = 5.0507837393(16) \times 10^{-27} JT^{-1}$. (These values are from the NIST data base.)

So $\mu_n \approx 2.8\frac{e\hbar}{2m_p}$ and $\mu_n \approx -1.9\frac{e\hbar}{2m_p}$.

7.6.13.1. *The magnetic moments of the proton and neutron are roughly 3 and −2 times the nuclear magneton*

The magnetic moments were among the first indications that the neutron and proton were not elementary particles.

7.6.14. *The static quark model explains the anomalous magnetic moments of the nucleons*

In this model the quarks are elementary particles and the neutron and proton are bound states of them. The accuracy is about 10%. This is remarkable, because the static model ignores the binding mechanism or any relativistic effect.

We now present a calculation of the nucleon magnetic moment in the static quark moment.It is easiest conceptually to do this calculation for an arbitrary odd number of colors N_c; we can put $N_c = 3$ towards the end. So the baryon number of a quark is $\frac{1}{N_c}$. Neglecting the binding energy and kinetic energy as well as spin/isospin dependence,

7.6.14.1. *In the static quark model, the mass of the quark is approximately* $\frac{1}{N_c}$ *of the nucleon mass*

If we denote by τ_3 the Pauli matrix of isospin,

7.6.14.2. *The charge of a quark is* $Q = e\left[\frac{B}{2} + I_3\right] = e\left[\frac{1}{2N_c} + \frac{\tau_3}{2}\right]$

This formula will give the right answer when $N_c = 3$: for the up quark, $Q = \frac{2}{3}$ and for the down quark $Q = -\frac{1}{3}$.

For each quark (a runs over $1, 2 \cdots N_c$ since there are N_c quarks in the baryon) the component of magnetic moment along some direction (say third) is

$$\mu_a = \frac{e\hbar}{\frac{2m}{N_c}}\left[\frac{1}{2N_c} + \frac{\tau_{3a}}{2}\right]\sigma_{3a} = \frac{e\hbar}{4m}\left\{\sigma_{3a} + N_c\tau_{3a}\sigma_{3a}\right\}.$$

Here $\frac{1}{2}\sigma_{3a}$ is the spin of the ath quark. We are treating the quarks as fundamental particles and using the Dirac prediction for its magnetic moment.

We must sum over all the quarks in a baryon to get its magnetic moment

$$\mu = \frac{e\hbar}{4m}\sum_{a=1}^{N_c}\left\{\sigma_{3a} + N_c\tau_{3a}\sigma_{3a}\right\}$$

The first term is independent of isospin and the second depends on isospin.

$$\mu = \frac{e\hbar}{4m}\left(\mu_0 + N_c\mu_1\right), \quad \mu_0 = \sum_a \sigma_{3a} = J_3, \quad \mu_1 = \sum_{a=1}^{N}\tau_{3a}\sigma_{3a}$$

We need the matrix elements for the neutron and the proton. The static quark model ignores all degrees of freedom (such as position) **except** spin, isospin and color of the quarks. The baryon state must be invariant under color, so must be completely anti-symmetric in color. Quarks being fermions, this means in spin and isospin the state must be symmetric. It is as if the static quarks (devoid of color) are bosons with four independent states (two for spin and two for isospin). We can conveniently think of the four variables as the elements of a 2×2 matrix. Under isospin and spin this matrix transforms as

$$z \mapsto gzh^\dagger \; g, h \in SU(2)$$

The left action is isospin and the right action is spin (say). Now, we can express the spin and isospin operators as differential operators acting on these polynomials. Since spin acts on the right

$$z \overset{\sigma_3}{\mapsto} z\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix} = \begin{pmatrix}z_{11} & -z_{12}\\ z_{21} & -z_{22}\end{pmatrix}$$

Also under the action of isospin and spin (acting on the left and right respectively)

$$z \overset{\tau_3\sigma_3}{\mapsto} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} z \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} z_{11} & -z_{12} \\ -z_{21} & z_{22} \end{pmatrix}$$

7.6.14.3. *The states of the baryon are in one-one correspondence with N_cth degree polynomials in four complex variables $z_{11}, z_{12}, z_{21}, z_{22}$*

The norm of a polynomial is given by a gaussian integral of the last section:

$$||\psi||^2 = \int |\psi(z)|^2 e^{-\mathrm{tr}\, z^\dagger z} \frac{d^4 z}{\pi^2}$$

In our current notation, an orthonormal basis is

$$\frac{z_{11}^{n_{11}}}{\sqrt{n_{11}!}} \frac{z_{12}^{n_{12}}}{\sqrt{n_{12}!}} \frac{z_{21}^{n_{21}}}{\sqrt{n_{21}!}} \frac{z_{22}^{n_{22}}}{\sqrt{n_{22}!}} \tag{7.6.2}$$

with

$$n_{11} + n_{12} + n_{21} + n_{22} = N_c.$$

The matrices $\sum_a \sigma_{3a}$ and $\sum_{a=1}^{N} \tau_{3a}\sigma_{3a}$ can be written as differential operators acting on the polynomials:

$$\mu_0 = z_{11}\frac{\partial}{\partial z_{11}} + z_{21}\frac{\partial}{\partial z_{21}} - z_{12}\frac{\partial}{\partial z_{12}} - z_{22}\frac{\partial}{\partial z_{22}}$$

$$\mu_1 = z_{11}\frac{\partial}{\partial z_{11}} + z_{22}\frac{\partial}{\partial z_{22}} - z_{12}\frac{\partial}{\partial z_{12}} - z_{21}\frac{\partial}{\partial z_{21}}$$

7.6.14.4. *The polynomials that describe the nucleon states are $z(\det z)^k$ where $N_c = 2k + 1$*

Under isospin and spin, the four nucleon states transform with $I = J = \frac{1}{2}$. That is, they transform just like the four matrix elements of z. But, the nucleon contains N_c quarks, so the states must be polynomials of order N_c. We need to multiply z by some polynomial of degree $N_c - 1$ that is invariant under spin and isospin to get the nucleon states. When $N_c = 2k + 1$ is odd, that polynomial is $(\det z)^k$: Recall that $\det z$ is a quadratic polynomial in its matrix elements.

The magnetic moment operator μ commutes with J_3 (spin) and I_3 (isospin). So, we look at the effect of μ on eigenstates of these operators.

The spin up states of the proton and the neutron are[1]

$$|p\rangle = z_{11}(\det z)^k$$

$$|n\rangle = z_{21}(\det z)^k$$

It is important to note that neither the proton nor the neutron is an eigenstate of μ; what we call the magnetic moment of the proton or neutron is the expectation value of μ in these states. For example, μ can take a nucleon to some other baryon state; for example, it can mix $\mid p\rangle$ with $\mid \Delta^+\rangle$ or $\mid n\rangle$ with $\mid \Delta^0\rangle$. These lead to electromagnetic decays such as $\Delta^+ \to p\gamma$ or $\Delta^0 \to n\gamma$. We could use the static quark model to also predict these "transition magnetic moments". But it will take us too far afield.

To get just the magnetic moments of the nucleons we need the matrix elements $\langle p \mid \mu \mid p\rangle$, $\langle n \mid \mu \mid n\rangle$ since $\langle p \mid \mu \mid n\rangle = 0$; (Recall that μ commutes with I_3). Also the norms $\langle p \mid p\rangle$ and $\langle n \mid n\rangle$, which are the same.

7.6.14.5. *Calculation of the expectation values of μ when $N_c = 3$*

To proceed further it is convenient to specialize to the physically relevant value[2] $N_c = 3$. Since $k = 1$

$$|p\rangle = z_{11}\det z = z_{11}(z_{11}z_{22} - z_{12}z_{21}) = z_{11}^2 z_{22} - z_{11}z_{12}z_{21}$$

$$|n\rangle = z_{21}\det z = z_{21}(z_{11}z_{22} - z_{12}z_{21}) = z_{21}z_{11}z_{22} - z_{12}z_{21}^2$$

Since we know the orthonormal basis (7.6.2) we can find the norms of these states:

$$\langle p \mid p\rangle = 2 + 1 = 3, \quad \langle n \mid n\rangle = 1 + 2 = 3$$

Also

$$\mu_0 \mid p\rangle = \mid p\rangle, \quad \mu_0 \mid n\rangle = \mid n\rangle$$

and

$$\mu_1 \mid p\rangle = z_{11}(z_{12}z_{21} + 3z_{11}z_{22}), \quad \mu_1 \mid n\rangle = z_{21}(3z_{12}z_{21} + z_{11}z_{22})$$

Thus,

$$\langle p \mid \mu_1 \mid p\rangle = (-1 + 3 \times 2) = 5$$

$$\langle n \mid \mu_1 \mid n\rangle = (-3 \times 2 + 1) = -5$$

[1] Note that these states are nor normalized to have length one.

[2] It is of some interest to calculate the magnetic moments in the static quark model for arbitrary N_c for comparison with other models for nucleons such as the Skyrme model. But it is a bit harder. Could be an interesting research project for an enterprising student.

Since $\langle p \mid p\rangle = 3 = \langle n \mid n\rangle$ we get the expectation values

$$\frac{\langle p \mid \mu_0 \mid p\rangle}{\langle p \mid p\rangle} = 1 = \frac{\langle n \mid \mu_0 \mid n\rangle}{\langle n \mid n\rangle}$$

$$\frac{\langle p \mid \mu_1 \mid p\rangle}{\langle p \mid p\rangle} = \frac{5}{3}, \quad \frac{\langle n \mid \mu_1 \mid n\rangle}{\langle n \mid n\rangle} = -\frac{5}{3}$$

Using $\mu = \frac{e\hbar}{4m}(\mu_0 + 3\mu_1)$ when $N_c = 3$ we get

$$\frac{\langle p \mid \mu \mid p\rangle}{\langle p \mid p\rangle} = \frac{e\hbar}{4m}(1+5) = 3\frac{e\hbar}{2m}$$

$$\frac{\langle n \mid \mu \mid n\rangle}{\langle n \mid n\rangle} = \frac{e\hbar}{4m}(1-5) = -2\frac{e\hbar}{2m}$$

This agrees with experiment to about 10%.

7.6.14.6. *The agreement with observation is remarkable, considering we have completely ignored all interactions among quarks*

You can also check directly that $\langle p \mid \mu \mid n\rangle = 0$.

7.7. K mesons

7.7.1. *$K^{\pm}$ are pseudo-scalar, isospin $\frac{1}{2}$ particles of mass* 494 *Mev that only decay by weak interactions*

They were called "strange particles" when they were discovered. What was strange about them is that they were unusually long lived ($10^{-8}s$): Suggesting that they carry a new quantum number that is approximately conserved. This number was called "strangeness" (Gell–Mann). They are each other's anti-particles. K^+ was assigned strangeness $S = +1$ and therefore K^- would have $S = -1$. Unlike $\pi^{\pm}$ the $K^{\pm}$ form an isospin $\frac{1}{2}$ doublet.

7.7.2. *$K^0, \bar{K^0}$ is another pair of pseudo-scalar, isospin $\frac{1}{2}$ particles of mass* 498 *MeV that are also stable under strong interactions*

K^0 has $I_3 = -\frac{1}{2}, S = 1$ and $\bar{K^0}$has $I_3 = \frac{1}{2}, S = -1$ .The charges of all the K-mesons can be fit by changing the formula for electric charge (Gell–Mann–Nishijima)

$$Q = I_3 + \frac{B+S}{2}$$

7.7.3. *The new quantum number is accounted for by a new kind quark, the strange quark*

By a twist of fate, the strange quark has $S = -1$ and the strange anti-quark has $S = +1$. It has baryon number $\frac{1}{3}$ like the u and d quarks. From the above formula we see that its electric charge is $-\frac{1}{3}$. That is the same charge as the d quark. Thus we have the constituents of the Kaons:

$$K^+ = \bar{s}u,\ K^- = \bar{u}s,\ K^0 = \bar{s}d\ \bar{K^0} = \bar{d}s$$

7.7.4. *There is also a neutral pseudoscalar meson that has strangeness zero and isospin zero*

$$\eta^0 = \bar{s}s$$

with a mass ≈548 MeV. It decays mostly into 2γ which can be thought of as the strange quark and anti-quark annihilating each other. A more accurate description of the η^0 includes mixing with $\bar{u}u$ and $\bar{d}d$.

7.7.5. *The s quark is heavier than the u and d quarks*

This explains why particles that contain it are a few hundred MeV heavier than corresponding particles made from u and d quarks alone. For example, $m_{K^+} - m_{\pi^+} \approx 350$ MeV. Recall that the d quark is slightly heavier (by a few MeV) than the u quark to explain the neutron-proton mass difference. For strong interactions, the three quarks behave the same way. If we also ignore their mass differences, the isospin symmetry is enlarged to a symmetry that rotates three quarks into each other. Since these transformations can involve complex matrices, the symmetry must involve 3×3 complex matrices. One natural choice is to generalize the $SU(2)$ of isospin to $SU(3)$. This is not the only possibility: there several Lie groups with two commuting quantum numbers that could be I_3 and S. But $SU(3)$ is a reasonable guess and it worked.

7.7.6. *The* **8** *pseudo-scalar mesons* $\pi^\pm, \pi^0, K^\pm, K^0, \bar{K^0}, \eta^0$ *form a representation of* $SU(3)$

Analogous to the way the pions form a three-dimensional representation of $SU(2)$.

7.8. $SU(3)$

Recall the definition of the Lie algebra $su(n)$:

7.8.1. *The space of traceless hermitian $n \times n$ matrices is called $su(n)$*

We use lower case letters the Lie algebra and upper case letters for the group.

7.8.2. *The number of linearly independent elements of $su(n)$ is $n^2 - 1$*

A hermitian matrix has n^2 independent components: There are n real entries along the diagonal and $\frac{n(n-1)}{2}$ complex numbers above the diagonal. The entries below the diagonal are not independent because they are just complex conjugates of the ones above, so the total is $n + 2\frac{n(n-1)}{2} = n^2$. Since an anti-Hermitian matrix is simply i times a hermitian one, its number of independent components is also n^2. This is called the dimension of $u(n)$.

The condition of being traceless imposes one condition among the diagonal entries, so the number of independent components of $su(n)$ is $n^2 - 1$. In particular,

7.8.2.1. *The dimension of $su(3)$ is* 8

7.8.3. *The Gell–Mann matrices provide a basis for $su(3)$*

$$A = a_1\frac{\lambda_1}{2} + a_2\frac{\lambda_2}{2} + a_3\frac{\lambda_3}{2} + a_4\frac{\lambda_4}{2} + a_5\frac{\lambda_5}{2} + a_6\frac{\lambda_6}{2} + a_7\frac{\lambda_7}{2} + a_8\frac{\lambda_8}{2}$$

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}$$

$$\lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}$$

$$\lambda_8 = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}$$

These are normalized such that

$$\operatorname{tr} \lambda_\alpha \lambda_\beta = 2\delta_{\alpha\beta}$$

in imitation of the Pauli matrices.

7.8.3.1. *λ_3 and λ_8 are diagonal. They are related to isospin and strangeness of quarks*

Exercise 20. Derive the commutation relations of $su(3)$ in the Gell–Mann basis. That is, write the commutators $\left[\frac{\lambda_\alpha}{2}, \frac{\lambda_\beta}{2}\right] = i f_{\alpha\beta\gamma} \frac{\lambda_\gamma}{2}$ as linear combinations of the Gell–Mann matrices. Decompose the 8 dimensional space $su(3)$ into irreducible representations of $SU(2)$.

Partial Solution

Using the normalization of the trace above, we get

$$f_{\alpha\beta\gamma} = -i\mathrm{tr}\left(\left[\frac{\lambda_\alpha}{2}, \frac{\lambda_\beta}{2}\right]\lambda_\gamma\right)$$

Obviously this is anti-symmetric in α, β. It is also anti-symmetric in β and γ since:

$$\mathrm{tr}\left(\left[\lambda_\alpha, \lambda_\beta\right]\lambda_\gamma\right) + \mathrm{tr}\left(\left[\lambda_\alpha, \lambda_\gamma\right]\lambda_\beta\right) =$$

(using$\mathrm{tr}AB = \mathrm{tr}BA$)

$$\mathrm{tr}\left(\left[\lambda_\alpha, \lambda_\beta\right]\lambda_\gamma\right) + \mathrm{tr}\left(\lambda_\beta\left[\lambda_\alpha, \lambda_\gamma\right]\right)$$

(using $[A, BC] = [A, B]C + B[A, C]$)

$$= \mathrm{tr}\left(\left[\lambda_\alpha, \lambda_\beta\lambda_\gamma\right]\right) = 0$$

since $\mathrm{tr}[A, B] = 0$.

Thus $f_{\alpha\beta\gamma}$ is completely anti-symmetric. So it is enough to calculate it when $\alpha < \beta < \gamma$: All other components are either zero or are given by a signed permutation.

7.9. Gell–Mann–Okubo Formula

7.9.1. *The most obvious consequence of the strange quark is that there is a spin $\frac{3}{2}$ baryon sss*

This is analogous to $ddd = \Delta^{--}$ and so should be negatively charged. It is called the Ω^-. The spin is $\frac{3}{2}$ because the quark is a fermion and its wavefunction is anti-symmetric in color. Same argument as for uuu or ddd. Obviously, Ω^- should not carry isospin and has strangeness -3.

7.9.2. ***Next there should be an isospin $\frac{1}{2}$ pair ssu and ssd of charges zero and -1 respectively***

These are called Ξ^{*0}, Ξ^{*-}. They have strangeness $S = -2$. The star is to distinguish it from another similar particle of spin $\frac{1}{2}$; which coincidentally is called Ξ.

7.9.3. ***There is an isospin 1 triplet suu, sud, sdd of charges $1, 0, -1$.***

These are called $\Sigma^{*+}, \Sigma^{*0}, \Sigma^{*-}$.

7.9.4. ***Along with the original quartet $\begin{pmatrix} \Delta^{++} \\ \Delta^{+} \\ \Delta^{0} \\ \Delta^{-} \end{pmatrix} = \begin{pmatrix} uuu \\ uud \\ udd \\ ddd \end{pmatrix}$ we get a set of ten spin $\frac{3}{2}$ baryons.***

To a good approximation we can think of the u and d quarks as having the same mass, but the s quark is heavier so that $SU(3)$ is broken down to its isospin subgroup.If we ignore the breakdown of isospin (a smaller effect), Δ are all of the same mass. The three Σ^*, would have somewhat larger mass, then $\Xi^*$and Ω^-. larger still. The static quarks model would say that each time we add replace a u or d quark by an s quark we are increasing the mass of a particle by some fixed amount: The $s-u$ mass difference. Thus we get a particular case of the Gell–Mann–Okubo mass relations

$$m_{\Sigma^*} - m_{\Delta} = m_{\Xi^*} - m_{\Sigma} = m_{\Omega^-} - m_{\Xi^*}$$

At the time that this relation was discovered (by deeper group theoretic arguments rather than the static quark model) it was known that (all masses are in MeV)

$$m_{\Delta} = 1230,\ m_{\Sigma^*} = 1385,\ m_{\Xi^*} = 1530$$

but the Ω^- had not been seen yet. Thus, the first equality is a post-diction that could be verified

$$m_{\Sigma^*} - m_{\Delta} = 155 \approx 145 = m_{\Xi^*} - m_{\Sigma^*}$$

And there is a prediction of the mass of the Ω^-:

$$m_{\Omega^-} \approx 1675$$

The discovery of the Ω^- with a mass of 1672MeV at Brookhaven was spectacular confirmation of the Gell–Mann–Okubo relations.

7.9.5. *Together they form the 10 dimensional representation of SU(3)*

The quarks belong to the fundamental representation of dimension three of $SU(3)$.

There are three quarks in each baryon. The anti-symmetry in color implies that the wavefunction must be symmetric in spin and "flavor" $SU(3)$. The spin $\frac{3}{2}$ baryons are completely symmetric in spin, so are also completely symmetric in flavor. The completely symmetric third rank tensor representation of $SU(3)$ is of dimension 10.

7.9.6. *Another way of thinking of this is in terms of an even more approximate SU(6) symmetry*

The static quark model with three quarks will have $SU(6)$ symmetry: Not only do the u, d, s quarks all have the same mass, we treat there spin up or down states as having the same energy: Altogether there are six states for the quarks with the same energy. After removing color, the wavefunction of a baryon is in the completely symmetric third rank tensor representation of $SU(6)$. This is of dimension $\frac{6(6+1)(6+2)}{3!} = 56$.

When we reduce with respect to spin and flavor $SU(2) \times SU(3) \subset SU(6)$ we can first identify a spin $\frac{3}{2}$ representation which is completely symmetric in flavor (as seen above). The number of independent states of the third rank symmetric tensor in dimension three is $\frac{3(3+1)(3+2)}{3!} = 10$. So, altogether these account for $4 \times 10 = 40$ of the states.

7.9.7. *The spin $\frac{1}{2}$ Baryons form the eight-dimensional representation of SU(3)*

There remains 16 states of spin $\frac{1}{2}$. Spin alone accounts for a factor of $2\frac{1}{2} + 1 = 2$. Therefore these states must belong to an 8 dimensional representation of $SU(3)$. Of these 8 , we can identify $n = udd$ and $p = uud$ as the states without any strange quark. There are three particles of strangeness -1 containing one strange quark:

$$\Sigma^- = dds, \quad \Sigma^0 = uds, \quad \Sigma^- = uus$$

Here the ud quarks are combined so as to produce $I = 1, I_3 = 0$. There is another combination uds that has $I = 0, I_3 = 0$ called the Λ baryon. Finally, there are the doubly strange baryons

$$\Xi^- = ssd, \quad \Xi^0 = ssu$$

7.9.8. *The spin zero fields form also the eight dimensional (adjoint) representation of SU*(3)

It is a coincidence of $SU(3)$ that the mesons and the spin $\frac{1}{2}$ baryons transform the same way.

Chapter 8

BOSONS AND FERMIONS

Many physical systems are made of a large number of copies of identical particles. The electrons inside a metal or a white dwarf star, particles of light (photons) inside a resonant cavity or the cosmos, the atoms in liquid Helium are all examples. This idea is surprisingly versatile: Vibrations of a solid can be thought of as a collection of particles of sound (phonons); the electrostatic field of a charged particle can be resolved as a superposition of photons.

Let us begin by considering free particles, so that the total energy is the sum of the energies of each of them. The simplest situation is that there is just one state available with energy[1] $\hbar\omega$. If there are n free particles occupying this state, the total energy is $n\hbar\omega$. The empty state with $n = 0$ has the lowest possible energy (is the "ground state") and is sometimes called the "vacuum". The number n of particles in the state is called the "occupation number".

If there are many energy levels, $\hbar\omega_k$ for some range of values of k, each of them will have some occupation number n_k and the total energy will be

$$E = \hbar\omega_1 n_1 + \hbar\omega_2 n_2 + \cdots = \sum_k \hbar\omega_k n_k.$$

The single particle energies $\hbar\omega_k$ are often determined by solving an eigenvalue problem for a differential operator with appropriate boundary conditions. For now we don't need to know these details.

At first, it was thought that n can be any number $n = 0, 1, 2, 3, \ldots$. This is true for photons, the quantum many body system originally studied by

[1] Recall that $\hbar$ has units of Joule*secs. So ω has units of frequency. The physical meaning of this frequency will become clear later.

Bose. Many other particles (phonons, He_4 atoms, pi mesons, W, Z particles, the Higgs particle etc.) are also of this type. Such particles are now called bosons. It turns out that any particle whose angular momentum is an integer multiple of $\hbar$ is a boson. This relation with angular momentum is not at all obvious and is one of the deepest theorems ("spin-statistics theorem") of Field Theory. But we do not go into that here.

Electrons are not bosons. The Exclusion Principle, which Pauli deduced by looking at atomic energy levels, says that each atomic energy level can be occupied by at most one electron. That is, either $n = 0$ or $n = 1$. Now, we know that this is related to the fact that electrons have half-integer angular momentum: The minimum is $\frac{\hbar}{2}$. In fact all particles with half integer angular momentum satisfy the exclusion principle: The other half of the Spin-Statistics theorem. Such particles are called fermions. Electrons, protons, neutrons, muons, quarks, neutrinos etc. are examples.

To summarize, there are two kinds of particles. Those with integer angular momentum are bosons and those with half -integer angular momentum are fermions. For Fermions, the occupation numbers can be $n_k = 0, 1$ and for bosons, $n_k = 0, 1, 2, \ldots$.

If the particles interact with each other (are not free) the energies do not just add. Determining the energies of such interacting many body systems are among the hardest problems in physics. Many of the most interesting phenomena (scattering, superconductivity, symmetry breaking) arise this way. We will return to some of this later.

8.1. Partition Function

Counting is hard. It is harder than calculus or linear algebra. So we can often use ideas from these disciplines to help with counting. This is the mathematical field of combinatorics.

Suppose we want to know the degeneracy $d(E)$ of a state of a free bosonic system with energy E. That is, how many ways can we choose n_k such that $E = \sum_k \hbar\omega_k n_k$. (Suppose that ω_k are already known.) A function helps with this counting, called the "Partition Function".

$$Z(\beta) = \sum_E d(E) e^{-\beta E}$$

We can think of $\frac{e^{-\beta E}}{Z(\beta)}$ as the probability that energy is equal to E. The quantity β actually has a physical meaning: $\beta = \frac{1}{k_B T}$ where T is the

absolute temperature and k_B is Bolzmann's constant. When $T > 0$, the states of high energy are less likely. (There are some situations where the temperature is negative and the opposite is true.) Usually $d(E)$ grows with E like a power of E and the most important contributions to $Z(\beta)$ come from the value of E where the summand has a maximum: The two factors pull in opposite directions.

Exercise 21. Show that the mean value of energy is $\langle E \rangle = -\frac{\partial}{\partial \beta} \log Z(\beta)$. What is its variance?

Once we know this function, we can can extract $d(E)$ by various tricks (e.g., Fourier analysis). It is a very efficient way of packing the information in $d(E)$. In the simplest case with just one frequency ω

$$d(E) = 1, \quad E = n\hbar\omega$$

$$Z_B(\beta) = \sum_{n=0}^{\infty} e^{-\beta n \hbar \omega} = \frac{1}{1 - e^{-\hbar\omega\beta}}$$

Just the geometric series. For fermions,

$$Z_F(\beta) = 1 + e^{-\hbar\omega\beta}.$$

When there are many frequencies, we can still write

$$Z_B(\beta) = \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \cdots e^{-\beta[n_1\hbar\omega_1 + n_2\hbar\omega_2 + \cdots]}$$

This summand factorizes:

$$Z_B(\beta) = \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \cdots e^{-\beta n_1 \hbar\omega_1} e^{-\beta n_2 \hbar\omega_2} \cdots$$

Each factor depends only one of the occupation numbers

$$Z_B(\beta) = \sum_{n_1=0}^{\infty} e^{-\beta n_1 \hbar\omega_1} \sum_{n_2=0}^{\infty} e^{-\beta n_2 \hbar\omega_2} \cdots$$

and so each one can be evaluated a geometric series

$$Z_B(\beta) = \frac{1}{1 - e^{-\hbar\omega_1\beta}} \frac{1}{1 - e^{-\hbar\omega_2\beta}} \cdots$$

In more compact notation

$$Z_B(\beta) = \prod_k \frac{1}{1 - e^{-\beta\hbar\omega_k}}$$

Similarly for fermions

$$Z_F(\beta) = \prod_k \left[1 + e^{-\beta\hbar\omega_k}\right]$$

Exercise 22. A more subtle partition function also keeps track of the total number of particles $N = \sum_k n_k$, not just the total energy $Z(\beta, \mu) = \sum_{E,N} d(E, N)e^{-\beta E - \mu N}$. Show that

$$Z_B(\beta, \mu) = \prod_k \frac{1}{1 - e^{-\beta\hbar\omega_k - \mu}}, \quad Z_F(\beta, \mu) = \prod_k \left[1 + e^{-\beta\hbar\omega_k - \mu}\right]$$

Any sensible person seeing such a product will convert into a sum by taking logarithms.

$$\log Z_B(\beta) = -\sum_k \log\left[1 - e^{-\beta\hbar\omega_k}\right]$$

$$\log Z_F(\beta) = \sum_k \log\left[1 + e^{-\beta\hbar\omega_k}\right]$$

Often the sum over k can be converted to an integral, when there are many frequencies closely packed together. We will see examples later.

Exercise 23. This has little physical meaning, but makes a connection to the most famous function in Number Theory. Suppose $\omega_k = \log p_k$ where $p_k = 2, 3, 5, 7 \cdots$ are the prime numbers. What is the Bosonic partition function?

Solution

$$\begin{aligned} Z(\beta) &= \sum_{n_k=0}^{\infty} e^{-\beta \sum_k n_k \log p_k} \\ &= \Pi_{k=0}^{\infty} \left(\sum_{n_k=0}^{\infty} e^{-\beta n_k \log p_k} \right) \\ &= \Pi_{k=0}^{\infty} \frac{1}{1 - e^{-\beta \log p_k}} \\ Z(\beta) &= \Pi_{k=0}^{\infty} \frac{1}{1 - p_k^{-\beta}} \end{aligned}$$

On the other hand, any number can be written as a product of prime powers in a unique way:

$$N = \prod_k p_k^{n_k} \iff \log N = \sum_k n_k \log p_k$$

So

$$Z(\beta) = \sum_{n_k=0}^{\infty} e^{-\beta \sum_k n_k \log p_k} = \sum_{N=1}^{\infty} e^{-\beta \log N} = \sum_{N=1}^{\infty} \frac{1}{N^\beta}$$

This is the Riemann zeta function;the variable β is usually called s in number theory.

We just derived a famous identity of Euler for the Riemann zeta function

$$\Pi_{k=0}^{\infty} \frac{1}{1 - p_k^{-\beta}} = \zeta(\beta) \equiv \sum_{N=1}^{\infty} \frac{1}{N^\beta}$$

8.1.1. *The Planck spectrum*

Quantum mechanics began with Planck's resolution of a paradox in combining statistical physics with the new theory of electromagnetic radiation. If you imagine a large metal box of side L, the electric and magnetic fields inside can have wave numbers $\frac{2\pi}{L}\boldsymbol{n}$ where $\boldsymbol{n} \in \mathbb{Z}^3$ is a triple of integers. (Solve the wave equation with the boundary conditions at the sides of the box).

8.2. The Harmonic Oscillator

This review is too brief if you have not seen this previously. See the book [7] for an excellent, detailed presentation.

8.2.1. *The classical simple harmonic oscillator*

The simple harmonic oscillator is a system with one degree of freedom with hamiltonian

$$H = \frac{1}{2}p^2 + \frac{1}{2}\omega^2 q^2$$

ω is the natural frequency. In classical mechanics, the momentum p and position q satisfy the Poisson Brackets

$$\{p, q\} = 1$$

The equations of motion are

$$\frac{dq}{dt} = \{H, q\} = p, \quad \frac{dp}{dt} = \{H, p\} = -\omega^2 q$$

It is useful to introduce complex linear combinations

$$\alpha = \frac{1}{\sqrt{2}}\left(\sqrt{\omega}q + \frac{1}{\sqrt{\omega}}ip\right), \quad \alpha^* = \frac{1}{\sqrt{2}}\left(\sqrt{\omega}q - \frac{1}{\sqrt{\omega}}ip\right)$$

so that

$$H = \omega\alpha^*\alpha, \quad \{\alpha, \alpha^*\} = i$$

satisfying

$$\frac{d\alpha^*}{dt} = \{H, \alpha^*\} = i\omega\alpha^*, \quad \frac{d\alpha}{dt} = \{H, \alpha\} = -i\omega\alpha$$

The solutions are periodic (oscillating) functions

$$\alpha = Ae^{-i\omega t}, \quad \alpha^* = A^* e^{i\omega t}$$

or, after some change of variables,

$$q(t) = Q\sin(\omega t + \theta)$$

for some constants A, Q, θ. Thus, the period of the oscillation is $\frac{2\pi}{\omega}$.

8.2.1.1. *Partition function*

In classical statistical mechanics, the probability density of states in phase space is proportional to $e^{-\beta H}$ where $\beta = \frac{1}{k_B T}$ is the inverse temperature. (Boltzmann's constant k_B converts T from units of temperature to units of energy.) The thermodynamic properties of a system are given by the partition function

$$Z(\beta) = Z_0 \int e^{-\beta H} dpdq$$

The overall constant Z_0 is undetermined in the classical theory.

For the harmonic oscillator this is a gaussian integral

$$Z(\beta) = Z_0 \int e^{-\beta\left[\frac{1}{2}p^2 + \frac{1}{2}\omega^2 q^2\right]} dpdq = Z_0 \frac{2\pi}{\omega\beta}$$

We will see that $Z_0 = \frac{1}{2\pi\hbar}$ from comparison with the quantum theory.

8.2.2. *The quantum simple harmonic oscillator*

In quantum mechanics p and q are operators satisfying the Heisenberg commutation relations

$$[p, q] = -i\hbar$$

In the Schrodinger picture, $p = -i\hbar\frac{\partial}{\partial q}$. The hamiltonian is now a differential operator

$$H = \frac{1}{2}\left[-\hbar^2\frac{\partial^2}{\partial q^2} + \omega^2 q^2\right]$$

Its eigenvalues E_n are the allowed values of energy. They are the values for which the ODE

$$H\psi_n = E_n\psi_n$$

has non-zero solutions of finite norm:

$$\int |\psi_n(q)|^2 dq < \infty.$$

To find them, it is useful to introduce two related operators

$$a = \frac{1}{\sqrt{2}}\left[\sqrt{\frac{\omega}{\hbar}}q + \sqrt{\frac{\hbar}{\omega}}\frac{\partial}{\partial q}\right], \quad a^\dagger = \frac{1}{\sqrt{2}}\left[\sqrt{\frac{\omega}{\hbar}}q - \sqrt{\frac{\hbar}{\omega}}\frac{\partial}{\partial q}\right]$$

so that

$$[a, a^\dagger] = 1$$

and

$$H = \frac{1}{2}\hbar\omega\left(a^\dagger a + aa^\dagger\right) = \hbar\omega a^\dagger a + \frac{1}{2}\hbar\omega$$

Exercise. Verify the above relations. Pay special attention to the term $\frac{1}{2}\hbar\omega$ in the hamiltonian.

So, how does this help to find eigenvalues of H? The key idea are the relations

$$[H, a^\dagger] = \hbar\omega a^\dagger, \quad [H, a] = -\hbar\omega a$$

Thus, if we have one solution to the eigenvalue equation $H\psi_n = E_n\psi_n$ we can find another:

$$\begin{aligned} Ha^\dagger\psi_n &= [H, a^\dagger]\psi_n + a^\dagger H\psi_n \\ &= \hbar\omega a^\dagger\psi_n + E_n a^\dagger\psi_n \end{aligned}$$

That is, the application of $a^\dagger$ to ψ_n gives us another eigenstate with energy $E_n + \hbar\omega$:

$$Ha^\dagger\psi_n = (E_n + \hbar\omega)\, a^\dagger\psi_n$$

as long as $a^\dagger\psi_n \neq 0$ and has finite norm. This is why $a^\dagger$ is called a "raising operator": It raises the energy. Similarly, a is a lowering operator.

$$Ha\psi_n = (E_n - \hbar\omega)\, a\psi_n$$

That is, we can lower energy by application of a unless $a\psi_n = 0$. The lowest value of energy corresponds to a solution of the equation

$$a\psi_0 = 0$$

We can see that

$$H\psi_0 = \frac{1}{2}\hbar\omega\psi_0$$

so that the "ground state" has energy $\frac{1}{2}\hbar\omega$.

It is easy enough to solve this ODE

$$\left[\sqrt{\frac{\omega}{\hbar}}q + \sqrt{\frac{\hbar}{\omega}}\frac{\partial}{\partial q}\right]\psi_0(q) = 0 \implies$$

$$\psi_0 = Ce^{-\frac{1}{2}\frac{\omega}{\hbar}q^2}$$

for some constant C.This function is normalizable; i.e., $\int |\psi_0|^2 dq$ is finite. We could choose C so that $\int |\psi_0|^2 dq = 1$. But this is not important.

Example. What would have happened if we had solved $a^\dagger\chi_0 = 0$ instead? Why is there no "highest value" for energy?

Now, we can find the next highest value of energy by applying the raising operator:

$$\begin{aligned} \psi_1 &= a^\dagger\psi_0 \\ H\psi_1 &= \hbar\omega\left[1 + \frac{1}{2}\right]\psi_1 \end{aligned}$$

and so on:

$$\psi_n = a^{\dagger n}\psi_0$$
$$H\psi_n = \left[n + \frac{1}{2}\right]\hbar\omega\psi_n, \quad n = 0, 1, 2\cdots$$

In many situations, a constant added to all the energy levels does not matter: Only differences of energy levels matter. In that case we can redefine the hamiltonian so that the ground state has energy zero:

$$\tilde{H} = H - \frac{1}{2}\hbar\omega$$
$$\psi_n = a^{\dagger n}\psi_0$$
$$\tilde{H}\psi_n = n\hbar\omega\psi_n, \quad n = 0, 1, 2\cdots$$

Thus the energy levels of a harmonic oscillator and equally spaced, multiples of $\hbar\omega$.

8.2.2.1. *Partition function*

In quantum statistical physics, the probability of a state is proportional to its energy. The partition function is then $Z(\beta) = \sum_E d(E)e^{-\beta E}$ where $d(E)$ is the degeneracy of energy E (i.e., the number of states with energy E). For the SHO, this is a sum we already evaluated: $Z(\beta) = \sum_{n=0}^{\infty} e^{-n\hbar\omega\beta} = \frac{1}{1-e^{-\hbar\omega\beta}}$. We can see that in the limit $\hbar \to 0$

$$Z(\beta) \to \frac{1}{\hbar\omega\beta}$$

which agrees with the classical result if the normalization constant $Z_0 = \frac{1}{2\pi\hbar}$.

8.2.3. *More degrees of freedom*

Any oscillating system with small displacements can be reduced to a sum of independent simple harmonic oscillators by a change of variables [13].

$$H = \frac{1}{2}\sum_k p_k^2 + \frac{1}{2}\sum_k \omega_k^2 q_k^2$$
$$\{p_k, q_l\} = \begin{cases} 1 & k = l \\ 0 & k \neq l \end{cases}$$
$$\{p_k, p_l\} = 0 = \{q_k, q_l\}$$

Complex combinations again are useful

$$\alpha_k = \frac{1}{\sqrt{2}}\left(\sqrt{\omega_k}q_k + \frac{1}{\sqrt{\omega_k}}ip_k\right), \quad \alpha_k^* = \frac{1}{\sqrt{2}}\left(\sqrt{\omega_k}q_k - \frac{1}{\sqrt{\omega_k}}ip_k\right)$$

$$\{\alpha_k, \alpha_l^*\} = \begin{cases} i & k = l \\ 0 & k \neq l \end{cases}$$

$$\{\alpha_k, \alpha_l\} = 0 = \{\alpha_k^*, \alpha_l^*\}$$

so that

$$H = \sum_k \omega_k \alpha_k^* \alpha_k.$$

The quantum theory also decomposes into a sum of independent systems:

$$H = \sum_k \hbar\omega a_k^\dagger a_k + \frac{1}{2}\sum_k \hbar\omega_k$$

$$\left[a_k, a_l^\dagger\right] = \begin{cases} 1 & k = l \\ 0 & k \neq l \end{cases}$$

$$[a_k, a_l] = 0 = \left[a_k^\dagger, a_l^\dagger\right]$$

Again, if ψ_n is an eigenstate of energy E_n , $a_k^\dagger\psi_n$ is one of energy $E_n + \hbar\omega_k$. As long as $a_k\psi_n \neq 0$, it is an eigenstate of energy $E_n - \hbar\omega_k$.The ground state is the simultaneous solution of

$$a_k\psi_0 = 0$$

which is a set of PDEs

$$\left[\sqrt{\frac{\omega_k}{\hbar}}q_k + \sqrt{\frac{\hbar}{\omega_k}}\frac{\partial}{\partial q_k}\right]\psi_0(q) = 0$$

The solution is separable as a product of functions of single variables:

$$\psi_0 = C\prod_k e^{-\frac{1}{2}\frac{\omega_k}{\hbar}q_k^2}$$

We can verify

$$H\psi_0 = \left[\sum_k \frac{1}{2}\hbar\omega_k\right]\psi_0$$

This is the state of lowest energy. The other states obtained by acting on ψ_0 with some raising operators

$$\psi_n = a_1^{\dagger n_1} a_2^{\dagger n_2} \cdots \psi_0$$
$$H\psi_n = \left[\sum_k \left(n_k + \frac{1}{2}\right)\hbar\omega_k\right]\psi_n$$

Again, we can add a constant to H so that the ground state energy is zero:

$$\tilde{H} = H - \sum_k \frac{1}{2}\hbar\omega_k$$
$$\tilde{H} = \sum_k \hbar\omega_k a_k^\dagger a_k$$
$$\tilde{H}\psi_n = \left[\sum_k \hbar\omega_k n_k\right]\psi_n$$

8.3. Free Bosons are Harmonic Oscillators

You must have caught on by now to the point I am trying to make. The energies of a system of harmonic oscillators are exactly the same as those of a system of bosons.

$$E_n = \sum_k n_k \hbar\omega_k, \quad n_k = 0, 1, 2 \cdots$$

So, the mathematical description of bosons are as oscillators. The natural frequencies of the oscillators are (up to a factor of $\hbar$) the energies of single bosons. The operator $a_k^\dagger$ can now be thought as creating a boson in the state of energy $\hbar\omega_k$. Similarly, a_k is the annihilation operator of such a boson.

An immediate application is to vibrations in a solid. At equilibrium, the atoms of a crystalline solid are arranged in a regular periodic pattern: A lattice. Small oscillations of the atoms around these equilibrium positions propagate as sound waves. They have a characteristic frequencies which are determined by the spring constants (second derivative at equilibrium of the potential between nearby atoms). Quantum mechanically, these sound waves behave like bosonic particles. They are called phonons.

Similarly, the electromagnetic field inside a cavity with conducting walls has oscillations. Quantum mechanically, these oscillations correspond to particles of light: Photons.

We are onto a fundamental physical fact: Particles are quantum excitations of fields. Particles and fields arose as different ideas in classical physics: Electrons where thought of as particles and electromagnetism as a field. In quantum theory, these ideas come together.

8.3.1. *Occupation number basis*

To describe particles, it is better to think of oscillator states directly in terms of occupation numbers rather than in terms of wave functions.

Again start with a single degree of freedom. We want operators satisfying

$$[a, a^\dagger] = 1$$

Denote the empty state by the symbol $|0\rangle$ (pronounced "ket 0")

$$a|0\rangle = 0$$

More generally, introduce states satisfying

$$a^\dagger a|n\rangle = n|n\rangle$$

for some real numbers n. (Since $a^\dagger a$ is hermitian, its eigenvalues are real.)

For different values of n these states must be orthogonal.

$$\langle n|m\rangle = 0, \quad n \neq m$$

$\langle n|$ is the conjugate vector to $|n\rangle$ and is pronounced "bra n". So, $\langle n|m\rangle$ is the bra-ket[2] of n and m. We can normalize the eigenstates to have length one.

$$\langle n|n\rangle = 1$$

These are simply the eigenstates of the hamiltonian. Now,

$$a^\dagger a\left(a^\dagger|n\rangle\right) = a^\dagger\left(1 + a^\dagger a\right)|n\rangle = (n+1)a^\dagger|n\rangle$$

[2]This terrible pun is due to Dirac. His only excuse is that he was quite young when he invented this.

Thus, $a^\dagger|n\rangle$ is proportional to $|n+1\rangle$. What is the proportionality factor?Let

$$a^\dagger|n\rangle = \xi_n|n+1\rangle$$

As an infinite dimensional matrix,

$$a^\dagger = \begin{pmatrix} 0 & \xi_0 & 0 & 0 & \cdots \\ 0 & 0 & \xi_1 & 0 & \cdots \\ 0 & 0 & 0 & \xi_2 & \cdots \\ 0 & 0 & 0 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

We can derive recursion relation relation for ξ_n using the commutation relations and the requirement that $a, a^\dagger$ are conjugate to each other.

$$a = \begin{pmatrix} 0 & 0 & 0 & \cdots \\ \xi_0^* & 0 & 0 & \cdots \\ 0 & \xi_1^* & 0 & \cdots \\ 0 & 0 & \xi_2^* & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

This can also be written as

$$a|0\rangle = 0, \quad a|n\rangle = \xi_{n-1}^*|n-1\rangle, \quad n > 0$$

Then

$$a^\dagger a = \begin{pmatrix} |\xi_0|^2 & 0 & 0 & \cdots \\ 0 & |\xi_1|^2 & 0 & \cdots \\ 0 & 0 & |\xi_2|^2 & \cdots \\ 0 & 0 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix}, \quad aa^\dagger = \begin{pmatrix} 0 & 0 & 0 & \cdots \\ 0 & |\xi_0|^2 & 0 & \cdots \\ 0 & 0 & |\xi_1|^2 & \cdots \\ 0 & 0 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

Thus

$$[a, a^\dagger] = 1 \implies |\xi_0|^2 = 1, \quad |\xi_n|^2 - |\xi_{n-1}|^2 = 1, \quad n > 0$$

The solution to this recursion relation is $|\xi_n|^2 = n+1$. So, we can choose $\xi_n = \sqrt{n+1}$.

$$a^\dagger|n\rangle = \sqrt{n+1}|n+1\rangle, \quad a|n\rangle = \begin{cases} \sqrt{n-1}|n-1\rangle & n > 0 \\ 0 & n = 0 \end{cases}$$

Or, as infinite dimensional matrices

$$a = \begin{pmatrix} 0 & 0 & 0 & \cdots \\ 1 & 0 & 0 & \cdots \\ 0 & \sqrt{2} & 0 & \cdots \\ 0 & 0 & \sqrt{3} & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix}, \quad a^\dagger = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & \sqrt{2} & 0 & \cdots \\ 0 & 0 & 0 & \sqrt{3} & \cdots \\ 0 & 0 & 0 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

Exercise 24. Find normalizable states satisfying $a^\dagger \chi(z) = z a^\dagger \chi(z)$, $\chi(z) = \sum_{n=0} \chi_n(z)|n\rangle$ These "coherent states" are especially useful in quantum optics.

Exercise 25. Recall that $|0\rangle$ is $Ce^{-\frac{1}{2}\frac{\omega}{\hbar}q^2}$. Find polynomials $P_n(q)$ such that $|n\rangle$ is $P_n(q)Ce^{-\frac{1}{2}\frac{\omega}{\hbar}q^2}$. (**Hint** Derive recursion relations for $P_n(q)$).

8.4. Are Free Fermions some kind of Oscillators too?

Is there some kind of oscillator whose excitation number takes just two values $n = 0, 1$?

Let us start with a system with just one natural frequency ω. It is either empty or is filled by a fermion. If empty, the energy is zero. If filled the energy is $\hbar\omega$.

$$H|0\rangle = 0, \quad H|1\rangle = \hbar\omega|1\rangle$$

There are no other states. By analogy to the creation annihilation operators for bosons we can introduce operators

$$b|1\rangle = |0\rangle, \quad b^\dagger|0\rangle = |1\rangle$$
$$b|0\rangle = 0, \quad b^\dagger|1\rangle = 0$$

The above relations imply

$$\left[bb^\dagger + b^\dagger b\right]|0\rangle = |0\rangle$$
$$\left[bb^\dagger + b^\dagger b\right]|1\rangle = |1\rangle$$

so that

$$bb^\dagger + b^\dagger b = 1.$$

Notice that the sign is + : We have an anti-commutator rather than a commutator as for bosons. We get

$$\tilde{H}_F = \hbar\omega b^\dagger b.$$

Since we cannot put more than one particle in the state (the exclusion principle) we must have

$$b^{\dagger 2} = 0$$

Since there is never more than one particle to annihilate,

$$b^2 = 0$$

as well. It is not difficult to find matrices and vectors that represent these relations:

$$|0\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad |1\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$b = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad b^\dagger = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

You might remember the Pauli spin matrices:

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

and then recognize

$$b^\dagger = \frac{\sigma_1 + i\sigma_2}{2}, \quad b = \frac{\sigma_1 - i\sigma_2}{2}.$$

The hamiltonian of the fermionic oscillator can be thought of as

$$\hat{H}_F = \frac{1}{2}\omega \left(b^\dagger b - b b^\dagger \right) = \omega b^\dagger b - \frac{1}{2}\omega$$

The extra $-\frac{1}{2}\omega$ is the analogue of the zero point energy. We can remove it by adding a constant to the hamiltonian as before (to get $\tilde{H}_F$) but it is more symmetric to leave it in. Then you can see that

$$\hat{H}_F = \frac{1}{2}\omega\sigma_3$$

in terms of the Pauli matrices.This connection between spin matrices and fermionic creation operators is not a coincidence; but we won't delve deeper into this matter here.

Exercise 26. Find the partition function of the fermionic oscillator above.

Solution $2\cosh[\frac{\omega\tau}{2}]$.

Is there an analogue of the representation of $a, a^\dagger$ as differential operators? We can invent one, but it needs a new kind of number.

8.4.1. *Grassmann numbers*

At some point you had to extend your idea of a number to include i , which solves $i^2 = -1$. In the same spirit, suppose there is a number that satisfies

$$\theta^2 = 0$$

but $\theta \neq 0$. A function of such a number can be thought of as a linear combination

$$\psi(\theta) = \psi_0 + \theta\psi_1$$

for real numbers ψ_0, ψ_1. Thus

$$\theta\psi = \theta\psi_0$$

We can define

$$\partial_\theta \psi = \psi_1$$

Clearly

$$\partial_\theta^2 = 0$$

Moreover

$$\theta\partial_\theta\psi = \theta\psi_1$$

$$\partial_\theta\left(\theta\psi\right) = \psi_0$$

so that

$$\left(\theta\partial_\theta + \partial_\theta\theta\right)\psi = \psi$$

Thus, these differential operators satisfy exactly the relations of $b, b^\dagger$.

We can now extend this trick to fermions with many degrees of freedom. Introduce variables that satisfy:

$$\theta^k\theta^l + \theta^l\theta^k = 0$$

In particular $(\theta^k)^2 = 0$ etc. A function of such variables will be a series

$$\psi(\theta) = \psi_0 + \psi_k\theta^k + \frac{1}{2!}\psi_{kl}\theta^k\theta^l + \cdots$$

the coefficients are anti-symmetric

$$\psi_{kl} = -\psi_{lk}$$

etc. Define the differentiation

$$\partial_k\psi(\theta) = \psi_k + \psi_{kl}\theta^l + \cdots$$

Then we can verify that

$$\partial_k \partial_l + \partial_l \partial_k = 0$$
$$\partial_k \theta^l + \theta^l \partial_k = \delta_k^l$$

Exercise. How would you define the integral $\int d\theta \psi(\theta)$ in order that it is linear and "translation invariant": $\int d\theta \psi(\theta + \alpha) = \int d\theta \psi(\theta)$ for some constant Grassmann number α?

Solution $\int d\theta \, [\psi_0 + \theta \psi_1] = \psi_1$is the only quantity that is linear and satisfies translation invariance. Incidentally, this means integration and differentiation are the same thing for functions of a Grassmann variable!

8.5. Beyond Free Particles: The Jaynes–Cummings Model*

To get a first taste of an interacting system, let us consider a boson with just one energy level $\hbar\omega_B$ and a fermion with a single level as well $\hbar\omega_F$. If they do not interact with each other the hamiltonian is

$$H_0 = \hbar\omega_B a^\dagger a + \hbar\omega_F b^\dagger b$$

The operators satisfy, as before,

$$[a, a^\dagger] = 1, \quad b^\dagger b + b b^\dagger = 1, \quad b^2 = 0 = b^{\dagger 2}$$

In addition, the fermion and boson creation annihilation operators do not affect each other's action; i.e., they commute:

$$[a, b] = 0 = [a^\dagger, b] = [a, b^\dagger] = [a^\dagger, b].$$

The energy levels are now

$$\hbar\omega_B n_B + \hbar\omega_F n_F$$

with $n_B = 0, 1, 2, \ldots$ and $n_F = 0, 1$. There can be any number of bosons, but at most one fermion.

This model was originally invented to represent an atom inside an electromagnetic resonant cavity of frequency ω_B.

We ignore all except one pair of energy levels of the atom. This pair has energies differing from each other by $\hbar\omega_F$ which is approximately equal to $\hbar\omega_B$. Since there are just two possible states for the atom, mathematically it is identical to the states of a single fermion: The empty fermion state is the ground state of the atom and the occupied fermion state is the excited state of the atom. This is the interpretation we will give it here.

8.5.1. *Interactions*

The atom can emit a photon and make a transition from its excited state to the ground state. In the other interpretation, a fermion is annihilated and a photon created. To have detailed balance (time reversal symmetry) we must allow for the opposite process as well: The atom absorbs a photon and climbs to the excited state, or the photon is annihilated and a fermion created.

This can represented by an additional "interaction term" in the hamiltonian

$$H_1 = \hbar g \left(a^\dagger b + a b^\dagger\right)$$
$$H = H_0 + H_1$$

Here g is a constant that measures the strength of the interaction ("coupling constant"). This model is simple enough that it can be exactly solved. The true eigenstates are some combinations of fermionic and bosonic states. Much can be learned by working out such simple examples in detail.

Exercise 27. Find a conserved quantity for the Jaynes–Cummings model. Use it to find the eigenvalues and eigenvectors of the hamiltonian.

Solution Let

$$a^\dagger a|n,\nu\rangle = n|n,\nu\rangle, \quad b^\dagger b|n,\nu\rangle = \nu|n,\nu\rangle, \quad n = 0,1,2,\ldots, \quad \nu = 0,1$$

Each term in the hamiltonian leaves $n+\nu$ unchanged so it is a conserved quantity. For a given value m of this quantity, there are two independent states in the above basis if $m > 0$

$$|m,0\rangle, \quad |m-1,1\rangle,$$

and just one, $|0,0\rangle$ if $m = 0$.

Conserved quantities are the key to solving any quantum system. If $m = 0$ we get an eigenstate with eigenvalue zero:

$$H|0,0\rangle = 0$$

Now, if $m > 0$

$$H|m,0\rangle = \hbar\omega_B m|m,0\rangle + \hbar g\sqrt{m}|m-1,1\rangle$$
$$H|m-1,1\rangle = [\hbar\omega_B(m-1) + \hbar\omega_F]|m-1,1\rangle + \hbar g\sqrt{m}|m,0\rangle$$

Thus in this two dimennsional subpsace H reduces to the matrix

$$\begin{pmatrix} \hbar\omega_B(m-1)+\hbar\omega_F & \hbar g\sqrt{m} \\ \hbar g\sqrt{m} & \hbar\omega_B m \end{pmatrix}$$

It is convenient to define

$$\Delta = \frac{\omega_F - \omega_B}{2}$$

and write this as

$$[\hbar\omega_B m + \Delta]\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} \hbar\Delta & \hbar g\sqrt{m} \\ \hbar g\sqrt{m} & -\hbar\Delta \end{pmatrix}$$

The eigenvalues are

$$\hbar\left\{\omega_B m + \Delta \pm \sqrt{\Delta^2 + g^2 m}\right\}$$

corresponding to eigenvectors

$$\begin{pmatrix} \cos\theta_\pm \\ \sin\theta_\pm \end{pmatrix}, \quad \tan\theta_\pm = \frac{\pm\sqrt{\Delta^2+g^2m}-\Delta}{g\sqrt{m}}$$

8.6. Heisenberg Lie Algebra

8.6.1. *The Heisenberg algebra is a three-dimensional Lie algebra with basis p, q, c satisfying*

$$[p,q] = c = -[q,p], \quad [q,c] = 0 = [p,c]$$

Since all double commutators vanish, Jacobi identity is easily verified. These are also called Canonical Commutation Relations.

8.6.1.1. *This algebra has no faithful finite dimensional irreducible representations*

Faithful means the basis elements are represented by linearly independent matrices; no information is lost in the representation. In an irreducible reresentation, c would have to be a multiple of the identity (Schur's lemma) since it commutes with everything. But we would have a contradiction:

$$\mathrm{tr}[p,q] = 0 \neq \mathrm{tr}c$$

In infinite dimensions, this argument doesn't work because each side can have infinite trace (i.e., the trace doesn't exist).

8.6.1.2. *The Heisenberg algebra has a unitary irreducible representation*

Unitary means that p and q are represented by hermitian operators and c by an anti-Hermitian operator. Being irreducible, it must be a multiple of the identity. To be unitary and irreducible, $c = i\hbar\hat{1}$, where $\hbar$ is some real number. It has the dimensions of angular momentum (momentum times position). In quantum mechanics it is Planck's constant.

The Schrodinger representation

$$p = -i\hbar\frac{\partial}{\partial q}$$

and q is represented by position. With respect to the inner product

$$\langle\psi, \chi\rangle = \int \psi^*(q)\chi(q)dq$$

p, q are hermitian (ignoring some technical issues about domains etc.) There is another approach more suited for our purposes:

8.6.2. *An equivalent formulation is in terms of the creation-annihilation operators*

$$[a, a^\dagger] = 1, \quad a = \frac{1}{\sqrt{2}}[q + ip]$$

8.6.3. *In the Schrodinger representation the states are described as functions of position*

The annihilation operator $a = \frac{1}{\sqrt{2}}\left[q + \frac{d}{dq}\right]$ in units with $\hbar = \omega = 1$ (which we use from now on). The equation $a|0\rangle = 0$ becomes a differential equation

$$\left[q + \frac{d}{dq}\right]\psi_0(q) = 0$$

which has solution

$$\psi_0(q) = Ce^{-\frac{1}{2}q^2}$$

The constant is chosen such that

$$\int |\psi_0(q)|^2 dq = 1 \implies C = \frac{1}{\sqrt{\pi}}$$

Then $|1>= a^\dagger|0>$ corresponds to the function

$$\begin{aligned}\psi_1(q) &= \frac{1}{\sqrt{2}}\left[q - \frac{d}{dq}\right]\psi_0(q) \\ &= \sqrt{2}q\psi_0(q)\end{aligned}$$

The rest of the states are given by the recursion relation

$$\psi_n(q) = \frac{1}{\sqrt{n}} a^\dagger \psi_{n-1}(q)$$

leading to

$$\psi_n(q) = H_n(q)\psi_0(q)$$

where $H_n(q)$ is a polynomial (Hermite) of order n.

Exercise. Derive a recursion relation giving H_n in terms of H_{n-1} and its derivative.

Solution

We have $|n\rangle = \frac{1}{\sqrt{n}} a^\dagger |n-1\rangle$ which translates to

$$H_n(q)\psi_0(q) = \frac{1}{\sqrt{2n}}\left(q - \frac{d}{dq}\right)\left[H_{n-1}(q)e^{-\frac{1}{2}q^2}\right]$$

This simplifies to

$$H_n(q) = \frac{1}{\sqrt{2n}}\left[2qH_{n-1}(q) - H'_{n-1}(q)\right]$$

8.6.4. *Another equivalent point of view is in terms of complex functions*

See the book by Klauder and Sudarshan [18] for more.

The creation operator is just multiplication by z; destruction is differentiation

$$a^\dagger = z, \quad a = \frac{\partial}{\partial z}$$

We can calculate

$$[a, a^\dagger]\psi(z) = \frac{\partial}{\partial z}(z\psi) - z\frac{\partial \psi}{\partial z} = \psi$$

so that the commutation relation is satisfied. Comparing

$$\frac{\partial}{\partial z} z^n = n z^{n-1}$$

with

$$a|n\rangle = \sqrt{n}|n-1\rangle$$

we get the correspondence

$$|n\rangle = \frac{z^n}{\sqrt{n!}}.$$

The inner product is given by integration with some weight function (measure)

$$||\psi(z)||^2 = \int \psi^*(z)\psi(z)\rho(z)d^2z$$

Here $\int d^2z$ denotes integration over the whole complex plane. $\rho(z)$ needs to vanish sufficiently rapidly at infinity so that z^n has a finite length. We need

$$\langle m|n\rangle = \delta_{mn}$$

so that

$$\int \frac{z^{*m}}{\sqrt{m!}} \frac{z^n}{\sqrt{n!}} \rho(z) d^2z = \delta_{mn}$$

This determines $\rho(z)$ uniquely:

$$\rho(z) = \frac{1}{\pi} e^{-|z|^2}$$

Exercise 28. Prove that

$$\int e^{-|z|^2} \frac{z^{*m}}{\sqrt{m!}} \frac{z^n}{\sqrt{n!}} \frac{d^2z}{\pi} = \delta_{mn}$$

Solution

First of all, we make the changes of variable

$$z = x + iy, \quad r = \sqrt{x^2 + y^2}, \quad u = r^2$$

to get

$$\int e^{-|z|^2} d^2z = \int e^{-x^2-y^2} dxdy = \int_0^\infty e^{-r^2} 2\pi r dr = \pi \int_0^\infty e^{-u} du = \pi$$

More generally

$$z = re^{i\theta}$$

gives

$$\begin{aligned} \int e^{-|z|^2} z^{*m} z^n d^2z &= \int_0^\infty e^{-r^2} r^m e^{-im\theta} r^n e^{in\theta} r dr d\theta \\ &= \int_0^\infty e^{-r^2} r^{m+n} r dr \int_0^{2\pi} e^{i(n-m)\theta} d\theta \end{aligned}$$

The last factor is only non-zero if $m = n$:

$$\begin{aligned} \int_0^{2\pi} e^{i(n-m)\theta} d\theta &= 2\pi\delta_{mn} \\ \int e^{-|z|^2} z^{*m} z^n d^2z &= 2\pi\delta_{mn} \int_0^\infty e^{-r^2} r^{m+n} r dr \\ &= \pi\delta_{mn} \int_0^\infty e^{-u} u^m du \\ \int e^{-|z|^2} z^{*m} z^n d^2z &= \pi m! \delta_{mn} \end{aligned}$$

from which the results follows.

Exercise. With this inner product show that z and $\frac{\partial}{\partial z}$ are hermitian conjugates of each other

$$\begin{aligned} \left\langle \phi, \frac{\partial}{\partial z} \psi \right\rangle^* &= \left[\int \phi^*(z) \frac{\partial \psi}{\partial z} e^{-z^* z} \frac{d^2z}{\pi} \right]^* \\ &= \int \phi(z) \frac{\partial \psi^*}{\partial z^*} e^{-z^* z} \frac{d^2z}{\pi} \end{aligned}$$

$$
\begin{aligned}
&= \int \frac{\partial}{\partial z^*}\left[\phi(z)\psi^* e^{-z^* z}\right]\frac{d^2z}{\pi} - \int \phi(z)\psi^* \frac{\partial\left[e^{-z^* z}\right]}{\partial z^*}\frac{d^2z}{\pi} \\
&= -\int \phi(z)\psi^*(z^*) z e^{-z^* z}\frac{d^2z}{\pi} \\
&= \langle \psi, z\phi\rangle
\end{aligned}
$$

as needed.

8.6.5. *All irreducible unitary representations of the Heisenberg algebra are equivalent to each other*

The Schrodinger representation, the Heisenberg representation, the complex function representation are all equivalent to each other:

$$H_n(q)e^{-\frac{q^2}{2}} \equiv |n\rangle \equiv \frac{z^n}{\sqrt{n!}}$$

where H_n are the Hermite polynomials [7].

8.7. Bosonic States as Polynomials

8.7.1. *If the space of states of a single boson is V, that of a pair of bosons is $S^2(V)$, the space of symmetric matrices*

In some orthonormal basis, single particle states are given by vectors $\psi = (\psi_1, \ldots, \psi_M)$ while two boson states are

$$\psi_{ij} = \psi_{ji}.$$

For fermions we would have anti-symmetric matrices. Suppose $i = 1, \ldots, M$: the single boson has some finite number M of states available to it. Then there are $\frac{M(M+1)}{2}$ independent two-boson states. More generally

8.7.2. *The space of states of n bosons is $S^n(V)$ the space of symmetric tensors*

$$\psi_{i_1 \cdots i_a \cdots i_b \cdots i_n} = \psi_{i_1 \cdots i_b \cdots i_a \cdots i_n}$$

invariant under any interchange. For fermions we would get anti-symmetric tensors that change sign under odd permutations.

8.7.3. *The total state space of bosons is* $S(V) = \bigoplus_{n=0}^{\infty} S^n(V)$

$S^0(V) = C$ is the vacuum or empty state, represented by a tensor of rank zero: A scalar. This space of symmetric tensors you can build out of a vector space is called its "Bosonic Fock space".

8.7.4. *We can also think of* $S(V)$ *as the space of polynomials*

$$\psi(z) = \sum_{n=0}^{\infty} \psi_{i_1 \cdots i_n} z_{i_1} \cdots z_{i_n}$$

The degree of the polynomial is the total number of bosons. In any given state this is a finite number, but we allow it to be as large as needed. Because the components of the complex numbers commute the coefficients are symmetric tensors.

8.7.4.1. *In the special case* $\dim V = 1$ *there is a correspondence between free bosonic states and the states of a simple harmonic oscillator*

There is just one polynomial of degree n, namely z^n. It corresponds to the state with occupation number n:

$$\mid n\rangle = \frac{z^n}{\sqrt{n!}}$$

The factor $\frac{1}{\sqrt{n!}}$ ensures that it has norm one with respect to

$$||\psi(z)||^2 = \int e^{-|z|^2} \psi^*(z)\psi(z) \frac{d^2 z}{\pi}$$

8.7.5. *The space of polynomials* $S(V)$ *carries a representation of the canonical commutation relations*

$$[a_i, a_j^\dagger] = \delta_{ij}, \quad [a_i, a_j] = 0 = [a_i^\dagger, a_j^\dagger]$$

We are to think of $a_i^\dagger$ as the multiplication of the polynomial representing the state by z_i; and a_i is the differentiation of this polynomial w.r.t. z_i:

$$a_i^\dagger = z_i, \quad a_i = \frac{\partial}{\partial z^i}$$

We can

$$|n_1, n_2 \cdots\rangle = \frac{z_1^{n_1}}{\sqrt{n_1!}} \frac{z_2^{n_2}}{\sqrt{n_2!}} \cdots$$

$$\psi(z) = \sum_{n_i=0}^{\infty} \psi_{n_1 n_2 \cdots} |n_1, n_2 \cdots\rangle$$

$$||\psi(z)||^2 = \int e^{-\sum_i |z|^2} \psi^*(z)\psi(z) \frac{d^2 z_1 d^2 z_2 \cdots}{\pi}$$

8.8. The Symplectic Lie Group and its Lie Algebra

Recall that momentum and position play symmetric roles in classical mechanics. We can even mix them in performing canonical transformations, as long as the Poisson brackets (canonical commutation relations) are preserved. Let us study them a little more in depth.

Since we want to treat momentum and position at the same footing let us introduce a common notation

$$\xi^a = \begin{pmatrix} p_1 \\ p_2 \\ \cdots \\ q^1 \\ q^2 \\ \cdots \end{pmatrix}, \quad a = 1, \ldots 2n$$

The canonical Poisson brackets become

$$\left\{\xi^a, \xi^b\right\} = \Omega^{ab}$$

the matrix Ω is anti-symmetric (because the Poisson bracket is anti-symmetric).

$$\Omega^{ab} = -\Omega^{ba}.$$

It is non-degenerate; i.e., its determinant is non-zero. Its inverse is often called the "symplectic form".

In the co-ordinate system above it has the explicit form $\Omega = \begin{pmatrix} 0_n & 1_n \\ -1_n & 0_n \end{pmatrix}$ where 1_n is the $n \times n$ identity matrix. But we can use other systems as well.

A linear canonical transformation will be a $2n \times 2n$ matrix Λ such that the matrix Ω unchanged. That is

$$\xi^a \mapsto \Lambda^a_c \xi^c$$

$$\{\xi^a, \xi^b\} \mapsto \Lambda^a_c \Lambda^b_d \{\xi^c, \xi^d\}$$

so that

$$\Lambda^a_c \Lambda^b_d \Omega^{cd} = \Omega^{ab}$$

or in matrix notation

$$\Lambda \Omega \Lambda^T = \Omega.$$

The set of matrices that satisfy this condition form a group called the Symplectic Group $Sp(n)$. It is analogous to the Orthogonal group, except that it preserves an anti-symmetric tensor rather than a symmetric one.

Exercise 29. Show that an anti-symmetric matrix with non-zero determinant must have even dimension.

Solution $\det \Omega = \det \Omega^T$ and $\Omega^T = -\Omega$ imply that $\det \Omega = (-1)^{\dim \Omega} \det \Omega$.

Exercise 30. Show that there is a real linear transformation which can brings an anti-symmetric non-degenerate matrix to the standard above $\Omega = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. (**Hint** i times an anti-symetric matrix is hermitian. Diagonalize it using a complex transformation and then rewrite in terms of real and imaginary parts.)

8.8.1. *The Symplectic Lie algebra*

As always the Lie algebra is easier to understand. A symplectic matrix infinitesimally close to the identity

$$\Lambda = 1 + \epsilon \lambda$$

will satisfy

$$\lambda \Omega + \Omega \lambda^T = 0$$

If we define

$$s = \lambda \Omega$$

this is the condition that

$$s = s^T$$

(The point is that Ω is anti-symmetric, so that $s^T = \Omega^T \lambda^T = -\Omega \lambda^T$). Thus, the symplectic Lie algebra $sp(n)$ consists of symmetric matrices. Compare with $so(n)$ which is made of anti-symmetric matrices.

In particular

$$\dim sp(n) = \frac{2n(2n+1)}{2} = n(2n+1).$$

The simplest case is $sp(1)$ which is three-dimensional.

There is another way to understand the Symplectic Lie algebra. Given a quadratic function

$$h(\xi) = \frac{1}{2} h_{bc} \xi^b \xi^c$$

the infinitesimal canonical transformation generated by it is a linear transformation

$$\{h, \xi^a\} = \frac{1}{2} h_{bc} \Omega^{ba} \xi^c + \frac{1}{2} h_{bc} \xi^b \Omega^{ca} = (\lambda \xi)^a$$

where

$$\lambda^a_c = h_{bc} \Omega^{ba}.$$

Such infinitesimal transformations preserve the Poisson brackets. (The proof is straightforward: It uses the Jacobi identity. Since the above formula is a one-one correspondence between quadratic functions and symmetric tensors, we see that all elements of $sp(n)$ arises this way from some function (called the generator).

Moreover, the Poisson bracket of two quadratic functions is another quadratic function:

$$\{f, h\} = \Omega^{ab} \partial_b f \partial_c h = \Omega^{bc} f_{bd} h_{ce} \xi^d \xi^e$$

Since the Poisson bracket is anti-symmetric and satisfies the Jacobi identity, the set of quadratic functions is a Lie algebra. This Lie algebra is $sp(n)$.

8.8.2. *A Representation of* $sp(1)$

The functions p^2, q^2, pq form a basis for $sp(1)$. The operators

$$\psi \mapsto -\frac{\partial^2 \psi}{\partial q^2}, \quad q^2 \psi(q), \quad -i \left[\frac{\partial}{\partial q} (q\psi) + q \frac{\partial \psi}{\partial q} \right]$$

provide a representation (the Schrodinger representation) for these basis elements of $sp(1)$. This representation is unitary w.r.t. the inner product

$$\langle \psi, \phi \rangle = \int \psi^*(q) \phi(q) dq$$

Of course, this is an infinite dimensional representation. One of the elements of $sp(1)$ is the hamiltonian of the

harmonic oscillator

$$H = \frac{1}{2}(p^2 + q^2)$$

The partition function

$$Z(\beta) = \mathrm{tr} e^{-\beta H} = \frac{2}{\sinh \frac{\beta}{2}}$$

can now be seen to be the character of this representation. The trace does not converge for all quadratic operators; only the positive ones. (In general the character makes sense as a formal power series or a distribution.)

More generally, the Bosonic Fock space carries a representation of $sp(n)$.

Exercise 31. Show that the above representation of $sp(1)$ is reducible as the direct sum of two invariant subspaces; one corresponding to even functions of q and other to odd functions. Each of these subspaces is irreducible, although this is harder to prove.

8.8.3. **Exponentiating to a Representation of the Group*

Recall that the odd spin representations of the orthogonal Lie algebra do not provide a representation of $SO(3)$ but of its extension $SU(2)$. Something similar happens for $SP(1)$. The above representation of the Lie algebra does not lead to representation of $SP(1)$, but of its extension by Z_2 called the "Metaplectic" group. To understand this, consider the element $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \in SP(1)$. This is a rotation in the (p, q) plane. As θ varies from 0 to 2π it describes a closed curve in $SP(1)$ starting and ending at the identity. In the above representation, the infinitesimal generator of a rotation in the p, q plane is the hamiltonian of the harmonic oscillator. (Recall that the orbits of the harmonic oscillator are circles). So, the corresponding curve is the $e^{i\theta H}$. At $\theta = 0$ it is the identity operator. If the eigenvalues of H had been integers, $e^{i\theta H}$ would have become identity at $\theta = 2\pi$. But we know that the eigenvalues of H are half-integers; the ground state energy is $\frac{1}{2}$, the excited energy is $\frac{3}{2}$ and so on. Thus, $e^{i\theta H} = -1$ at $\theta = 2\pi$. As we increase θ to 4π, $e^{i\theta H}$does become the identity again.

Remark 32. Often we add a multiple of the identity to the hamiltonian so that the ground state has zero energy. This does not affect many physical results. But it will change the commutation relations: The commutation relations would be spoilt by the addition of a term proportional to the identity. There are some situations (e.g. when there are an infinite number of

degrees of freedom) where we *have* to add a constant to the hamiltonian to make the vacuum expectation value finite ("normal order" the hamiltonian). The resulting Lie algebra (with an extra generator which commutes with everything) is a "central extension" of the Symplectic Lie algebra. So, instead of an extensionof the group SP by Z_2 (see below) we have an extension[14] by $U(1)$.

This is all similar to the odd spin representation of the Lie algebra $so(3)$; its exponentiation gives a representation not of $SO(3)$ but of its double cover $SU(2)$. In the same way the unitary matrices obtained by exponentiating the representation of $sp(1)$ on the harmonic oscillator states give a representation of a double cover of $SP(1)$ called $MP(1)$. Weil coined the word "Metaplectic" to describe this group (as well as the representation of $sp(1)$ on the harmonic oscillator). We can rephrase this in the language of group extensions: There is an exact sequence of group homomorphisms

$$\{1\} \to Z_2 \to MP(1) \to SP(1) \to \{1\}.$$

We omit the proofs of these statements. The exquisite mathematics (with deep connections to number theory) can be found in the book by Folland [16] and in the original paper of Bargmann [17].

There is an important difference with the case of $so(3)$ however. The group $SO(3)$ is doubly connected (has fundamental group Z_2), so that $SU(2)$ (its double cover) is simply connected. By contrast, $SP(1)$ has fundamental group $\mathbb{Z}$: It is infinitely connected. Indeed, as a manifold $SP(1)$ is diffeomorphic to $\mathbb{R}^2 \times \mathbb{S}^1$; the "noncompact" directions $\left(\text{which are tangential to } \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \text{ and } \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right)$ correspond the contractible part $\mathbb{R}^2$; the rotations $\left(\text{tangential to } \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\right)$ correspond to the circle. The Metaplectic group is also diffeomorphic to $\mathbb{R}^2 \times \mathbb{S}^1$; except that its circle is a double cover of the circle in $SP(1)$.

There is a simply connected Lie group $\tilde{SP}(1)$ which is diffeomorphic to $\mathbb{R}^3$; it has the same Lie algebra as $SP(1)$. But it does not have an easy description. See [17].

8.9. The Orthogonal Lie Algebra

We now return to the case of linear transformations that preserve a positive inner product (i.e., a symmetric positive matrix or metric tensor). There is always an orthogonal co-ordinate system (i.e., cartesian) in which the metric tensor is simply the identity matrix.

An infinitesimal rotation is generated by the analogues of the orbital angular momentum operators

$$L_{ab} = x_a \partial_b - x_b \partial_a$$

Their effect on the co-ordinates are easily worked out

$$[L_{ab}, x_c] = \delta_{bc} x_a - \delta_{ac} x_b$$

This says that L_{ab} only affects x_a or x_b which are rotated into each other: It is a rotation in the ab plane.

The commutation relations of the Lie algebra follows also by direct computation:

$$[L_{ab}, L_{cd}] = \delta_{bc} L_{ad} - \delta_{ac} L_{bd} - \delta_{bd} L_{ac} + \delta_{ad} L_{ac}$$

If the dimension of the underlying Euclidean space is even, (say $2n$) the following generators form a maximal set of commuting elements:

$$L_{12}, L_{34}, \ldots, L_{2n-1,2n}$$

Since there are n of them we see that is the rank of the $o(2n)$ Lie algebra. The dimension is

$$\dim o(2n) = \frac{2n(2n-1)}{2} = n(2n-1)$$

If the dimension is odd ($2n+1$ say) the maximal commuting set is again

$$L_{12}, L_{34}, \ldots, L_{2n-1,2n}$$

and the dimension is

$$\dim o(2n+1) = \frac{2n(2n+1)}{2} = n(2n+1)$$

The case $o(3)$ is of rank one and dimension three, as we already know.

There are in addition, polynomials in the generators (analogous to L^2) which also commute with all the L_{ab}. The number of them that are algebraically independent is equal to the rank:

$$L_{ab} L_{ab}, \quad L_{ab} L_{bc} L_{ca}, \quad \cdots L_{a_1 a_2} L_{a_3 a_4} \cdots L_{a_n a_{n+1}}$$

The irreducible representations of the Orthogonal group are given by traceless tensors of various symmetry types: unlike in the case $n = 1$, it is in general possible to make tensors that are neither symmetric nor anti-symmetric (mixed type). The irreducible representations of mixed type are constructed using an elaborate theory of "Young Tableaux" which we skip.

As in the case $m = 3$, there are representation of the Lie algebras $o(m)$ which do not exponentiate to a representation of $O(m)$; instead they exponentiate to representations of the double cover of the group (extension by Z_2) called the "spin group".

$$1 \to Z_2 \to \mathrm{Spin}(m) \to O(m) \to 1$$

These spin representations are of great physical interest, for all values of m (and even in the limit $m \to \infty$).

8.10. Clifford Algebra

Recall that the Pauli matrices satisfy

$$\sigma_1\sigma_2 + \sigma_2\sigma_1 = 0, \quad \sigma_1^2 = 1 = \sigma_2^2$$

Given σ^1, σ^2 we can get

$$\sigma_3 = i\sigma_1\sigma_2$$

and verify that the algebra extends to three dimensions:

$$\sigma_i\sigma_j + \sigma_j\sigma_i = 2\delta_{ij}, \quad i, j = 1, 2, 3$$

Then we find that the matrices provide a representation of the rotation Lie algebra. Also,

$$b^\dagger = \frac{\sigma_1 + i\sigma_2}{2}, \quad b = \frac{\sigma_1 - i\sigma_2}{2}$$

satisfy the Canonical anti-commutation relations (are the creation and annihilation operators for fermions).

All this generalizes to higher dimensions, and is central to our understanding of spin and fermions.

8.10.1. *The Clifford algebra on $\mathbb{R}^{2n}$ is generated by matrices satisfying the relations*

$$\Gamma_a\Gamma_b + \Gamma_b\Gamma_a = \delta_{ab}, \quad a, b = 1, 2, \ldots, 2n$$

Let us first consider the case where a, b take the range of values $1, 2, 3, 4$. Once we understand this, we will see how to extend to any even dimension $2n$. A simple trick will then allow us to pass to odd dimension $2n + 1$ as well.

Let us start with the ansatz

$$\Gamma_1 = \sigma_1 \otimes 1, \quad \Gamma_2 = \sigma_2 \otimes 1$$

If we choose also

$$\Gamma_3 = \sigma_3 \otimes \sigma_1, \quad \Gamma_4 = \sigma_3 \otimes \sigma_2$$

we get what we want! The point is that σ_3 anti-commutes with both σ_1 and σ_2 so that the last two matrices anti-commute with the first pair.

Now, if the dimension is six, we choose

$$\Gamma_1 = \sigma_1 \otimes 1 \otimes 1, \quad \Gamma_2 = \sigma_2 \otimes 1 \otimes 1$$

$$\Gamma_3 = \sigma_3 \otimes \sigma_1 \otimes 1, \quad \Gamma_4 = \sigma_3 \otimes \sigma_2 \otimes 1$$

and

$$\Gamma_5 = \sigma_3 \otimes \sigma_3 \otimes \sigma_1, \quad \Gamma_6 = \sigma_3 \otimes \sigma_3 \otimes \sigma_2$$

Now, you get the idea for the general case.

$$\Gamma_{2k-1} = \sigma_3 \otimes (k-1 \text{ times}) \cdots \otimes \sigma_3 \otimes \sigma_1 \otimes 1(n-k \text{ times}) \cdots \otimes 1,$$

$$\Gamma_{2k} = \sigma_3 \otimes (k-1 \text{ times}) \cdots \otimes \sigma_3 \otimes \sigma_2 \otimes 1(n-k \text{ times}) \cdots \otimes 1$$

for

$$k = 1, \ldots, n$$

This is a representation of the Clifford algebra on an even dimensional Euclidean space. The representation matrices act on a vector space of dimension 2^n. The space on which they act consists of *spinors*.

Note that the Clifford matrices are hermitian (direct products of hermitian matrices.)

8.10.2. *The matrix product $\Gamma_1 \cdots \Gamma_{2n}$ anti-commutes with all the elements Γ_a*

The point is each Γ_a commutes with itself but anti-commutes with the others; so we get an odd number of negative signs as we pass Γ_a through $\Gamma_1 \cdots \Gamma_{2n}$. Moreover

$$(\Gamma_1 \cdots \Gamma_{2n})^\dagger = \Gamma_{2n} \cdots \Gamma_1 = (-1)^{2n-1+2n-2+\cdots} \Gamma_1 \cdots \Gamma_{2n}$$

$$= (-1)^{n(2n+1)} \Gamma_1 \cdots \Gamma_{2n} = \begin{cases} -\Gamma_1 \cdots \Gamma_{2n} & n \text{ odd} \\ \Gamma_1 \cdots \Gamma_{2n} & n \text{ even} \end{cases}$$

So, if we define

$$\Gamma = i^n \Gamma_1 \cdots \Gamma_{2n}$$

we will find that

$$\Gamma^\dagger = \Gamma, \quad \Gamma^2 = 1.$$

8.10.3. *Bilinears of the Clifford matrices provide a unitary representation of $o(2n)$*

$$\Sigma_{ab} = \frac{1}{2}[\Gamma_a, \Gamma_b]$$

They satisfy the commutation relations of $o(2n)$

$$[\Sigma_{ab}, \Sigma_{cd}] = \delta_{bc}\Sigma_{ad} - \delta_{ac}\Sigma_{bd} - \delta_{bd}\Sigma_{ac} + \delta_{ad}\Sigma_{bc}$$

Exercise. Prove this by direct calculation.

Moreover, they are anti-hermitian matrices

$$\Sigma_{ab}^\dagger = -\Sigma_{ab}$$

Also, the Clifford matrices transform as a collection of vectors:

$$[\Sigma_{ab}, \Gamma_c] = \delta_{ac}\Gamma_b - \delta_{bc}\Gamma_a$$

while Γ is a scalar:

$$[\Sigma_{ab}, \Gamma] = 0.$$

Since Γ is not a multiple of the identity, the representation of $o(2n)$ is reducible. The two eigenspaces of Γ (with eigenvalues ± 1) do provide irreducible representations of $o(2n)$. (Takes a bit more work to prove this.)

8.10.4. *Using Γ_a we can construct a representation of the Clifford algebra over odd dimensional Euclidean spaces $\mathbb{R}^{2n+1}$*

We just have to define

$$\Gamma_{2n+1} = \Gamma$$

Then

$$\Gamma_{2n+1}^2 = 1$$

and it anti-commutes with all the others:

$$\Gamma_a\Gamma_b + \Gamma_b\Gamma_a = 2\delta_{ab}, \quad a, b = 1 \cdots 2n+1$$

8.10.5. *Using the Clifford matrices we can construct the Dirac Operator on spinors, which is a "square root" of the Laplace operator*

Suppose we have spinors that depend on position. Define

$$D\psi = \Gamma_a \frac{\partial \psi}{\partial x^a}$$

This is a rotation invariant[3] operator: both $\frac{\partial \psi}{\partial x^a}$ and Γ_a transform as vectors. It is called the Dirac operator.

Then

$$D^2\psi = \Gamma_a \frac{\partial}{\partial x^a}\left(\Gamma_b \frac{\partial \psi}{\partial x^b}\right)$$

$$= \Gamma_a \Gamma_b \frac{\partial^2 \psi}{\partial x^a \partial x^b}$$

since $\frac{\partial^2 \psi}{\partial x^a \partial x^b}$ is symmetric in ab,

$$= \frac{1}{2}\left(\Gamma_a \Gamma_b + \Gamma_b \Gamma_a\right) \frac{\partial^2 \psi}{\partial x^a \partial x^b}$$

Thus the square of the Dirac operator is the Laplace operator:

$$D^2\psi = \delta_{ab} \frac{\partial^2 \psi}{\partial x^a \partial x^b}$$

This will be important when we study relativistic wave equations. The Dirac operator can be extended to Riemannian manifolds and provides much subtle topological information not contained in the Laplace operator.

[3]Dirac originally discovered this operator while looking for a wave equation for the electron. The group involved there is not an orthogonal group, but the Lorentz group $O(1,3)$.

This page intentionally left blank

Chapter 9

THE ISING MODEL

A phase transition is a spectacular phenomenon in thermodynamics. The most familiar example is the boiling of water as it turns to steam. This is a first order transition: There is a discontinuous change in the energy, called the latent heat: The amount of energy needed to convert water to steam of the same temperature. As we increase the pressure, the latent heat decreases. At a critical pressure and temperature, the latent heat vanishes so that the energy is continuous. But the derivative w.r.t. temperature (specific heat) is infinite at that point: This is called a *critical phase transition*. (For water the the critical temperature is 374 C and the critical pressure is 218 atmospheres; for nitrogen it is -147 C and 34 atmospheres.) You cannot liquefy a gas above the critical temperature no matter how much pressure you apply.

Another important example of a critical phase transition occurs in a magnet. At low temperatures a ferromagnet (indeed iron is an example) has most of its magnetic moments point in the same direction. But above a critical temperature (the Curie point) these magnets are randomly oriented. Remarkably, the singularities at the critical point is the same for gases and magnets, even though the underlying physical processes are very different. Understanding of this surprising phenomenon of universality is one of the great achievements of twentieth century physics.

Onsager made the first step towards understanding critical phase transitions, by solving a two dimensional model for a magnet exactly. This is possible because of a surprising connection with spinors: The transfer matrix of the magnet on an $M \times M$ square lattice can be shown to be a spinor representative of a rotation in $2M$dimensions. Using the representation theory of $o(2M)$ the partition function can be determined exactly. As $M \to \infty$ the free energy has a singularity, corresponding to a critical phase transition.

This showed for the first time that summing over molecular degrees of freedom can lead to singularities in the free energy: Until then it was only a conjecture that phase transitions could be explained this way.

9.1. The Hamiltonian

Our aim is to build the simplest model of a large number of molecules with magnetic moments. The collective behavior of these magnets must lead to a low temperature phase in which most moments point in the same direction (the ordered phase). At high temperatures, there should be a disordered phase in which the magnetic moments average to zero. A molecule has a magnetic moment that is proportional to its angular momentum (spin). The proportionality constant (the gyromagnetic ratio) depends on the structure of the molecule and is not important for us. In the simplest model, the Ising model, this magnetic moment (or spin) can point in one of two directions: We have a variable $s = \pm 1$ at each molecule that describes this orientation. There are more intricate models where the spin can lie on a circle (the XY model) or on a sphere (the Heisenberg model) but we look only at the simplest case. The molecules are arranged on a cubic lattice (other lattices can be chosen as well) of L sites in each direction.

Two neighboring spins will interact with an energy $-\tilde{J}ss'$. If the constant $\tilde{J} > 0$, the spins will have a tendency to align producing a ferromagnet at low temperatures[1]. This interaction is due to an intricate quantum mechanical phenomenon involving tunneling, (the exchange interaction) and decays exponentially with distance (as chemical bonds do as well). So we can ignore the interaction except for molecules that are very close together. That is why we only include interactions among nearest neighbors.

Thus the magnetic energy of the Ising model is

$$H(\sigma) = -\tilde{J} \sum_{x-y} s_x s_y - \tilde{B} \sum_x s_x$$

where $x - y$ denotes two positions on the lattice that are connected by a nearest neighbor bond. The quantity $\tilde{B}$ is the externally imposed magnetic field (times the gyromagnetic ratio of the molecule). The partition function of this system is

$$Z_L(J) = \sum_{s=\pm 1} e^{J \sum_{x-y} s_x s_y + B \sum_x s_x}$$

[1] If $\tilde{J} < 0$ the spins will try to be opposite: An anti-ferromagnet.

where

$$J = \frac{\tilde{J}}{k_B T}, \quad B = \frac{\tilde{B}}{k_B T}$$

and k_B is Boltzmann's constant and T is temperature.

Recall that the thermodynamic free energy is the limit of a large number of molecules

$$W(J, B) = \lim_{L \to \infty} -\frac{\log Z_L(J, B)}{L^3}$$

From this every other thermodynamic quantity (specific heat, magnetization) etc. can be calculated by standard formulas in thermodynamics.

No one has been able to get an analytic formula for the free energy for the cubic lattice. It is commonly accepted that this is impossible in terms of the usual functions known to mathematical physicists.

Onsager solved the two dimensional Ising model; i.e., on a square lattice. He was building on a technique developed by Ising for the one dimensional lattice, called the transfer matrix method. Although the one dimensional model is too simple (it does not have a phase transition) it is a good place to start.

9.2. Transfer Matrix of the 1D Ising Model

Imagine a long chain of L spins arranged at regular intervals along a line. Each spin interacts with its two nearest neighbors. We will eventually take the limit $L \to \infty$. The hamiltonian can be written as

$$H = -\tilde{J} \sum_{x=1}^{L-1} s_x s_{x+1} - \tilde{B} \sum_x s_x$$

A more symmetric way to write this is

$$H = -\tilde{J} \sum_{x=1}^{L-1} s_x s_{x+1} - \frac{1}{2} \tilde{B} \sum_x (s_x + s_{x+1})$$

It is convenient to split the term involving a single spin as an average of nearest neighbors.[2] The partition function is then

$$Z_L(J, B) = \sum_{s=\pm 1} e^{J s_1 s_2 + \frac{1}{2} B(s_1 + s_2)} e^{J s_2 s_3 + \frac{1}{2} B(s_2 + s_3)} e^{J s_3 s_4 + \frac{1}{2} B(s_3 + s_4)}$$

$$\cdots e^{J s_{L-1} s_L + \frac{1}{2} B(s_{L-1} + s_L)}$$

[2]The boundary spins are counted with half the strength; we will see that they don't matter much anyway in the limit $L \to \infty$

Define the 2×2 matrix labeled by $s, s' = \pm 1$

$$T_{ss'} = e^{Jss' + \frac{1}{2}B(s+s')}$$

That is,

$$T = \begin{pmatrix} e^{J+B} & e^{-J} \\ e^{-J} & e^{J-B} \end{pmatrix} \equiv \begin{pmatrix} \frac{1}{uv} & u \\ u & \frac{v}{u} \end{pmatrix}$$

with

$$u = e^{-J}, \quad v = e^{-B}.$$

This is called the transfer matrix: It transfers us by one step along the lattice. Then

$$Z_L(J) = \sum_{s_1, s_2 \cdots = \pm 1} T_{s_1 s_2} T_{s_2 s_3} T_{s_3 s_4} \cdots T_{s_{L-1} s_L}$$

The simplifying feature of the one dimensional model is that the s_2 only appears in the first two factors, s_3 only in the second and third and so on. The sum over $s_2, s_3 \cdots s_{L-1}$ can be thought of as matrix multiplication.

$$Z_L(J, B) = \sum_{s_1, s_L} T^{L-1}_{s_1 s_L}$$

The power of a symmetric matrix T^{L-1} can be calculated conveniently in terms of its eigenvalues (which are real) and eigenvectors (which can be chosen to be orthonormal).

$$T\psi_1 = \lambda_1 \psi_1, \quad T\psi_2 = \lambda_2 \psi_2$$

$$\psi_1^T \psi_1 = 1 = \psi_2^T \psi_2, \quad \psi_1^T \psi_2 = 0.$$

$$T^L = \lambda_1^L \psi_1 \psi_1^T + \lambda_2^L \psi_2 \psi_2^T .$$

Suppose $|\lambda_1| > |\lambda_2|$.

$$T^L = \lambda_1^L \left[\psi_1 \psi_1^T + \left(\frac{\lambda_2}{\lambda_1} \right)^L \psi_2 \psi_2^T \right] \to \lambda_1^L \psi_1 \psi_1^T$$

as $L \to \infty$. Thus the free energy is simply the log of the largest eigenvalue of the transfer matrix:

$$W = - \lim_{L \to \infty} \left[\log \lambda_1 + \frac{1}{L} \log \psi_1 \psi_1^T \right] = - \log \lambda_1$$

A little algebra gives

$$\lambda_1 = \frac{1 + u^2 + \sqrt{4u^4v^2 + u^4 - 2u^2 + 1}}{2uv}$$

The only singularities (poles or branch cuts) in the physical region $u, v \geq 0$ are at $u = 0, 1$ or $v = 0, 1$; these correspond to zero or infinite temperature. The one dimensional Ising model does not have a phase transition at any intermediate values. Still it is a starting point for more intricate models.

9.2.1. *A Trick with Pauli Matrices*

It will be useful to write T (with $B = 0$) in terms of the Pauli matrices:

$$T = \begin{pmatrix} e^J & e^{-J} \\ e^{-J} & e^J \end{pmatrix} = e^J 1_2 + e^{-J}\sigma_1$$

Since

$$e^{J'\sigma_1} = \cosh J' + \sigma_1 \sinh J'$$

we can write

$$T = ae^{J'\sigma_1}$$

where

$$a \cosh J' = e^J, \quad a \sinh J' = e^{-J}$$

Solving,

$$\tanh J' = e^{-2J}, \quad a^2 = e^{2J} - e^{-2J} = 2 \sinh 2J$$

This trick with Pauli matrices has a useful generalization for more complicated models.

9.3. Ising Model on an $L \times 2$ Ladder

The next setup is to consider the Ising model on a ladder of two rows of spins, coupling nearest neighbors along the row and columns (again we ignore the external magnetic field):

$$H = -\tilde{J} \sum_{x=1}^{L-1} \left[s_{x,1}s_{x+1,1} + s_{x,2}s_{x+1,2} \right] - \tilde{J} \sum_{x=1}^{L-1} s_{x,1}s_{x,2}$$

The same idea as before can work except we must think of the transfer matrix as a 4×4 matrix.

$$Z_L(J) = \sum_{s=\pm 1} e^{Js_{11}s_{21}+Js_{12}s_{22}+Js_{11}s_{12}} e^{Js_{21}s_{31}+Js_{22}s_{32}+Js_{21}s_{22}}$$
$$\dots e^{Js_{L-1,1}s_{L1}+Js_{L-1,2}s_{L2}+Js_{L-1,1}s_{L-1,2}}$$

Note that s_{21} and s_{22} only appear in the first two factors, s_{31} and s_{32}only in the second and third factor and so on. So we can again think of this as a matrix product, but we need matrices to be labelled by pairs of spins, each taking values ± 1 independently.: A 4×4 transfer matrix.

Define a "spin vector"

$$\boldsymbol{s} = (s_1, s_2)$$

with each component taking values ± 1 independently. Define the 4×4 matrix

$$\boldsymbol{T}_{\boldsymbol{ss}'} = e^{J\boldsymbol{s}\cdot\boldsymbol{s}'} e^{Js_1s_2}$$

This transfer matrix is a product of two matrices, a diagonal matrix

$$Q_{\boldsymbol{ss}'} = \delta_{\boldsymbol{ss}'} e^{J's_1s_2}$$

and a non-diagonal matrix

$$P_{\boldsymbol{ss}'} = e^{J\boldsymbol{s}\cdot\boldsymbol{s}'}$$

The precise way we treat the boundary spins s_1 and s_L will not matter for most values of temperature. Periodic bpundary conditions are mathematically convenient; i.e., impose $s_1 = s_L$ and sum over s_1. In that case

$$Z_N = \mathrm{tr}\boldsymbol{T}^{L-1}.$$

9.4. The Ising Model on an $L \times M$ lattice

If we have M rows, the same idea works except that

$$\boldsymbol{s} = (s_1, s_2, \dots s_M)$$

is a vector with M components and each one takes ± 1 values. The transfer matrix is a $2^M \times 2^M$ matrix which is again the product of a diagonal matrix

$$Q_{\boldsymbol{ss}'} = \delta_{\boldsymbol{ss}'} e^{J' \sum_{k=1}^{M} s_k s_{k+1}}$$

and a non-diagonal matrix

$$P_{\boldsymbol{ss}'} = e^{J \boldsymbol{s} \cdot \boldsymbol{s}'}$$

The problem is to find the eigenvalues of

$$\boldsymbol{T} = PQ.$$

As $L \to \infty$, the largest eigenvalue dominates.[3] The difficulty is that the size of the matrix grows exponentially with M.

9.4.1. *Clifford Algebra to the rescue*

Now we recall that the Clifford matrices in $\mathbb{R}^{2M}$ are also $2^M \times 2^M$ dimensional. Kaufmann found a surprising use for them in the Ising model. (Onsager had solved the problem earlier by a somewhat opaque method. Kaufmann, an associate of Onsager, clarified the solution greatly. Mere mortals could understand Onsager's idea after that.) Kerson Huang's wonderful book [20] on Statistical Mechanics has a much deeper discussion than ours.

9.4.1.1. *Transfer matrix in terms of Pauli matrices*

Recall that the transfer matrix of the one dimensional Ising model can be written in terms of Pauli matrices:

$$T = a e^{J' \sigma_1}$$
$$\tanh J' = e^{-2J}, \quad a = \sqrt{2 \sinh 2J}$$

The matrix P is the direct product of M copies of this T.

$$P_{\boldsymbol{ss}'} = T_{s_1 s_1'} T_{s_2 s_2'} \ldots T_{s_M s_M'}$$
$$P = a^M e^{J' \sigma_1} \otimes e^{J' \sigma_1} \cdots \otimes e^{J' \sigma_1}$$

[3] A more symmetrical choice of transfer matrix would be $Q^{\frac{1}{2}} P Q^{\frac{1}{2}}$. But it is a bit simpler to work with this PQ in this case.

In other words

$$P = a^M e^{J' \sum_{k=1} \sigma_k^1}$$

where

$$\sigma_k^1 = 1 \otimes 1 \otimes \cdots \otimes \sigma_1 \otimes \cdots \otimes 1$$

with 1 everywhere except at the kth site.

More obviously, the diagonal factor can be written in terms of σ_3:

$$Q = e^{J' \sum_{k=1}^{M} \sigma_k^3 \sigma_{k+1}^3}$$

9.4.1.2. *Clifford Bilinears in terms of Pauli matrices*

We can choose a representation of the Clifford matrices[4]:

$$\Gamma_1 = \sigma_3 \otimes 1 \otimes 1 \cdots, \qquad \Gamma_2 = \sigma_2 \otimes 1 \otimes 1 \cdots$$
$$\Gamma_3 = \sigma_1 \otimes \sigma_3 \otimes 1 \cdots, \qquad \Gamma_4 = \sigma_1 \otimes \sigma_2 \otimes 1 \cdots$$
$$\Gamma_5 = \sigma_1 \otimes \sigma_1 \otimes \sigma_3 \cdots, \quad \Gamma_6 = \sigma_1 \otimes \sigma_1 \otimes \sigma_2 \cdots$$

You can check that it still satisfies the Clifford Algebra; this representation is a bit more convenient for our current purpose.

Then

$$\Gamma_1 \Gamma_2 = -i\sigma_1 \otimes \cdots$$
$$\Gamma_3 \Gamma_4 = -i1 \otimes \sigma_1 \otimes \cdots$$
$$\Gamma_5 \Gamma_6 = -i1 \otimes 1 \otimes \sigma_1 \otimes \cdots$$

etc.

Thus

$$\sigma_k^1 = i\Gamma_{2k-1}\Gamma_{2k}$$
$$P = a^M e^{iJ' \sum_{k=1}^{M} \Gamma_{2k-1}\Gamma_{2k}}$$

Moreover

$$\Gamma_2 \Gamma_3 = -i\sigma_3 \otimes \sigma_3 \otimes \cdots$$
$$\Gamma_4 \Gamma_5 = -i1 \otimes \sigma_3 \otimes \sigma_3 \otimes \cdots$$

[4]Compared to section (8.10.1) we are interchanging σ_1 and σ_3

so that

$$Q = e^{iJ\sum_{k=1}^{M-1}\Gamma_{2k}\Gamma_{2k+1}}$$

Thus

$$\boldsymbol{T} = PQ = a^M e^{iJ'\sum_{k=1}^{M}\Gamma_{2k-1}\Gamma_{2k}} e^{iJ\sum_{k=1}^{M-1}\Gamma_{2k}\Gamma_{2k+1}}$$

Since Clifford bilinears represent infinitesimal rotations, and $\boldsymbol{T}$ is a product of their exponentials, it can be thought of as a spinor representation of a rotation matrix $O(2M)$. Strictly speaking, it is a rotation by an imaginary angle (if J is real) but that won't affect our algebraic considerations: We can always analytically continue the formulas.

9.4.1.3. *Self-duality*

Before we solve the problem completely, we can already notice a surprising symmetry. It is obvious that interchanging the matrices $\Gamma_{2k-1}\Gamma_{2k} \rightleftarrows \Gamma_{2k}\Gamma_{2k+1}$ is a symmetry: It amounts to a relabeling of indices. If we combined with $J \leftrightarrow J'$ this will change PQ to QP. Now, the partition function

$$Z_L = \mathrm{tr} PQPQ\cdots PQ$$

will go over to

$$\mathrm{tr} QPQP\cdots QP$$

which (by cyclic symmetry of the trace) is the same as Z_L.

So, the partition function of the Ising model has a symmetry

$$a^{-M} Z_M(J) = a'^{-M} Z_M(J')$$

This is a symmetry that relates high temperature to low temperature ($J \to 0 \implies J' \to \infty$). This "duality" was discovered by Elliott Montroll (by a different method) and was one of the first indications that the Ising model can be solved. In particular, it gives the exact location of the phase transition to be at the self-dual point: $J = J' \implies J = \frac{1}{2}\log(1+\sqrt{2}) \approx 0.440687$.

9.4.1.4. *From spinor to vector representation*

Recall that $\Sigma_{ab} = \frac{1}{4}[\Gamma_a, \Gamma_b]$ provide a representation of $o(2M)$ Lie algebra. The defining representation is by matrices generating rotations in $\mathbb{R}^{2M}$. Explicitly they

are matrices of the form

$$[\rho_{ab}]_{cd} = \delta_{ac}\delta_{bd} - \delta_{ad}\delta_{bc}$$

That is the cd component of the matrix ρ_{ab} is 1 if $ab = cd$, equal to -1 of $ab = dc$ and zero otherwise. For example,

$$\rho_{12} = \begin{pmatrix} 0 & 1 & 0 & \cdot & 0 \\ -1 & 0 & 0 & . & 0 \\ 0 & 0 & 0 & . & 0 \\ . & . & . & . & . \\ 0 & 0 & 0 & . & 0 \end{pmatrix}$$

The transfer matrix of the Ising model is a $2^M \times 2^M$ matrix:

$$\boldsymbol{T} = a^M R, \quad R = e^{2iJ' \sum_{k=1}^{M} \Sigma_{2k-1,2k}} e^{2iJ \Sigma_{2k,2k+1}}$$

It is the spin representative of the matrix

$$r = e^{2iJ' \sum_{k=1}^{M} \rho_{2k-1,2k}} e^{2iJ \rho_{2k,2k+1}}$$

This is a $2M \times 2M$ matrix: much smaller when M is large. There is an orthogonal matrix v which can reduce r to a canonical form

$$r = v e^{\sum_{k=1}^{M} \theta_k \rho_{2k-1,2k}} v^{-1}$$

This means R has the canonical form by the spin representative of V:

$$R = V e^{\sum_{k=1}^{M} \theta_k \Sigma_{2k-1,2k}} V^{-1}$$

Without knowing V explicitly, we can read off the eigenvalues of R. The matrices $\Sigma_{2k-1,2k}$ commute with each other and have eigenvalues $\pm\frac{1}{2}$. So, the eigenvalues of R are $e^{\pm \epsilon_k \frac{\theta_k}{2}}$ where $\epsilon_k = \pm 1$.

The partition function is given by the trace:

$$\mathrm{tr}R = \sum_{\epsilon_k = \pm 1} e^{\pm \epsilon_k \frac{\theta_k}{2}} = \prod_{k=1}^{M} \left[2 \cosh \frac{\theta_k}{2} \right]$$

Thus the problem of finding the partition function can be solved if we can find the "characteristic values" θ_k of r .

9.4.1.5. *Diagonalizing Cyclic matrices*

To determine the characteristic values θ_k of r, the trick is note that each factor has a simple form

$$e^{2iJ'\sum_{k=1}^{M}\rho_{2k-1,2k}} = \begin{pmatrix} \cosh 2J' & \sinh 2J' & 0 & 0 & \cdot & \cdot & \cdot \\ \sinh 2J' & \cosh 2J' & 0 & 0 & \cdot & \cdot & \cdot \\ 0 & 0 & \cosh 2J' & \sinh 2J' & \cdot & \cdot & \cdot \\ 0 & 0 & \sinh 2J' & \cosh 2J' & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \cdot & \cdot & \cdot \end{pmatrix}$$

$$e^{2iJ\sum_{k=1}^{M}\rho_{2k,2k+1}} = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdot & \cdot & \cdot \\ 0 & \cosh 2J & \sinh 2J & 0 & \cdot & \cdot & \cdot \\ 0 & \sinh 2J & \cosh 2J' & 0 & 0 & \cdot & \cdot \\ 0 & 0 & 0 & \cosh 2J & \sinh 2J & \cdot & \cdot \\ 0 & 0 & 0 & \sinh 2J & \cosh 2J & \cdot & \cdot \\ 0 & 0 & 0 & 0 & 0 & \cdot & \cdot \\ 0 & 0 & 0 & 0 & 0 & \cdot & \cdot \end{pmatrix}$$

Another way to write this is

$$\left[e^{2iJ'\sum_{k=1}^{M}\rho_{2k-1,2k}}\right]_{ab} = c_1\delta_{ab} + s_1\sum_k\left[\delta_{a,2k-1}\delta_{b,2k} + \delta_{a,2k}\delta_{b,2k-1}\right]$$

$$\left[e^{2iJ\sum_{k=1}^{M}\rho_{2k-1,2k}}\right]_{cd} = c_2\delta_{cd} + s_2\sum_l\left[\delta_{c,2l}\delta_{d,2l+1} + \delta_{c,2l+1}\delta_{b,2l}\right]$$

The product is some matrix of the form (setting $c_1 = \cosh J'$, $s_1 = \sinh J'$, $c_2 = \cosh J$, $s_2 = \sinh J$)

$$\begin{aligned}\left[e^{2iJ'\sum_{k=1}^{M}\rho_{2k-1,2k}}e^{2iJ\sum_{k=1}^{M}\rho_{2k-1,2k}}\right]_{ad} &= c_1c_2\delta_{ad} \\ &+ s_1c_2\sum_k\left[\delta_{a,2k-1}\delta_{d,2k} + \delta_{a,2k}\delta_{d,2k-1}\right] \\ &+ s_2c_1\sum_l\left[\delta_{a,2l}\delta_{d,2l+1} + \delta_{a,2l+1}\delta_{b,2l}\right] \\ &+ s_1s_2\sum_{kl}\left[\delta_{a,2k-1}\delta_{2l,2k}\delta_{d,2l+1} + \delta_{a,2k}\delta_{2k-1,2l+1}\delta_{b,2l}\right]\end{aligned}$$

The last term simplifies to

$$s_1 s_2 \sum_{kl} \left[\delta_{a,2k-1}\delta_{d,2k+1} + \delta_{a,2k}\delta_{b,2k-2}\right]$$

This matrix has the form (displayed below for with $M = 7$, but the pattern is valid generally)

$$\begin{pmatrix}
c_1c_2 & c_1s_2 & s_1s_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_2s_1 \\
c_1s_2 & c_1c_2 & c_2s_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & s_1s_2 \\
0 & c_2s_1 & c_1c_2 & c_1s_2 & s_1s_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & s_1s_2 & c_1s_2 & c_1c_2 & c_2s_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & c_2s_1 & c_1c_2 & c_1s_2 & s_1s_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & s_1s_2 & c_1s_2 & c_1c_2 & c_2s_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & c_2s_1 & c_1c_2 & c_1s_2 & s_1s_2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & s_1s_2 & c_1s_2 & c_1c_2 & c_2s_1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & c_2s_1 & c_1c_2 & c_1s_2 & s_1s_2 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & s_1s_2 & c_1s_2 & c_1c_2 & c_2s_1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_2s_1 & c_1c_2 & c_1s_2 & s_1s_2 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & s_1s_2 & c_1s_2 & c_1c_2 & c_2s_1 & 0 \\
s_1s_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_2s_1 & c_1c_2 & c_1s_2 \\
c_2s_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & s_1s_2 & c_1s_2 & c_1c_2
\end{pmatrix}$$

Each row is repeated two steps below, shifted by two permutations. With periodic boundary conditions this will become a symmetry under the cyclic group $\mathbb{Z}_M$. So it is a good guess that it has eigenvectors of the form

$$\psi = \begin{pmatrix} u \\ zu \\ z^2u \\ z^3u \\ \cdot \\ \cdot \\ \cdot \end{pmatrix}, \quad z^M = 1$$

where $u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$ is a two dimensional vector. This ansatz reduces the problem of diagonalizing r to an eigenvalue equation for a 2×2 matrix[20].

$$\begin{pmatrix} c_1c_2 + s_1s_2z & c_2s_1z^{-1} + c_1s_2 \\ c_2s_1z + c_1s_2 & c_1c_2 + s_1s_2z^{-1} \end{pmatrix} u = \lambda u$$

Extracting the thermodynamic information (e.g., the phase transition) from this is a rewarding exercise in statistical mechanics, but it does not involve any more group theory. Read more in [20].

In three dimensions, the Ising model has not been solved exactly; may be it is not possible. But the ideas introduced by Wilson (based on renormalization) have given us a deep understanding of the physics. Further ideas using conformal symmetry ("conformal bootstrap" of Rychkov) also have been effective. These are beyond the scope of this book.

This page intentionally left blank

Chapter 10

WAVE EQUATIONS

10.1. Lorentz Invariance

We saw that rotations are described by 3×3 matrices satisfying

$$R^T R = 1, \quad \det R = 1.$$

In relativity, space and time (similarly, energy and momentum) are combined into a single vector with four components. Our convention will be that time (or energy) is the zeroth component.

Its length^2 is given by the Minkowski rule:

$$p \equiv \begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \end{pmatrix}$$

$$p.p = E^2 - c^2 \mathbf{p}^2$$

It will be useful to think of this as

$$p.p = p^T \eta p, \quad \eta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -c^2 & 0 & 0 \\ 0 & 0 & -c^2 & 0 \\ 0 & 0 & 0 & -c^2 \end{pmatrix}$$

where the matrix η is called the "Minkowski metric". It will be convenient to use units with $c = 1$ in discussing relativistic physics. You can always convert

to ordinary units by dimensional analysis. We will use these units (mostly) from now on.

Thus mass is simply the length of the energy-momentum vector (up to a factor c^2). More generally the scalar product of a vector with itself may be positive (energy-momentum for massless particles), zero (massless particles) or negative. A vector with $p \cdot p < 0$ cannot be the energy-momentum of any particle: It would have imaginary mass. But such "space-like" vectors can represent other interesting things; like the separation between two points in space-time.

There are transformations analogous to rotations that preserve the length^2 of a vector. Obviously, any rotation is of this type. But there are also transformations that mix space and time. For example,

$$\Lambda = \begin{pmatrix} \cosh\theta & \sinh\theta & 0 & 0 \\ \sinh\theta & \cosh\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

This is an example of a Lorentz transformation. The variable θ (called rapidity) is analogous to an angle of rotation. Unlike an angle, it can vary from $-\infty$ to ∞.

Exercise 33. Verify that $(\cosh\theta p_0 + \sinh\theta p_1)^2 - (\sinh\theta p_0 + \cosh\theta p_1)^2 = p_0^2 - p_1^2$ for any θ.

The condition that a linear transformation $p \mapsto \Lambda p$ preserve the Minkowski dot product is

$$(\Lambda)^T \eta(\Lambda p) = p^T \eta p \iff \Lambda^T \eta \Lambda = \eta$$

Exercise 34. Show that $\det\Lambda = \pm 1$.

Such transformations can be divided into four types, depending on the sign of Λ_{00} and of $\det\Lambda$. If $\Lambda_{00} < 0$ the Lorentz transformation also involves a reversal of time (called T). If $\det\Lambda = -1$ it involves a reflection of space and time (called PT). It turns out there are subtle effects in particle physics (weak interactions) that are not invariant under T or PT. The subset of Lorentz transformations ("proper Lorentz transformations") that have $\Lambda_{00} > 0, \det\Lambda = 1$ are true symmetries of nature.

It should not be surprising that the scalar wave equation and the Klein-Gordon equation (see below) are invariant under all Lorentz transformations; proper or

not. Later we will see wave equations for spinors that are only invariant under the proper Lorentz transformations. They describe neutrinos, for example.

10.1.1. *Index Notation*

It will be convenient to denote a 4-vector as p_μ with the subscript ranging over $0, 1, 2, 3$. The Minkowski inner product can be written as

$$p_0^2 - p_1^2 - p_2^2 - p_3^2 = \sum_{\mu\nu} p_\mu p_\nu \eta^{\mu\nu}$$

where $\eta^{00} = 1, \eta^{11} = \eta^{22} = \eta^{33} = -1$, all other components being zero. We can abbreviate it further as

$$p_0^2 - p_1^2 - p_2^2 - p_3^2 = p_\mu p_\nu \eta^{\mu\nu}$$

by dropping the summation symbol: Any index that is repeated will be summed over.

$$p_0 q_0 - p_1 q_1 - p_2 q_2 - p_3 q_3 = p_\mu q_\nu \eta^{\mu\nu}$$

Part of the deal is that an index can appear no more than twice. For example

$$(p_0 q_0 - p_1 q_1 - p_2 q_2 - p_3 q_3)^2 = p_\mu q_\nu \eta^{\mu\nu} p_\rho q_\sigma \eta^{\rho\sigma}$$

and not $p_\mu q_\nu \eta^{\mu\nu} p_\mu q_\nu \eta^{\mu\nu}$.

10.2. Lorentz Group and Its Lie Algebra

10.2.1. *The Lorentz Group consists of matrices that satisfy the condition* $\Lambda\eta\Lambda^T = \eta$

Here $\eta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$ is the Minkowski metric.

10.2.2. *The determinant of such a matrix is* ± 1

The set of matrices of determinant 1 forms subgroup $SO(1,3)$. Under continuous deformations, the sign of the determinant cannot change. There is a further sign that cannot change under such deformations:

10.2.3. *The set of matrices of determinant one and* $\Lambda_{00} > 0$ *is a subgroup* $SO_+(1,3)$

The $SO_+(1,3)$ is the subgroup that does not invert either time or space. Thus, the Lorentz group breaks up into four connected subsets, according to the signs of Λ_{00} and $\det\Lambda$. Of these, $SO_+(1,3)$ is a connected subgroup.

10.2.4. *Surprisingly, only* $SO_+(1,3)$ *is an exact symmetry of nature*

Space reversal (parity) is broken by weak interactions. The combination of space and time reversal $SO(1,3)$ is broken by a subtle phase in the quark mass matrix (Kobayashi-Maskawa). So, only the connected components of the Lorentz Lie group (the part determined by its Lie algebra) is an exact symmetry.

10.2.5. *Infinitesimal Lorentz transformations satisfy* $M\eta + \eta M^T = 0$

Here $\Lambda = 1 + \epsilon M$ where ϵ is an infinitesimally small quantity. Equivalently, ηM is anti-symmetric. There are $\binom{4}{2} = 6$ independent solutions for this condition. They form a six dimensional Lie algebra under commutation. It includes the Lie algebra of rotations as a sub-algebra. The remaining generators mix time and space (e.g., "boosts").

10.3. The Variational Principle for the Wave Equation

Let us start with the simplest case

10.3.1. *The* $1+1$ *Dimensional wave equation* $\ddot{\phi} - \phi'' = 0$ *is the condition for* $S[\phi] = \frac{1}{2}\int \left[\dot{\phi}^2 - \phi'^2\right] dtdx$ *to be an extremum*

Consider the class of functions satisfying the initial and final conditions

$$\phi(t_1, x) = q_1(x), \quad \phi(t_2, x) = q_2(x)$$

It is convenient also to put a cut-off on space ("box" boundary conditions)

$$\phi(t_1, -L) = 0 = \phi(t_1, L)$$

Usually we are interested in the case of unbounded space. The mathematically correct thing to do is to first consider the finite box, then take the limit. We won't bother to do that usually. But let us do it in this simple context to see how it is done.

We will consider variations that preserve these conditions. That is

$$\phi_\epsilon(t,x) = \phi(t,x) + \epsilon\xi(t,x)$$

with

$$\xi(t_1,x) = 0 = \xi(t_2,x)$$

We must also require the variations at the spatial boundary:

$$\xi(t_1,-L) = 0 = \xi(t_1,L)$$

Then

$$S[\phi_\epsilon] = S[\phi] + \epsilon\int_{t_1}^{t_2} dt\int_{-L}^{L}\left[\dot\phi\dot\xi - \phi'\xi'\right]dx + \frac{1}{2}\epsilon^2\int\left[\dot\xi^2 - \xi'^2\right]dtdx$$

By integration by parts (the boundary term vanishes because of the b.c. above) in space

$$\int_{-L}^{L}\phi'\xi' dx = [\phi'\xi]_{x=-L}^{x=L} - \int_{-\infty}^{\infty}\phi''\xi dx$$

and in time (we use the condition on ξ at initial and final times)

$$\int_{t_1}^{t_2} dt\dot\phi\dot\xi = \left[\dot\phi\xi\right]_{t=t_1}^{t=t_2} - \int_{t_1}^{t_2} dt\ddot\phi\xi$$

Thus the first order variation can be written as

$$\delta S = \epsilon\int_{t_1}^{t_2} dt\int_{-L}^{L}\left[\dot\phi\dot\xi - \phi'\xi'\right]dx = -\epsilon\int_{t_1}^{t_2} dt\int_{-L}^{L}\left[\ddot\phi - \phi''\right]\xi dx$$

At an extremum this must vanish for all ξ. That is possible precisely when the wave equation is satisfied:

$$\ddot\phi - \phi'' = 0.$$

You might wonder why we kept the second order term in S as we don't need it to prove the wave equation. But it helps us to understand the nature of the extremum.

10.3.2. *We can cast the action in a Lorentz invariant form*

$$S[\phi] = \frac{1}{2}\int\eta^{\mu\nu}\partial_\mu\phi\partial_\nu\phi dtdx$$

Its variation is

$$S[\phi_\epsilon] = S[\phi] + \epsilon\int_{t_1}^{t_2} dt\int_{-L}^{L}\eta^{\mu\nu}\partial_\mu\phi\partial_\nu\xi dx + \frac{1}{2}\epsilon^2\int\eta^{\mu\nu}\partial_\mu\xi\partial_\nu\xi dtdx$$

The integration by parts we did is based on the identity

$$\eta^{\mu\nu}\partial_\mu\phi\partial_\nu\xi = \partial_\nu\left[\eta^{\mu\nu}\partial_\mu\phi\xi\right] - \left[\eta^{\mu\nu}\partial_\mu\partial_\nu\phi\right]\xi$$

and the following theorem

10.3.3. ***Gauss' theorem: The integral of a divergence over a region Ω can be written as a surface integral on its boundary $\partial\Omega$***

$$\int_\Omega \partial_\mu j^\mu dx = \int_{\partial\Omega} j^\mu d\sigma_\mu$$

You have seen proofs of this in two and three dimensions. It holds in all dimensions. Here $d\sigma_\mu$is the area element on the boundary. It should be thought of as a vector pointing along the outward normal to the boundary.

In the above example, $\partial\Omega$ consists of four pieces:

$$t = t_1, \quad t = t_2, \quad x = \pm L$$

which works out to

$$-\int_{-L}^{L} j^0(t_1,x)dx + \int_{-L}^{L} j^0(t_2,x)dx - \int_{t_1}^{t_1} j^1(t,-L)dt + \int_{t_1}^{t_2} j^1(t,L)dx$$

The signs arise because the outward normal points backward in time (the first term) and to the left (in the third term).

In that case each of these pieces j^μ was zero. But later we will need the general case of Gauss' theorem.

10.3.4. ***Another important application of Gauss' theorem is to conservation laws***

If $\partial_\mu j^\mu = 0$, the surface integral of j on the boundary of any domain is zero. Applied to the above region it says that

$$\int_{-L}^{L} j^0(t_1,x)dx = \int_{-L}^{L} j^0(t_2,x)dx - \int_{t_1}^{t_1} j^1(t,-L)dt + \int_{t_1}^{t_2} j^1(t,L)dx$$

If we assume that the flux j^1 tends to zero as $L \to \infty$ we get

$$\int_{-\infty}^{\infty} j^0(t_1,x)dx = \int_{-\infty}^{\infty} j^0(t_2,x)dx$$

Think of j^0 as a charge density. The total initial charge is equal to the final charge. Thus, $\partial_\mu j^\mu = 0$ is the differential version of a conservation law.

10.3.5. ***Now we see that the wave equation in any dimension follows from the variational principle*** $S[\phi] = \frac{1}{2}\int \eta^{\mu\nu}\partial_\mu\phi\partial_\nu\phi dx$

Here $dx = dx^0 dx^1, \ldots$. The Lorentz invariant version of the argument carries over with no change to higher dimensions. Read it again with $\mu, \nu = 0, 1, 2, 3$. You may strain our geometric imagination a bit but the equations are the same.

10.4. The Klein-Gordon Equation

10.4.1. ***The set of four-momenta of a particle with mass m satisfy***

$$p_0^2 - c^2 p_1^2 - c^2 p_2^2 - c^2 p_3^2 = m^2 c^4,\ p_0 > 0$$

10.4.2. ***If we ignore the condition that the energy has to be positive, we can express this as a simple differential equation for its wave-function***

$$-\hbar^2\left[\frac{\partial^2}{\partial t^2} - c^2\frac{\partial^2}{\partial x_1^2} - c^2\frac{\partial^2}{\partial x_2^2} - c^2\frac{\partial^2}{\partial x_3^2}\right]\phi = m^2c^4\phi$$

Recall that $p_0 = i\hbar\frac{\partial}{\partial t}$, $p_1 = -i\hbar\frac{\partial}{\partial x^1}$ etc. in quantum mechanics.

10.4.3. ***In Lorentz invariant notation*** $\eta^{\mu\nu}\partial_\mu\partial_\nu\phi + \mu^2\phi = 0$ ***where*** $\mu = \frac{mc}{\hbar}$

Note that μ has dimensions of length^{-1}. The static (time-independent) solutions satisfy Yukawa's equation.

10.4.4. ***The equation is invariant under Lorentz transformations-including parity and time reversal***

The equation allows for negative energy solutions. To properly interpret this situation, we need quantum field theory. We will return to this topic later.

10.4.5. ***If*** ϕ ***is complex-valued, the Klein-Gordon equation implies a conservation law***

$$j_\mu = 2\mathrm{Im}\phi^*\partial_\mu\phi$$

The point is that

$$\eta^{\mu\nu}\partial_\nu\left[\phi^*\partial_\mu\phi\right] = \eta^{\mu\nu}\partial_\nu\phi^*\partial_\mu\phi + \phi^*\eta^{\mu\nu}\partial_\nu\partial_\mu\phi$$

The second term is zero because of the equation of motion:

$$\eta^{\mu\nu}\partial_\nu\left[\phi^*\partial_\mu\phi\right] = \eta^{\mu\nu}\partial_\nu\phi^*\partial_\mu\phi$$

The quantity of the r.h.s. is real (complex conjugation interchanges the two factors up to a switching of indices $\mu \rightleftarrows \nu$). So,

$$\partial^\mu j_\mu = 2\mathrm{Im}\partial^\mu\left[\phi^*\partial_\mu\phi\right] = 0.$$

10.5. Noether's Theorem

10.5.1. *This conservation law can also be understood as a consequence of a symmetry of the action under $\phi \to e^{i\alpha}\phi$*

$$S = \int \left[\eta^{\mu\nu}\partial_\nu\phi^*\partial_\mu\phi + \mu^2\phi^*\phi\right] dx$$

Varying w.r.t. ϕ^* gives the equation of motion. The action is clearly invariant under the above transformation when α is a constant. To derive the conservation law, we use a deep idea of Noether. Consider the change of S under infinitesimal changes $\delta\phi(x) = i\epsilon\alpha(x)\phi(x)$ where $\alpha(x)$ can depend on x. We already know that S is invariant under such changes when α is a constant. So, δS must be of the form

$$\delta S = \int \partial_\mu\alpha j^\mu dx$$

for some j^μ which depends quadratically on ϕ and involves one derivative of ϕ. By a straightforward calculation we can see that it is

$$j^\mu = -i\partial^\mu\phi^*\phi + i\phi^*\partial^\mu\phi$$

which is just the j^μ we defined earlier.

To proceed further we need two facts:

- $\delta S = \int \partial_\mu\left[\alpha j^\mu\right] dx - \int \alpha\partial_\mu j^\mu dx$. The first term is a surface integral by Gauss's theorem and hence can be set to zero (the fields and hence j^μ must vanish at infinity).
- When the equation of motion is satisfied, $\delta S = 0$ for *any* variation.

Together they imply that

$$\partial_\mu j^\mu = 0$$

which is the conservation law.

10.5.2. *The power of Noether's theorem is that it holds for non-linear equations as well: Any symmetry implies a conservation law*

For example, the action

$$S = \int \left[\eta^{\mu\nu} \partial_\nu \phi^* \partial_\mu \phi + \frac{\lambda}{2} \left(\phi^* \phi - a^2 \right)^2 \right] dx$$

also has the symmetry under $\phi \mapsto e^{i\alpha}\phi$. The same argument as above implies again the conservation of

$$j^\mu = -i\partial^\mu \phi^* \phi + i\phi^* \partial^\mu \phi$$

But this action leads to a more complicated non-linear equation of motion,

$$\eta^{\mu\nu} \partial_\nu \partial_\mu \phi + \lambda(\phi^* \phi - a^2)\phi = 0$$

A version of this occurs is the standard model, in connection with the Higgs boson.

10.5.3. *Noether's theorem also applies to non-abelian symmetries: Every symmetry of the action under a Lie group implies the conservation of a current valued in its Lie algebra*

For example, let $\phi : \mathbb{R}^{1,3} \to \mathbb{R}^n$ be a scalar field as far as Lorentz transformations are concerned; but it transforms as a vector under some "internal" $O(n)$ symmetry. An action of the form

$$S = \int \left[\eta^{\mu\nu} \partial_\nu \phi_a \partial_\mu \phi_a + V(|\phi|^2) \right] dx, \quad |\phi|^2 = \phi_a \phi_a$$

is invariant under $O(n)$. This leads to the conservation of a current

$$j^\mu_{ab} = \partial^\mu \phi_a \phi_b - \phi_a \partial_\mu \phi_b$$

The argument is the same as above, applied to $\delta\phi_a(x) = \epsilon \alpha_{ac}(x)\phi_c(x)$ where $\alpha_{ab}(x)$ is anti-symmetric in ab. Note that the current is an anti-symmetric matrix for each μ (in the indices ab): It is valued in the Lie algebra $o(n)$.

In fact the earlier case of a complex scalar field is just the particular case $n = 2$.

10.6. Fermionic Wave Equations

10.6.1. *The Weyl Equation*

A scalar field represents a particle with spin zero. This is partly why the Klein–Gordon equation cannot describe the electron. Even when it is not moving, the electron carries an angular momentum of $\frac{\hbar}{2}$. Before we discuss the Dirac equation, let us look at an even more basic equation, the Weyl equation for a *massless* spin half particle. After that we will see how to add mass to it. There are two ways of doing this, the so called "Majorana mass" and the "Dirac mass". The electron mass is of Dirac type, but the neutrinos (at least two kinds of neutrinos are massive; we don't know for sure yet which two) may well have a Majorana mass.

To understand this, note that the length^2 of a four-vector can be written as a determinant:

$$p_0^2 - p_1^2 - p_2^2 - p_3^2 = \det \begin{pmatrix} p_0 + p_3 & p_1 - ip_2 \\ p_1 + ip_2 & p_0 - p_3 \end{pmatrix}$$

In terms of Pauli matrices

$$\sigma^0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma^1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

The matrix above is hermitian and can be written as a linear combination of Pauli matrices:

$$\hat{p} = \begin{pmatrix} p_0 + p_3 & p_1 - ip_2 \\ p_1 + ip_2 & p_0 - p_3 \end{pmatrix} = p_0\sigma^0 + p_1\sigma^1 + p_2\sigma^2 + p_3\sigma^3 \equiv \sigma^\mu p_\mu.$$

Conversely, every 2×2 hermitian matrix can be thought of as a four-vector. (The number of independent components of a 2×2 hermitian matrix is four.)

If p is the momentum of a massless particle it would be a null vector; i.e.,

$$\det \hat{p} = 0.$$

In this case there will be a spinor $\tilde{u}(p)$such that

$$\hat{p}\tilde{u}(p) = 0.$$

If we pass to the Fourier transform

$$u(x) = \int \tilde{u}(p) e^{ip\cdot x} \frac{dp}{(2\pi)^4}$$

this becomes a differential equation

$$\sigma^\mu \frac{\partial u}{\partial x^\mu} = 0.$$

This is the Weyl equation: an analogue of the wave equation for a massless spin half particle. Itis the most elementary of wave equations.

10.6.2. *The Weyl Equation implies the wave equation*

This is obvious in the momentum picture. If there is a non-zero spinor satisfying $\hat{p}\tilde{u}(p) = 0$, we must have $\det p = 0$ which implies the wave equation. We can derive it another way. Define the "parity" conjugate Pauli matrices

$$\check{\sigma}^{\mu} = (1, -\sigma^1, -\sigma^2, -\sigma^3)$$

Then

$$\check{\sigma}^{\mu}\sigma^{\nu} + \check{\sigma}^{\nu}\sigma^{\mu} = 2\eta^{\mu\nu}$$

Equivalently

$$\check{\sigma}^{\mu}\sigma^{\nu}p_{\mu}p_{\nu} = \eta^{\mu\nu}p_{\mu}p_{\nu}$$

We can apply the operator $\check{\sigma}^{\mu}\frac{\partial}{\partial x^{\mu}}$ to the Weyl equation to derive the wave equation

$$\sigma^{\nu}\frac{\partial u}{\partial x^{\nu}} = 0 \implies \check{\sigma}^{\mu}\frac{\partial}{\partial x^{\mu}}\sigma^{\nu}\frac{\partial u}{\partial x^{\nu}} = \check{\sigma}^{\mu}\sigma^{\nu}\frac{\partial}{\partial x^{\mu}}\frac{\partial u}{\partial x^{\nu}} = \eta^{\mu\nu}\frac{\partial^2 u}{\partial x^{\mu}\partial x^{\nu}} = 0$$

10.6.2.1. *Lorentz invariance transformation of spinors*

A Lorentz transformation must take $\hat{p}$ to another hermitian matrix. This suggests that there is a 2×2 complex matrix corresponding to every Lorentz transformation such that

$$\widehat{\Lambda p} = \lambda\hat{p}\lambda^{\dagger}$$

The r.h.s. is hermitian for any λ. The condition that the length be unchanged becomes $\det\hat{p} = \det\left[\lambda\hat{p}\lambda^{\dagger}\right]$; i.e.,

$$|\det\lambda|^2 = 1.$$

Now, if you change λ by multiplying it by a complex number of magnitude one, $\lambda\hat{p}\lambda^{\dagger}$ is unchanged. We can use this phase freedom to choose

$$\det\lambda = 1.$$

So, we suspect that to every Lorentz transformation Λ, there is a 2×2 complex matrix of determinant one such that

$$\widehat{\Lambda p} = \lambda\hat{p}\lambda^{\dagger} \tag{10.6.1}$$

for all four-vectors. This is true, but with some important caveats.

- Both λ and $-\lambda$ give the same Λ. Thus we have a 2 to 1 map, λ is much like a "matrix square root" of Λ.
- Λ must be a proper Lorentz transformation. That is, $\Lambda_{00} > 0$ and $\det\Lambda = 1$. This subgroup of of Lorentz transformations is called $SO_+(1,3)$. To see the first condition consider the special case $p = (1,0,0,0)$. Then $\left(\widehat{\Lambda p}\right)_{00} = \Lambda_{00} = |\lambda_{00}|^2$. Similarly, we can show that $\det\Lambda > 0$ as well if it arises from a λ as above. So, we cannot find such a λ for parity (space reflection) or time reversal: For the first case $\det\Lambda < 0$ and for the second $\Lambda_{00} < 0$.

The precise statement is,

Theorem 35. *There is a homomorphism $\Lambda : SL_2(\mathbb{C}) \to SO_+(1,3)$ such that*

$$\Lambda^\mu{}_\nu(\lambda)\sigma^\nu = \lambda\sigma^\mu\lambda^\dagger.$$

The kernel of this homomorpism is $Z_2 = \{1,-1\} \subset SL_2(\mathbb{C})$.

Explicitly, for $\lambda = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$,

$$\Lambda(\lambda) = \begin{pmatrix} \frac{1}{2}\left(|a|^2+|b|^2+|c|^2+|d|^2\right) & \frac{1}{2}\left(ba^*+ab^*+dc^*+cd^*\right) & \frac{1}{2}i\left(ba^*-ab^*+dc^*-cd^*\right) & \frac{1}{2}\left(|a|^2+|c|^2-bb^*-dd^*\right) \\ \frac{1}{2}\left(ca^*+ac^*+db^*+bd^*\right) & \frac{1}{2}\left(da^*+ad^*+cb^*+bc^*\right) & \frac{1}{2}i\left(da^*-ad^*+cb^*-bc^*\right) & \frac{1}{2}\left(ca^*+ac^*-db^*-bd^*\right) \\ -\frac{1}{2}i\left(ca^*-ac^*+db^*-bd^*\right) & -\frac{1}{2}i\left(da^*-ad^*+cb^*-bc^*\right) & \frac{1}{2}\left(da^*+ad^*-cb^*-bc^*\right) & -\frac{1}{2}i\left(ca^*-ac^*-db^*+bd^*\right) \\ \frac{1}{2}\left(|a|^2+|b|^2-cc^*-dd^*\right) & \frac{1}{2}\left(ba^*+ab^*-dc^*-cd^*\right) & \frac{1}{2}i\left(ba^*-ab^*-dc^*+cd^*\right) & \frac{1}{2}\left(|a|^2+|d|^2-bb^*-cc^*\right) \end{pmatrix}$$

This is the analogue of the homomorphism $R : SU(2) \to SO(3)$ we found in Sec.4.4. Indeed if we restrict to the subgroup $SU(2) \subset SL_2(\mathbb{C})$ it reduces to that case.

Exercise 36. Prove the above theorem.

Now if the spinor u transforms as

$$u \mapsto \lambda^{\dagger-1}u$$

the Weyl equation is invariant under proper Lorentz transformations $SO_+(1,3)$. It is not invariant under Parity, as a λ does not exist in this case, as noted above.

Pauli actually discovered the Weyl equation first. But he rejected it as unphysical, because at that time Parity was believed to be an exact symmetry of nature. But once physicists discovered that Parity was violated in reactions that involve neutrinos, it became natural to use this equation to describe a neutrino.

But this story is full of twists and turns. Now, we know that at least some kinds of neutrinos are massive. So they cannot be described by the Weyl equation any more.

How would we modify the Weyl equation so that the particle has a mass? That is, so that the condition for a solution is $\det \hat{p} = m^2$? There are two kind of masses: Majorana mass and the Dirac mass. The latter requires that we double the number degrees of freedom but has the added benefit that the equation becomes parity invariant.

10.6.3. *The Majorana Equation*

Let us write $\lambda = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, $ad - bc = 1$. Then $\lambda^{-1} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$. Now,

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} -b & a \\ -d & c \end{pmatrix} = \begin{pmatrix} d & -c \\ -b & a \end{pmatrix} = \lambda^{-1T}$$

so that matrices of determinant one satisfy

$$\lambda = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \lambda^{-1T} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Recall that the set of matrices of determinant one form a group, called $SL_2(\mathbb{C})$.

Now, $u \mapsto \lambda^{\dagger -1} u$ implies that

$$u^* \to \lambda^{-1T} u^*$$

and

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} u^* \mapsto \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \lambda^{-1T} u^*$$

$$= \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \lambda^{-1T} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} u^* = \lambda \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} u^*$$

That is, the combination $\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right) u^*$ transforms the same way as $i\sigma^\mu \partial_\mu u$:

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} u^* \mapsto \lambda \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} u^*, \quad i\sigma^\mu \partial_\mu u \mapsto \lambda i \sigma^\mu \partial_\mu u$$

Thus the equation

$$i\sigma^\mu \partial_\mu u + m \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} u^* = 0$$

is Lorentz invariant. Because of the presence of the complex conjugate in the second term, this equation is no longer complex linear (i.e., u being a solution need not mean iu is a solution.) It is better to think of as a system of equation for four real variables rather than two complex variables. This is the way Majorana thought of it originally.

This equation violates parity, as can be verified by writing it out explicitly in terms of spatial derivatives. There isn't even any way of implementing parity as a transformation on u.

10.6.4. *The Dirac Equation*

But there must be a way to give the fermion a mass without violating parity. The electron is massive and its most important interaction (electromagnetic) is parity preserving. The idea is to double the number of degrees of freedom, so that parity interchanges them. That is, we introduce two independent spinor fields u and v which transform as

$$u \mapsto \lambda^{\dagger -1} u, \quad v \mapsto \lambda v$$

Then

$$i\sigma^\mu \partial_\mu u + mv = 0$$

is invariant, as we saw earlier. We need an additional equation involving derivatives of v.

Define

$$\check{\sigma}^0 = \sigma^0, \quad \check{\sigma}^i = -\sigma^i, \quad i = 1, 2, 3$$

so that $\check{\sigma}^\mu$ is the Parity transform of σ^μ. Then we see that

$$i\check{\sigma}^\mu \partial_\mu v + mu = 0$$

is also Lorentz invariant.

Exercise 37. Show that $\Lambda^\mu_{\ \nu}(\lambda)\check{\sigma}^\nu = \lambda^{\dagger -1}\check{\sigma}^\mu \lambda^{-1}$

We can combine the two equations for u, v into a four component equation:

$$\psi = \begin{pmatrix} u \\ v \end{pmatrix}$$

to get the Dirac equation

$$i\gamma^\mu \partial_\mu \psi + m\psi = 0.$$

Here, the Dirac matrices are

$$\gamma^\mu = \begin{pmatrix} 0 & \check{\sigma}^\mu \\ \sigma^\mu & 0 \end{pmatrix}.$$

They satisfy the "Clifford algebra"

$$\gamma^\mu \gamma^\nu + \gamma^\nu \gamma^\mu = 2\eta^{\mu\nu}$$

There are other representations for the Dirac matrices which differ from the above by linear equivalence transformations (changes of basis). In calculations, it is best not to rely on any choice of basis, and use only properties that follow directly from the Clifford algebra.

Exercise 38. Show that the Dirac equation implies the Klein–Gordon equation for each component of ψ. This verifies the claim that it describes a massive free particle. Also, show that the Dirac equation is Parity invariant, if we let u and v interchanged under Parity.

Having doubled the number of degrees of freedom, you should expect that the equation describes twice as many particles as the electron itself. This is true: It also predicts the existence of an anti-particle with the same mass but the opposite (positive) charge. To understand this, we have to couple the Dirac equation to the electromagnetic field (see later) and also "second quantize" it [15].

10.6.5. *The Feuter Equation*

The analytic continuation of the scalar wave equation to Euclidean space (imaginary time) is the Laplace equation. What is the analytic continuation of the Weyl equation? We would replace the Pauli matrices by

$$\tilde{\sigma}^4 = \sigma^0 = 1, \quad \tilde{\sigma}^1 = \sqrt{-1}\sigma^1, \quad \tilde{\sigma}^2 = \sqrt{-1}\sigma^2, \quad \tilde{\sigma^3} = \sqrt{-1}\sigma^3$$

and

$$\frac{\partial \chi}{\partial x^4} + \tilde{\sigma}^1 \frac{\partial \chi}{\partial x^1} + \tilde{\sigma}^2 \frac{\partial \chi}{\partial x^2} + \tilde{\sigma}^3 \frac{\partial \chi}{\partial x^3} = 0$$

This has an interesting mathematical meaning. If we denote

$$i = \tilde{\sigma}^1, \quad j = \tilde{\sigma}^2, \quad k = \tilde{\sigma}^3$$

the algebraic relations satisfied by the "Eulcidean Pauli matrices" becomes

$$i^2 = -1 = j^2 = k^2$$

$$ij = k = -ji, \quad jk = i = -kj, \quad ki = j = -ik$$

These are the relations satisfied by quaternions: A generalization of complex numbers to higher dimensions discovered by Hamilton. Then the "Eucidean Weyl equation" is the quaternionic analogue of the Cauchy-Riemann equations (quaternionic analyticity).

$$\frac{\partial \chi}{\partial x^4} + i\frac{\partial \chi}{\partial x^1} + j\frac{\partial \chi}{\partial x^2} + k\frac{\partial \chi}{\partial x^3} = 0.$$

Independently of Dirac, Feuter discovered this equation in this context. The theory of quaternionic analytic functions is much more complicated than complex analytic functions, because quaternions do not commute. Still, it has been worked out in some detail. Most physicists just use Pauli matrices instead of thinking in terms of quaternions.

10.7. Variational Principle for Fermionic Wave Equations

Recall that the wave equation for spin zero particles can be deduced from a variational principle. There are similar principles for fermions as well.

Let us begin with the Dirac equation. Note that $\gamma^\mu$is hermitian for $\mu = 0$ and anti-Hermitian for $\mu = 1, 2, 3$. Yet, $\gamma^{\mu\dagger}$ satisfies the same anti-commutation relations as γ^μ. In fact

$$\gamma^\mu = \gamma^0 \gamma^{\mu\dagger} \gamma^0.$$

Define

$$\bar{\psi} = \psi^\dagger \gamma^0$$

We can use arguments similar to those in the section on the Dirac equation to see that

- $\bar{\psi}\psi$ transforms as a scalar
- $\bar{\psi}\gamma^\mu\psi$ transforms as a vector

when ψ transforms as a Dirac spinor. Thus

$$S_F = \int \bar{\psi} \left[i\gamma^\mu \partial_\mu + m \right] \psi \, dx$$

is Lorentz invariant. It serves as the variational principle: Varying w.r.t. $\bar{\psi}$ yields the Dirac equation. Similarly, for a Weyl spinor, $u^\dagger \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} u^*$ is a Lorentz scalar and $u^\dagger \sigma^\mu u$ is a Lorentz vector. Thus

$$S_W = \int u^\dagger \left[i\sigma^\mu \partial_\mu u + m \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} u^* \right] dx$$

is a variational principle for the Majorana equation.

10.8. Maxwell's Equations

The book by Jackson on *Classical Electrodynamics* has become a standard reference. The second volume of the series by Landau and Lifshitz *Classical Theory of Fields* shows greater physical insight.

10.8.1. *All magnetic fields must have zero divergence*

$$\nabla \cdot \mathbf{B} = 0$$

This means in particular that there is no analogue to an isolated electric charge in magnetism: A permanent magnet has to be a dipole. If you cut a dipole into two we will not get an isolated North pole and South pole. Instead we will get two dipoles again. Some theories that go beyond the standard model do allow for magnetic monopoles; but none have yet been observed.

10.8.2. *This equation can be solved by postulating that the magnetic field is a curl of a vector potential*

$$\mathbf{B} = \nabla \times \mathbf{A}$$

10.8.3. *Two vector potentials that differ only by the gradient of a scalar give the same magnetic field*

This is called a gauge transformation

$$\mathbf{A}' = \mathbf{A} + \nabla\Lambda, \quad \mathbf{B}' = \mathbf{B}$$

$$\nabla \times \nabla\Lambda = 0.$$

It turns out that invariance under this transfomation is a fundamental symmetry of nature. We will see that gauge transformations that generalize this are the fundamental symmetries of the standard model.

10.8.4. ***Another equation of Maxwell relates the time derivative of the magnetic field to the eletric field***

$$\frac{\partial \mathbf{B}}{\partial t} = -\boldsymbol{\nabla} \times \mathbf{E}$$

10.8.5. ***We can solve this by postulating in addition a scalar potential V***

$$\mathbf{E} = \frac{\partial \mathbf{A}}{\partial t} - \boldsymbol{\nabla} V$$

Remark 1. Recall that we are using units such that $c = 1$. Otherwise there will be some factors of $c$all over the place.

The gauge transformations must now change the scalar potential as well

$$V' = V + \frac{\partial \Lambda}{\partial t}$$

so that the electric field is unchanged.

$$\frac{\partial \boldsymbol{\nabla} \Lambda}{\partial t} = \boldsymbol{\nabla} \frac{\partial \Lambda}{\partial t}.$$

10.8.6. ***Under Lorentz transformations the scalar and vector potentials combine into a four-vector $A = (V, \mathbf{A})$.***

We will introduce an index $\mu = 0, 1, 2, 3$ such that

$$A_0 = V, \quad A = (A_0, A_1, A_2, A_3)$$

Then the gauge transformation can be written as

$$A'_\mu = A_\mu + \partial_\mu \Lambda$$

where ∂_μ denotes differentiation along the μ th direction. Gauge invariance is based on the identity

$$\partial_\mu \partial_\nu \Lambda = \partial_\nu \partial_\mu \Lambda.$$

The electric and magnetic fields are then

$$E_i = \partial_0 A_i - \partial_i A_0, \quad i = 1, 2, 3.$$

$$B_1 = \partial_2 A_3 - \partial_3 A_2, \quad B_2 = \partial_3 A_1 - \partial_1 A_3, \quad B_3 = \partial_1 A_2 - \partial_2 A_1$$

This suggests that we combine them into a single matrix $F_{\mu\nu}$

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$$

It is an anti-symmetric matrix:

$$F = \begin{pmatrix} 0 & E_1 & E_2 & E_3 \\ -E_1 & 0 & B_3 & -B_2 \\ -E_2 & -B_3 & 0 & B_1 \\ -E_3 & B_2 & -B_1 & 0 \end{pmatrix}$$

10.8.7. ***The remaining Maxwell's equations can be written in Lorentz invariant form as***

$$\partial^\mu F_{\mu\nu} = j_\nu$$

Expanded in terms of three-dimensional quantities

$$\frac{\partial \mathbf{E}}{\partial t} = -\boldsymbol{\nabla} \times \mathbf{B} + \mathbf{j}$$

$$\boldsymbol{\nabla} \cdot \mathbf{E} = j_0$$

The scalar j_0 is proportional to charge density and the vector $\mathbf{j}$ to current density.

10.8.8. ***The potential A satisfies a wave equation***

10.8.9. ***The electromagnetic field describes a particle of mass zero and spin one***

Mass zero because it travels at the velocity of light. (Duh. it is light.) Spin one because in three-dimensional language it includes a vector field, which has spin one.

10.9. Quantum Electrodynamics

So far we know the equations for the wave functions of spin $0, \frac{1}{2}$ and 1 particles. To understand the interactions of these particles with each other we must introduce non-linearities. The key is gauge invariance. A complete study of the resulting theory, quantum electrodynamics is well outside the scope of this course. Itzykson and Zuber *Introduction to Quantum Field Theory* is still a good reference. At a level closer to this course is the book by Kerson Huang, *Quarks and Leptons*.

Exercise. The Dirac equation implies the conservation of a current

$$j^\mu = \bar{\psi}\gamma^\mu\psi, \quad \bar{\psi} = (\chi^* \quad \phi^*)$$

That is,

$$\partial_\mu j^\mu = 0.$$

This implies that

$$\frac{\partial}{\partial t}\int j^0 d^3x = 0.$$

Thus we can think of $Q = e\int j^0 d^3x$ as the electric charge and $j^0, \mathbf{j}$ as the charge and current densities respectively. The constant e is the electric charge of the electron (or whatever other particle to which we will apply this equation).Thus

10.9.1. ***The Maxwell's equations in the presence of electrons is***

$$\partial^\mu F_{\mu\nu} = e\bar{\psi}\gamma_\mu\psi. \tag{10.9.1}$$

Just as electrons create electric and magnetic fields, these fields must affect their motion. The change in the Dirac equation due to the presence of electric and magnetic fields is more subtle. Gauge invariance is the key to understanding this. Recall that under gauge transformation

$$A'_\mu = A_\mu + \partial_\mu\Lambda$$

where Λ is an arbitrary function. We want to preserve this symmetry when we introuduce A_μ into the Dirac equation. We must transform ψas well so that the changes in ψ and $A_\mu$compensate for each other. Notice that if

$$\psi' = e^{ie\Lambda}\psi$$

$$\partial_\mu\psi' = e^{ie\Lambda}\left[\partial_\mu\psi + \left(ie\partial_\mu\Lambda\right)\psi\right]$$

Remark 40. Sensible people can handle the double use of the symbol e here. The e in the exponent is the electric charge and that below is the base of natural logarithms. Their values of course, have nothing to do with each other.

Thus in the combination below the derivatives of Λ cancel out:

$$\left[\partial_\mu - ieA'_\mu\right]\psi' = e^{ie\Lambda}\left[\partial_\mu - ieA_\mu\right]\psi$$

10.9.2. ***The Dirac equation in the presence of an electromagnetic field is***

$$\gamma^\mu\left[\partial_\mu - ieA_\mu\right]\psi = im\psi \tag{10.9.2}$$

Under a gauge transformation both sides are multiplied by the same factor, so it cancels out. The pair of equations (10.9.1,10.9.2) describe Quantum Electro Dynamics (QED) of charged spin one half particles and photons.

10.9.3. ***The equation of a charged massive spin zero particle is***

$$\eta^{\mu\nu}\left[\partial_\mu - ieA_\mu\right]\left[\partial_\nu - ieA_\nu\right]\phi = -m^2\phi$$

This also follows using gauge invariance. Of course, here ϕ is a scalar not a spinor.

10.9.4. ***The proper interpretation of the equations of Quantum Electrodynamics involves renormalization***

The trouble is that the equations as described above lead to infinities when quantum effects are fully included. They have to be removed by a strange set of rules called "renormalization". These rules work remarkably well and agree with experiments to high precision: Fifteen decimal point accuracy is the best science has ever achieved. Yet the correct mathematical formulation is still not clear. Dirac himself was very unsatisfied by this situation. New ideas in analysis are needed. But that is another story.

10.10. Lagrangian Formalism

10.10.1. ***Hamilton's Variation Principle gives a concise formulation of equations of motion***

Define the Lagrangian L to be some function of position and velocity; and action to be its integral:

$$S = \int L(q, \dot{q})dt$$

The condition that the action be stationary w.r.t. to small changes in q leads to the condition

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}} - \frac{\partial L}{\partial q} = 0$$

With the choice

$$L = \frac{1}{2}m\dot{q}^2 - V(q)$$

this gives the Newtonian equations of motion

$$m\ddot{q} = -\frac{\partial V}{\partial q}.$$

10.10.2. ***In a relativistic theory the dynamical quantities are fields: Functions of space and time***

10.10.3. ***The Lagrangian depends on the fields and their derivatives***

The Lagrangian is a Lorentz scalar.

10.10.4. ***The action is the integral of the Lagrangian over space and time***

$$S = \int L(\phi, \partial\phi) d^4x$$

$$\partial_\mu \left[\frac{\partial L}{\partial \left(\partial_\mu \phi\right)} \right] = \frac{\partial L}{\partial \phi}$$

The Lagrangian of a free massive scalar field is

$$L = \frac{1}{2}\eta^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - \frac{1}{2}m^2\phi^2$$

leading to the Klein–Gordon equation

$$\partial_\mu\partial^\mu\phi + m^2\phi = 0$$

More generally, an interacting scalar theory will have a lagrangian that has terms higher degree than two:

$$L = \tfrac{1}{2}\eta^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - V(\phi)$$

$$\partial_\mu\partial^\mu\phi + \tfrac{\partial V}{\partial\phi} = 0$$

For the Higgs field of the standard model (a complex doublet)

$$L = \eta^{\mu\nu}\partial_\mu\phi^\dagger\partial_\nu\phi - V(\phi), \quad V(\phi) = \frac{\lambda}{2}\left[\phi^\dagger\phi - v^2\right]^2$$

We can see directly that the ground states are on the sphere

$$\phi^\dagger\phi = v^2$$

10.10.5. ***The Lagrangian of Maxwell's theory is***

$$L = \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + j^\mu A_\mu, \quad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$$

leading to the equation

$$\partial^{\mu} F_{\mu\nu} = j_{\nu}$$

10.10.6. ***The Lagrangian of Dirac field is***

$$L = \bar{\psi}\left[i\gamma^{\mu}\partial_{\mu} + m\right]\psi$$

10.10.7. ***To get interacting theories we add the free lagrangians plus terms that depend on several fields***

For the Yukawa theory,

$$L = \bar{\psi}\left[i\gamma^{\mu}\partial_{\mu} + g\phi\right]\psi + \eta^{\mu\nu}\partial_{\mu}\phi^{\dagger}\partial_{\nu}\phi - V(\phi)$$

For QED

$$L = \bar{\psi} i\gamma^{\mu}\left[\partial_{\mu} + ieA_{\mu}\right]\psi + m\bar{\psi}\psi + \frac{1}{4}F_{\mu\nu}F^{\mu\nu}$$

For the Abelian Higgs Model

$$L = \eta^{\mu\nu}\left[\nabla_{\mu}\phi\right]^{*}\nabla_{\nu}\phi - V(\phi), \quad \nabla_{\mu}\phi = \partial_{\mu} - ie\phi \quad V(\phi) = \frac{\lambda}{2}\left[\phi^{*}\phi - v^{2}\right]^{2}$$

10.11. Yang–Mills Theory

10.11.1. ***Yang–Mills Theory is the foundation of the theory of elementary particles***

It describes the self-interaction of spin 1 particles: The photon, $Z, W^{\pm}$ and the gluons. The principle of gauge invariance also determines the interactions of these spin one particles with those of spin zero and spin $\frac{1}{2}$: The quarks and leptons.There is also a theory of interactions of spin zero particles (Higgs fields) and spin two particles (General Relativity).

10.11.2. ***Maxwell's theory of electromagnetism is invariant under an abelian gauge group***

Let $\Lambda : R^4 \longrightarrow R$ be a real valued function. Recall that under the gauge transformation

$$A_{\mu} \longrightarrow A_{\mu} + \partial_{\mu}\Lambda$$

the field strength

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$$

is unchanged. Thus, the Lagrangian

$$L = \frac{1}{4} F^{\mu\nu} F_{\mu\nu}$$

is invariant under both gauge and Lorentz transformations. Two successive gauge transformations is equivalent to one under the sum

$$\Lambda_1 + \Lambda_2$$

This is a commutative (abelian) group. Suppose a scalar field transforms as

$$\phi \to e^{i\Lambda} \phi$$

Then the covariant derivative

$$\nabla_\mu \phi = \partial_\mu \phi + i A_\mu \phi$$

transforms as

$$\nabla_\mu \phi \to e^{i\Lambda} \nabla_\mu \phi.$$

The Lagrangian

$$L = \frac{1}{4e^2} F^{\mu\nu} F_{\mu\nu} + \frac{1}{2} |\nabla \phi|^2 - V(|\phi|)$$

is gauge invariant. We saw a version of this in the discussion of the Higgs mechanism.

10.11.2.1. *The value of e determines the strength of the interaction*

We have chosen to define the gauge potential such that the coupling constant appears as a constant factor in the Lagrangian. For any e the gauge invariance holds. and is determined experimentally to be about a third. More precisely

$$\frac{e^2}{4\pi} \approx \frac{1}{137}.$$

10.11.2.2. *The commutator of covariant derivatives is just a multiplication by the field strength*

$$\nabla_\mu \nabla_\nu \phi - \nabla_\nu \nabla_\mu \phi = i F_{\mu\nu} \phi$$

This is similar to the definition of curvature in Riemannian geometry.

10.11.3. *In Yang–Mills theory, the gauge transformations are valued in a Lie group*

Let $g : R^4 \to G$ be a function from space-time into a Lie group. The cases of most physical interest are $G = SU(n)$ or $U(n)$. Suppose we have a scalar field transforming under some representation of this group. (Think of $G = U(n)$ and $\phi(x) \in C^n$.) Then

$$\phi \to g\phi$$

We can define a covariant derivative by analogy

$$\nabla_\mu \phi = \partial_\mu \phi + iA_\mu \phi$$

where A_μ is valued in (matrix representation of) the Lie algebra of G. For example, if $G = U(n)$, then each component of $A_\mu(x)$ is a hermitian matrix. How should A_μ transform in order that this covariant derivative transform as before?

$$\nabla_\mu \phi \to g\nabla_\mu \phi$$

A short calculation gives the answer

$$A_\mu \to gA_\mu g^{-1} + g\partial_\mu(g^{-1})$$

If $g = e^{ie\Lambda}$ this reduces to the transformation of Maxwell's theory. What then is the analogue of the field strength? We can calculate

$$\nabla_\mu \nabla_\nu \phi - \nabla_\nu \nabla_\mu \phi = iF_{\mu\nu}\phi$$

where

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + i[A_\mu, A_\nu]$$

The commutator term on the r.h.s. makes all the difference: It implies interactions among spin one particles that have no analogue in Maxwell's theory.

Under gauge transformations,

$$F_{\mu\nu} \to gF_{\mu\nu}g^{-1}.$$

10.11.3.1. *Recall that Compact Lie algebra is one that admits a positive invariant inner product*

That is, for non-zero elements of the Lie algebra

$$\langle u, u \rangle > 0$$

and under the adjoint action it is invariant:

$$\langle gug^{-1},\ gug^{-1}\rangle = \langle u, u\rangle.$$

On compact simple Lie algebras (e.g., $su(n)$) such an inner product is unique up to a scalar multiple. On $u(n) \approx su(n) \oplus u(1)$ there are two independent constants determining the general inner product. These constants are called coupling constants in the context of Yang-Mills theory

10.11.4. ***The Lagrangian of Yang-Mills theory is determined by a positive inner product on its Lie algebra***

$$L = \frac{1}{4}\langle F^{\mu\nu}, F_{\mu\nu}\rangle$$

10.11.5. ***Using covariant derivatives we can bring spin zero and spin one fields***

$$L = \frac{1}{4}\langle F^{\mu\nu}, F_{\mu\nu}\rangle + \frac{1}{2}|\nabla\phi|^2 + V(|\phi|)$$

$$L = \frac{1}{4}\langle F^{\mu\nu}, F_{\mu\nu}\rangle + \bar{\psi}[i\gamma^\mu\nabla_\mu + m]\psi$$

10.11.6. ***The Higgs Model with $U(2)$ invariance describes weak interactions***

$$L = \frac{1}{4}\langle F^{\mu\nu}, F_{\mu\nu}\rangle + \frac{1}{2}|\nabla\phi|^2 - V(|\phi|)$$

where $\phi = \begin{pmatrix}\phi_1\\ \phi_2\end{pmatrix}$ is a vector with two complex components. It is convenient to split the gauge field into a traceless 2×2 matrix L_μ and a multiple of the identity Y_μ since $u(2) \approx su(2) \oplus u(1)$

$$A_\mu = L_\mu + Y_\mu$$

$$L = \tfrac{1}{4e_1^2}\mathrm{tr}\, L^{\mu\nu}L_{\mu\nu} + \tfrac{1}{4e_2^2}Y^{\mu\nu}Y_{\mu\nu} + \tfrac{1}{2}|\nabla\phi|^2 - V(|\phi|)$$

where

$$\nabla_\mu\phi = \partial_\mu\phi + iL_\mu\phi + iqY_\mu\phi$$

The "hypercharge" q of the Higgs field and the coupling constants e_1, e_2 are experimentally determined. With the potential

$$V(\phi) = \frac{\lambda}{2}\left[\phi^\dagger\phi - v^2\right]^2$$

this describes a set of three massive particles $W^{\pm}, Z$ and a massless photon. (Higgs mechanism).

10.11.7. *Yang–Mills Theory with gauge group SU(3) is Quantum Chromodynamics, the theory of strong interactions*

$$L = \frac{1}{4\alpha}\langle F^{\mu\nu}, F_{\mu\nu}\rangle + \sum_{a=1}^{N_f} \bar{\psi}[i\gamma^{\mu}\nabla_{\mu} + m_a]\psi$$

Each quark field ψ_a is a three component vector under $SU(3)$in addition to being a Dirac spinor. There are six kinds of such quarks $a = 1, \ldots, 6$ corresponding to u, d, c, s, t, b with widely varying masses:

$$m_a \sim 5, 10, 1500, 250, 175000, 5000$$

in MeV. In most cases of interest in Nuclear Physics, only the lightest two or three quarks needs to be considered. For an excellent summary of the standard model see [10].

This page intentionally left blank

Chapter 11

RANDOM MATRICES

The main reference is the book "Random Matrices" by M. L. Mehta [21]. More mathematical developments are in the book by D. Gioev and P. Deift [23].

11.1. Sources of Random Matrix Theory

Random matrices arise in several disciplines of physics, probability theory and statistics.

11.1.1. *The Eigenvalues of a Matrix Whose Elements are Random Variables*

This is an idea, first pursued by Wigner in Nuclear Physics. The thousands of energy levels of nuclei defied any simple dynamical description. The hamiltonian was modeled as a hermitian or real symmetric matrix. Although the actual nuclear eigenvalues are not well described by this model, the level spacing (energy difference between two successive energy levels) fits with those of a random real symmetric matrix. (This happens when the spin-dependent part of the nuclear hamiltonian is important.) Part of the reason for the success is the universality of the spacing distribution: It is the same for a large class of random matrices.

This universality may be viewed as analogous to the central limit theorem, which says that the sum of a large number of random variables is Gaussian for more or less any collection of random variables. If the hamiltonian is a sum of a large number of terms of independent physical origin, each matrix element will tend to an independent gaussian random variable (the only relation between them comes from being hermitian or symmetric). Since there is no preferred basis,

the joint probability distribution must be invariant under unitary (or orthogonal) transformations.

11.1.2. *Covariance Matrix of Samples*

Imagine we make p measurements of some quantities $\xi_i, i = 1, \ldots, n$ (e.g., prices of stocks in the S&P 500, annual rain falls in the counties of NY state etc.) We can arrange these into an $n \times p$ rectangular matrix x_{ia}. By subtracting the mean values we get another vector:

$$y_{ia} = x_{ia} - \frac{1}{p}\sum_{b=1}^{p} x_{ib}$$

The covariance matrix of the data is

$$\Sigma_{ij} = \frac{1}{p}\sum_{b} y_{ib}y_{jb}$$

This is a positive[1] symmetric matrix of random variables. The eigenvector of the largest eigenvalue of this matrix is important in statistics: It captures most of the random variation (Principal Component Analysis). What is the distribution of this eigenvalue? This answer turns out to be universal (the same for a large class of random variables) and was found by Tracy and Widom. Such "extreme value distributions" are an active area of research in random matrix theory. But we won't go into that here.

11.1.3. *Operator algebras*

A different approach (due to Voiculescu), with surprising connections to Wigner's theory, started in the theory of operator algebras. The simplest example, is based on the Toeplitz algebra, defined by the relations

$$AA^\dagger = 1,$$

Note the similarity to the canonical commutation relations for creation-annihilation operators. Yet, an important difference is that it is the product $AA^\dagger$, not the commutator that is equal to one. The states on which the operators act can be obtained by acting with "creation operators" $A^\dagger$ on the "vacuum state" $|0\rangle$

[1] A matrix Σ is positive if $u^\dagger \Sigma u \geq 0$ for all vectors u. In our case this is the sum $\frac{1}{p}\sum_{ijb} u_i^* y_{ib} y_{jb} u_j = \frac{1}{p}\sum_b \mid \sum_i y_{jb} u_j \mid^2$ which is clearly positive.

defined by

$$A \mid 0\rangle = 0.$$

We will normalize this state so that

$$\langle 0 \mid 0\rangle = 1.$$

A convenient way to characterize a probability distribution is in terms of its moments. In operator algebras, this is the expectation value in some state of the powers of some self-adjoint operator ("observable"). We state the main result. Its proof will exploit the analogy with the simple harmonic oscillator. The analogues of the Hemite polynomials will be Chebyshev polynomials of the second kind. The analogue of the gaussian (the ground state) will be the semi-circular distribution.

Proposition 41. *The vacuum expectation values are*

$$\langle 0 \mid (A + A^\dagger)^n \mid 0\rangle = \begin{cases} \frac{1}{k+1}\binom{2k}{k} & n = 2k \\ 0 & n \text{ is odd} \end{cases}$$

The integers $\frac{1}{k+1}\binom{2k}{k}$ are called Catalan numbers. They arise as the solution to many counting problems.

Proof. Define

$$\mid n\rangle = A^{\dagger n} \mid 0\rangle, \quad n = 0, 1, \ldots$$

Using $AA^\dagger = 1$we get

$$\langle m \mid n\rangle = \langle 0 \mid A^m A^{\dagger n} \mid 0\rangle = \langle 0 \mid A^{m-1} A^{\dagger n-1} \mid 0\rangle$$

By iterating this we get

$$\langle m \mid n\rangle = \begin{cases} 1 & m = n \\ \langle 0 \mid A^{m-n} \mid 0\rangle = 0 & m > n \\ \langle 0 \mid A^{\dagger n-m} \mid 0\rangle = 0 & n > m \end{cases}$$

The last equality follows from $\langle 0 \mid A^{\dagger n-m} \mid 0\rangle = \langle 0 \mid A^{n-m} \mid 0\rangle^*$. Thus the states $\mid n\rangle$ are orthonormal and complete (being the eigenstates of the hermitian operator $A^\dagger A$):

$$\langle m \mid n\rangle = \delta_{mn}, \quad \sum_{m=0}^{\infty} \mid m\rangle\langle m \mid = 1$$

Now,

$$Q = A + A^\dagger$$

is a hermitian operator. There must be a family of states which are eigenvectors for this operator:

$$Q \mid q) = q \mid q)$$

Since Q is hermitean, the eigenvalues are real. The relation of these to the states $\mid n\rangle$ should be through some quantities $\psi_n(q)$

$$\mid q) = \sum_{n=0}^{\infty} \psi_n(q) \mid n\rangle$$

We have

$$q \mid q) = (A + A^\dagger) \mid q) = \sum_{n=0}^{\infty} \psi_n(q) \left[A + A^\dagger\right] \mid n\rangle$$

That is,

$$q \sum_{n=0}^{\infty} \psi_n(q) \mid n\rangle = \sum_{n=1}^{\infty} \psi_n(q) \mid n-1\rangle + \sum_{n=0}^{\infty} \psi_n(q) \mid n+1\rangle$$

$$q \sum_{n=0}^{\infty} \psi_n(q) \mid n\rangle = \sum_{n=0}^{\infty} \psi_{n+1}(q) \mid n\rangle + \sum_{n=1}^{\infty} \psi_{n-1}(q) \mid n\rangle$$

So, we have the recursion relations

$$q\psi_0(q) = \psi_1(q) \tag{11.1.1}$$

$$q\psi_n(q) = \psi_{n+1}(q) + \psi_{n-1}(q), \quad n > 1 \tag{11.1.2}$$

Now, recall the Chebyshev polynomials defined by

$$T_n(\cos\theta) = \cos(n\theta), \quad U_n(\cos\theta) = \frac{\sin([n+1]\theta)}{\sin\theta}$$

They satisfy a similar recursion (a direct consequence of the addition formula for cosines and sines):

$$\cos([n+1]\theta) + \cos([n-1]\theta) = 2\cos\theta\cos(n\theta) \iff T_{n-1}(x) + T_{n+1}(x) = 2xT_n(x) \tag{11.1.3}$$

$$\sin([n+1]\theta) + \sin([n-1]\theta) = 2\cos\theta\sin(n\theta) \iff U_{n-1}(x) + U_{n+1}(x) = 2xU_n(x) \tag{11.1.4}$$

Moreover

$$T_0(x) = 1, \quad T_1(x) = x, \quad U_0(x) = 1, \quad U_1(x) = 2x$$

What distinguishes U_n from T_n is the coefficient 2 in the formula for $U_1(x)$. The ansatz

$$\psi_n(q) = \left\{aT_n\left(\frac{q}{2}\right) + bU_n\left(\frac{q}{2}\right)\right\} f(q)$$

will then satisfy (11.1.2). Then (11.1.1) reduces to

$$\left\{a\frac{q}{2} + bq\right\} = q\{a+b\} \implies a = 0.$$

Thus we find that

$$\psi_n(q) = U_n\left(\frac{q}{2}\right)\psi_0(q).$$

The orthogonality relations for the Chebyshev polynomials follow from those of the sine functions:

$$\int_0^\pi \sin[(m+1)\theta]\sin[(n+1)\theta]d\theta = \frac{\pi}{4}\delta_{mn}, \quad m,n = 0,1,2\cdots$$

Putting $q = 2\cos\theta$ this becomes

$$\int_{-2}^{2} U_m\left(\frac{q}{2}\right)U_n\left(\frac{q}{2}\right)\sqrt{4-q^2}dq = 2\pi\delta_{mn}, \quad m,n = 0,1,2\cdots$$

Thus, it emerges that the spectrum of Q is continuous! In fact, it is the interval $[-2,2]$.

Being eigenstates of a self-adjoint operator the states $\mid q)$ satisfy the completeness and orthogonality relations:

$$(q \mid q') = \delta(q - q'), \quad \int \mid q)(q \mid dq = 1,$$

which translates to

$$\sum_{n=0}^{\infty} \psi_n^*(q)\psi_n(q') = \delta(q - q'), \quad \int \psi_m^*(q)\psi_n(q)dq = \delta_{mn}$$

To get $\int \psi_m^*(q)\psi_n(q)dq = \delta_{mn}$ we must choose

$$|\psi_0(q)|^2 = \begin{cases} \frac{1}{2\pi}\sqrt{4 - q^2} & |q| \leq 2 \\ 0 & |q| \geq 2 \end{cases}$$

This is known as the semi-circular distribution.

Now we can calculate the moments

$$\langle 0 \mid (A + A^\dagger)^n \mid 0\rangle = \int q^n \, |\psi_0(q)|^2 \, dq$$

By the symmetry $q \to -q$ this vanishes for odd n. For even n the integral can be calculated (e.g., use Mathematica) to be

$$\int_{-2}^{2} q^{2k} \frac{1}{2\pi}\sqrt{4 - q^2}dq = \frac{1}{k+1}\binom{2k}{k}$$

This proves the formula for the moments. □

11.2. The Gaussian Unitary Ensemble

This is the original model of a random matrix, due to Wigner.

11.2.1. *The Gaussian Unitary Ensemble (GUE) is a random hermitian matrix whose elements taken together have the joint probability distribution function*

$$\frac{1}{Z} e^{-\frac{1}{2\sigma^2}\mathrm{tr}A^2} dA$$

So, A is a hermitean matrix[2] whose matrix elements are gaussian random variables. These are N^2 independent real random variables, once the condition of hermiticity is imposed.

dA stands for the measure of integration over these N^2 components. Also, Z is a normalization constant.

We assume that the mean of each matrix element is zero. The probability distribution function (p.d.f.) is invariant under the action of the unitary group

$$A \to UAU^\dagger, \quad A \in U(N)$$

which explains the name.

11.2.1.1. *The Gaussian Orthogonal Ensemble (GOE) is a random real symmetric matrix with the analogous p.d.f.*

This time it is invariant under the orthogonal group action

$$A \to gAg^T, \quad g \in O(N)$$

There is also a Gaussian Symplectic Ensemble which is a quaternionic analogue. We won't pursue these models further here. See the book by Mehta for more on this.

11.2.2. *Although the matrix elements are independent variables, the eigenvalues are not*

In particular, the eigenvalues exhibit the phenomenon of level repulsion: Two eigenvalues of a matrix are unlikely to be close to each other. We can prove this by considering the special case of 2×2 hermitian matrices. Any such matrix can be expanded in terms of the Pauli matrices:

$$A = a_0 1 + a_1\sigma_1 + a_2\sigma_2 + a_3\sigma_3 := a_0 + \mathbf{a}\cdot\boldsymbol{\sigma}$$

The eigenvalues are

$$\lambda_{1,2} = a_0 \pm a, \quad a = |\mathbf{a}|$$

The joint pdf of the matrices elements is

$$e^{-\frac{1}{2\sigma^2}\left[a_0^2+|\mathbf{a}|^2\right]}\frac{da_0 d^3\mathbf{a}}{Z}$$

[2] It is a coincidence of notation that both the random matrix and the annihilation operator in the Topelitz algebra are called A. In fact, the random matrix A is more analogous to the quantity $Q = A + A^\dagger$ of that section.

In polar co-ordinates, after integrating over the angular co-ordinates, this becomes

$$e^{-\frac{1}{2\sigma^2}\left[a_0^2+a^2\right]} a^2 \frac{da_0 da}{Z}$$

We have omitted a constant factor of 4π : It can be absorbed into the normalization constant Z. The factor a^2 is the Jacobian for transformation to polar co-ordinates.

Let us express this in terms of eigenvalues

$$a_0 = \frac{\lambda_1 + \lambda_2}{2}, \quad a = \frac{\lambda_1 - \lambda_2}{2}$$

Thus joint of p.d.f. of the eigenvalues is (again absorbing a constant into Z)

$$e^{-\frac{1}{2\sigma^2}\left[\lambda_1^2+\lambda_2^2\right]} |\lambda_1 - \lambda_2|^2 \frac{d\lambda_1 d\lambda_2}{Z}$$

The p.d.f. of λ_1, λ_2 does not factorize as a product of functions of single variables λ_1 and λ_2 , because of the factor $|\lambda_1 - \lambda_2|^2$.; this factor is the Jacobian of the change of variables from matrix elements to eigenvalues. The "level repulsion" is the phenomenon that the joint probability density function for λ_1 and λ_2 vanishes as $\lambda_1 \to \lambda_2$. Another way of thinking of this is to rewrite this as

$$e^{-\left\{\frac{1}{2\sigma^2}\left[\lambda_1^2+\lambda_2^2\right]-2\log|\lambda_1-\lambda_2|\right\}} \frac{d\lambda_1 d\lambda_2}{Z}$$

Physicists are used to thinking of probability distributions in analogy to statistical mechanics., where the probability of a configuration is proportion to $e^{-\beta E}$ where β is the inverse of temperature and E is the energy. So the above distribution behaves as if the "energy" of the configuration λ_1, λ_2 is proportional to

$$E(\lambda_1, \lambda_2) = \frac{1}{2\sigma^2}\left[\lambda_1^2 + \lambda_2^2\right] - 2\log|\lambda_1 - \lambda_2|$$

As λ_1 and λ_2 approach each other the energy grows: As if they repel each other. This is not to be taken literally: There is no actual energy associated to eigenvalues. But this analogy to statistical mechanics gives powerful physical intuition. This is useful when we look at the case of N eigenvalues and let $N \to \infty$.

11.2.3. ***The joint p.d.f. of the eigenvalues of the GUE is***

$$P(\lambda_1, \ldots, \lambda_N) = e^{-\frac{1}{2\sigma^2}\sum_i \lambda_i^2} \prod_{i<j} |\lambda_i - \lambda_j|^2 \frac{d\lambda_1, \ldots, d\lambda_N}{Z}$$

The tricky part is to compute the Jacobian $\prod_{i<j} |\lambda_i - \lambda_j|^2$: the transformation to eigenvalues is analogous to that to polar co-ordinates. This factor can be thought of the volume of the set of all matrices of a given spectrum $\{\lambda_1, \ldots, \lambda_N\}$. Thus, there is a unitary matrix such that

$$A = U\mathrm{diag}(\lambda_1, \ldots, \lambda_N)U^\dagger$$

$$dA = U\mathrm{diag}(d\lambda_1, \ldots, d\lambda_N)U^\dagger + [dUU^\dagger, A]$$

$$\mathrm{tr}\, dA^2 = \sum_i d\lambda_i^2 + 2\sum_{i<j}(\lambda_i - \lambda_j)^2 |\left[U^\dagger dU\right]_{ij}|^2$$

This is similar to the formula of the metric of Euclidean space in polar co-ordinates. The eigenvalues are the "radial co-ordinates"; the unitary matrix is like a rotation, and contains the "angular co-ordinates". The analogue of the sphere is the coset space $U(N)/(U(1))^N$. Its volume is a finite constant times the factor $\prod_{i<j} |\lambda_i - \lambda_j|^2$.

The formula for the p.d.f. above follows from the usual formula for the volume element in Riemannian geometry (square root of the determinant of the metric.) An overall finite factor corresponding to the integral over the "angular co-ordinates (volume of) can be absorbed into the normalization factor Z. See Mehta's book [21] for a more detailed derivation.

11.2.4. ***Of special interest is the probability density of a single eigenvalue obtained by integrating all the others out:***

$$R(\lambda_1) = \int e^{-N\sum_i \lambda_i^2} \prod_{i<j} |\lambda_i - \lambda_j|^2 \frac{d\lambda_2 \cdots d\lambda_N}{Z}$$

Remarkably, this approaches a limit as $N \to \infty$. (We choose $2\sigma^2 \sim N^{-1}$ for this limit to exist.) More precisely,

11.2.5. ***The p.d.f. of $x = \lambda_1$ tends to the semi-circular distribution***

$$R(x) = \begin{cases} \frac{1}{2\pi}\sqrt{4 - x^2} & |x| < 2 \\ 0 & |x| > 2 \end{cases}$$

In particular, the probability for $|x| > 2$ is zero.

The proof of this uses some ideas from statistical mechanics. Consider the negative log of the integrand

$$E(\lambda) = N \sum_i \lambda_i^2 - \sum_{i \neq j} \log |\lambda_i - \lambda_j|$$

This can be thought of as the energy of a set of particles which repel each other by the log potential, but held in place by a background harmonic oscillator potential. (Remember that the probability of a configuration of energy E is proportional to $e^{-\beta E}$; so the negative log of the probability density can be interpreted as a constant times energy.)

If we define the density

$$\rho(\lambda) = \frac{1}{N} \sum_i \delta(\lambda - \lambda_i), \quad \int \rho(\lambda) d\lambda = 1$$

this can be written as

$$E[\rho] = N^2 \left[\int \rho(\lambda) \lambda^2 - \frac{1}{2} \mathcal{P} \int \rho(\lambda) \rho(\lambda') \log |\lambda - \lambda'| d\lambda d\lambda \right]$$

The principal value $\mathcal{P}$ means we are to exclude a small region near $\lambda = \lambda'$ from the integral. In the limit $N \to \infty$ we should expect that the density will tend to some continous function: The pair wise repulsion will push the λ_i apart while the confining potential will prevent them from going off to infinity. Since there are an infinite number of them, they have to approach each other and form a continuous distribution. The most likely configuration will have the least energy. Setting the variation w.r.t. ρ to zero (subject to the condition that $\int \rho(\lambda) d\lambda = 1$)

$$C + \lambda^2 = 2\mathcal{P} \int \rho(\lambda') \log |\lambda - \lambda'| d\lambda'$$

(C is the Lagrange multiplier enforcing the constraint.) Differentiating (to eliminate C) we get the singular integral equation

$$\lambda = \mathcal{P} \int \frac{\rho(\lambda')}{\lambda - \lambda'} d\lambda', \quad \int \rho(\lambda) d\lambda = 1$$

The operator on the r.h.s. is (a constant times) the Hilbert transform, a well known integral transform in complex function theory. The square of the Hilbert transform is the negative of the identity operator. This allows us to solve the equation by evaluating the Hilbert transform of the L.H.S. The solution is

$$\rho(\lambda) = \begin{cases} \frac{1}{2\pi} \sqrt{4 - \lambda^2} & |\lambda| < 2 \\ 0 & |\lambda| \geq 2 \end{cases}$$

Exercise 42. Verify that this is a solution to the integral equation. This needs some complex analysis (Hilbert transform and its connection to branch cuts of a multi-valued function).

Note the remarkable fact that the eigenvalue distribution of random hermitian matrices is the same as that of $A+A^\dagger$ in the Toeplitz algebra. This is the beginning of a new theory of non-commutative random variables. Voiculescu's "Free Probability Theory" extends this idea to relate random matrix theory to the algebra of the Free group. This is a frontier of research, outside the scope of this book. See [22].

11.2.6. ***Also of interest is the correlation function of a pair of eigenvalues T_2 defined by***

$$R_2(\lambda_1, \lambda_2) = \int e^{-N \sum_i \lambda_i^2} \prod_{i<j} |\lambda_i - \lambda_i|^2 \frac{d\lambda_3 \cdots d\lambda_N}{Z}$$

$$T_2(\lambda_1, \lambda_2) = R_2(\lambda_1, \lambda_2) - R(\lambda_1) R(\lambda_2)$$

11.2.7. ***This tends to a universal function of the normalized difference $r = \sqrt{2N}|\lambda - \lambda_1|$***

$$1 - \left[\frac{\sin \pi r}{\pi r}\right]^2$$

We will not study the correlations further here. See the book [23] by Deift and Gioev.

Although the formula is originally derived for Gaussians, the correlation turns out to be the same as this for more or less any ensemble of hermitian random matrices. This universality is reminiscent of that of critical phenomena in statistical physics.

11.2.8. ***Amazingly, numerical computations show that the zeros of the Riemann zeta function $\zeta(\frac{1}{2} + i\lambda)$ have the same correlation function***

Perhaps this is because the zeros λ_i are the eigenvalues of some hermitian matrix. This suggests certain strategies for proving the most famous problem in mathematics, the Riemann hypothesis. So far, they have not worked though.

This page intentionally left blank

Chapter 12

HARMONIC ANALYSIS ON FINITE GROUPS

12.1. Discrete Fourier Series

Harmonic analysis is a generalization of Fourier analysis. So let us begin with the simplest kind of Fourier analysis, based on a finite cyclic group Z_Λ, where $\Lambda \geq 2$ is an integer. We can identify $Z_\Lambda = \mathbb{Z}/\Lambda\mathbb{Z}$ as the quotient of the additive group of integers by the subgroup of numbers which are multiples of Λ. We can think of Z_Λ equivalently as the multiplicative group of complex numbers ζ satisfying $\zeta^\Lambda = 1$. Whether we mean the additive or multiplicative picture should be clear from the context; they are isomorphic to each other.

A function $\phi : Z_\Lambda \to \mathbb{C}$ is simply a sequence of complex numbers $\phi(n), n \in \mathbb{Z}$ satisfying the periodicity condition

$$\phi(n + \Lambda) = \phi(n),$$

which says that ϕ descends to a function on the quotient $Z_\Lambda = \mathbb{Z}/\Lambda\mathbb{Z}$.

A physical application could be a one-dimensional model of a lattice of atoms: Arranged at equal separation, with the (unphysical) boundary condition that the last atom be the nearest neighbor of the first.[1]

The obvious example of such a function is the exponential $e^{\frac{2\pi i}{\Lambda}n}$ or any of its powers $e^{\frac{2\pi i}{\Lambda}kn}$. Each exponential spans a one-dimensional representation of the cyclic group Z_Λ:

$$e^{\frac{2\pi i}{\Lambda}k(n+n')} = e^{\frac{2\pi i}{\Lambda}kn} e^{\frac{2\pi i}{\Lambda}kn'}.$$

Later we will see a generalization where Z_Λ is replaced by a finite non-abelian group. Then these exponentials are replaced by the matrix elements of a representation, possibly of dimension greater than one.

[1] Such periodic boundary conditions are commonly used as a prelude to taking $\Lambda \to \infty$ in the statistical physics of lattices.

Clearly, there are Λ independent complex numbers $\phi(0), \phi(1) \cdots \phi(\Lambda-1)$ that completely specify such a function $\phi : Z_\Lambda \to \mathbb{C}$. On the other hand, there are Λ independent exponentials:

$$1, e^{\frac{2\pi i}{\Lambda}n}, e^{\frac{2\pi i}{\Lambda}2n}, e^{\frac{2\pi i}{\Lambda}3n}, \cdots, e^{\frac{2\pi i}{\Lambda}(\Lambda-1)n}.$$

The constant function equal to one is the zeroth power of $e^{\frac{2\pi i}{\Lambda}n}$; it is also the Λth power. The basic theorem of Discrete Fourier series is that these exponentials are a basis in the space of all periodic functions of period Λ. That is, there is a periodic sequence $\tilde{\phi}(k)$, $k \in \mathbb{Z}$ with $\tilde{\phi}(k+\Lambda) = \tilde{\phi}(k)$ such that

$$\phi(n) = \frac{1}{\Lambda}\sum_{k=0}^{\Lambda-1} \tilde{\phi}(k) e^{\frac{2\pi i}{\Lambda}kn}.$$

The overall factor of Λ is put in for later convenience.

Indeed we can give a formula for these coefficients:

$$\tilde{\phi}(k) = \sum_{m=0}^{\Lambda-1} \phi(m) e^{-\frac{2\pi i}{\Lambda}km}$$

The proof is based on the following identity

$$\sum_{k=0}^{\Lambda-1} e^{\frac{2\pi i}{\Lambda}kn} = \begin{cases} \Lambda & n = 0 \\ 0 & n = 1, \ldots, \Lambda - 1 \end{cases} \tag{12.1.1}$$

The first statement is obvious: When $n = 0$ each term in the sum is equal to one. When $n = 1, \ldots, \Lambda - 1$ suppose

$$S_n = \sum_{k=0}^{\Lambda-1} e^{\frac{2\pi i}{\Lambda}kn}$$

In the sum we can shift $k \mapsto k - 1$:

$$S_n = \sum_{k=0}^{\Lambda-1} e^{\frac{2\pi i}{\Lambda}kn} = \sum_{k=1}^{\Lambda} e^{\frac{2\pi i}{\Lambda}(k-1)n} = e^{-\frac{2\pi i}{\Lambda}n} \sum_{k=1}^{\Lambda} e^{\frac{2\pi i}{\Lambda}kn} = e^{-\frac{2\pi i}{\Lambda}} \sum_{k=1}^{\Lambda} e^{\frac{2\pi i}{\Lambda}kn}$$

The $k = \Lambda$ term of the last sum is equal to one, so that

$$\sum_{k=1}^{\Lambda} e^{\frac{2\pi i}{\Lambda}kn} = \sum_{k=0}^{\Lambda-1} e^{\frac{2\pi i}{\Lambda}kn}.$$

That is,

$$S_n = e^{-\frac{2\pi i}{\Lambda}n} S_n$$

Since $n \neq 0$ this means that $S_n = 0$, as was to be proved.

This result can be restated as a

Lemma 43. *The exponentials satisfy the relation:*

$$\frac{1}{\Lambda}\sum_{k=0}^{\Lambda-1} e^{\frac{2\pi i}{\Lambda}k(m-n)} = \delta\,(m = n \mod \Lambda) \tag{12.1.2}$$

The r.h.s. is equal to 1 if $m = n \mod \Lambda$ and zero otherwise.

Now we can prove the fundamental theorem of Discrete Fourier analysis:

Theorem. *Given a function $\phi : Z_\Lambda \to \mathbb{C}$, define*

$$\tilde{\phi}(k) = \sum_{m=0}^{\Lambda-1} \phi(m) e^{-\frac{2\pi i}{\Lambda}km}.$$

Then

$$\phi(n) = \frac{1}{\Lambda}\sum_{k=0}^{\Lambda-1} \tilde{\phi}(k) e^{\frac{2\pi i}{\Lambda}kn}.$$

Proof. Substitute the definition of $\tilde{\phi}(m)$ into the series:

$$\phi(n) = \frac{1}{\Lambda}\sum_{k=0}^{\Lambda-1}\left(\sum_{m=0}^{\Lambda-1}\phi(m)e^{-\frac{2\pi i}{\Lambda}km}e^{\frac{2\pi i}{\Lambda}kn}\right) = \sum_{m=0}^{\Lambda-1}\left\{\frac{1}{\Lambda}\sum_{k=0}^{\Lambda-1}e^{\frac{2\pi i}{\Lambda}k(m-n)}\right\}\phi(m)$$

Using the Lemma above,

$$= \sum_{m=0}^{\Lambda-1}\delta\,(m = n \mod \Lambda)\,\phi(m) = \phi(n).$$

□

Exercise 44. Find the eigenvalues λ of the difference equation $\psi_{m+1} - 2\psi_m + \psi_{m-1} = \lambda\psi_m$ subject to the periodic boundary conditions $\psi_{m+\Lambda} = \psi_m$.

Solution First of all

$$\sum_{m=0}^{\Lambda-1} e^{-\frac{2\pi i}{\Lambda}km}\psi_{m+1} = \sum_{m=0}^{\Lambda-1} e^{-\frac{2\pi i}{\Lambda}k(m-1)}\psi_m = e^{\frac{2\pi i}{\Lambda}k}\sum_{m=0}^{\Lambda-1} e^{-\frac{2\pi i}{\Lambda}km}\psi_m = e^{\frac{2\pi i}{\Lambda}k}\tilde{\psi}_k$$

and similarly

$$\sum_{m=0}^{\Lambda-1} e^{-\frac{2\pi i}{\Lambda}km}\psi_{m-1} = e^{-\frac{2\pi i}{\Lambda}k}\tilde{\psi}_k.$$

After Fourier analysis the eigenvalue equation becomes

$$e^{\frac{2\pi i}{\Lambda}k}\tilde{\psi}_k - 2\tilde{\psi}_k + e^{-\frac{2\pi i}{\Lambda}k}\tilde{\psi}_k = \lambda\tilde{\psi}_k \iff$$

$$\left\{2\cos\tfrac{2\pi k}{\Lambda} - 2\right\}\tilde{\psi}_k = \lambda\tilde{\psi}_k \iff -4\sin^2\left(\frac{\pi k}{\Lambda}\right)\tilde{\psi}_k = \lambda\tilde{\psi}_k$$

Thus the eigenvalues are $\lambda_k = -4\sin^2\left(\frac{\pi k}{\Lambda}\right), \quad k = 0, 1, \cdots \Lambda - 1$. This is the discrete version of the eigenvalue equation of the Laplace operator on a circle.

12.1.1. *The Fourier transform of a convolution is a product*

What is the function corresponding to the product of Fourier coefficients of two functions ϕ and ψ? We can calculate

$$\frac{1}{\Lambda}\sum_{k=0}^{\Lambda-1}\tilde{\phi}(k)\tilde{\psi}(k)e^{\frac{2\pi i}{\Lambda}km} = \frac{1}{\Lambda}\sum_{k=0}^{\Lambda-1}\left\{\sum_{n=0}^{\Lambda-1}\phi(n)e^{-\frac{2\pi i}{\Lambda}kn}\right\}\left\{\sum_{l=0}^{\Lambda-1}\psi(l)e^{-\frac{2\pi i}{\Lambda}kl}\right\}e^{\frac{2\pi i}{\Lambda}km}$$

Re-arranging,

$$= \sum_{n,l=0}^{\Lambda-1}\phi(n)\psi(l)\left\{\frac{1}{\Lambda}\sum_{k}e^{\frac{2\pi i}{\Lambda}k(m-n-l)}\right\}$$

Using Eqn. (12.1.2) we get

$$\frac{1}{\Lambda}\sum_{k=0}^{\Lambda-1}\tilde{\phi}(k)\tilde{\psi}(k)e^{\frac{2\pi i}{\Lambda}km} = \sum_{n=0}^{\Lambda-1}\phi(n)\psi(m-n)$$

This operation is called the *convolution*:

$$\phi * \psi(m) = \sum_{n=0}^{\Lambda-1}\phi(n)\psi(m-n)$$

Inverting the Fourier transform we can write the result as

$$\sum_{m=0}^{\Lambda-1}\phi * \psi(m)e^{-\frac{2\pi i}{\Lambda}km} = \tilde{\phi}(k)\tilde{\psi}(k).$$

In the other direction we also have similarly:

$$\sum_{m=0}^{\Lambda-1}\phi(m)\psi(m)e^{-\frac{2\pi i}{\Lambda}km} = \frac{1}{\Lambda}\sum_{l}^{\Lambda-1}\tilde{\phi}(k-l)\tilde{\psi}(l)$$

The Fourier synthesis of a convolution is a product.

These are useful in many applications. Clearly, multiplication is a simpler operation than convolution. The Fourier transform reduces the latter to the former.

Thus, it plays a role similar to the logarithm, which reduces multiplication of numbers to the simpler operation of addition.

12.1.2. *The limit* $\Lambda \to \infty$

We can guess that as $\Lambda \to \infty$ the periodicity condition becomes irrelevant: We just have a function $\phi : \mathbb{Z} \to \mathbb{C}$. The sum

$$\tilde{\phi}(k) = \sum_{m=0}^{\Lambda-1} \phi(m) e^{-\frac{2\pi i}{\Lambda} km}$$

can be given a meaning for large Λ by a change of variable:

$$x = \frac{k}{\Lambda}.$$

These will get very close together as Λ gets large. Define

$$\tilde{\Phi}(x) = \sum_{m=0}^{\Lambda-1} \phi(m) e^{-2\pi i x m}$$

Clearly there is a periodicity $x \to x + 1$. So $x = \frac{\Lambda-2}{\Lambda}, x = \frac{\Lambda-1}{\Lambda}$ is the same as $x = -\frac{2}{\Lambda}, x = -\frac{1}{\Lambda}$, etc. respectively.

Using this we can enumerate the values it takes by starting at zero and going out in both directions:

$$x = \cdots, -\frac{2}{\Lambda}, -\frac{1}{\Lambda}, 0, \frac{1}{\Lambda}, \frac{2}{\Lambda} \cdots,$$

Roughly $\frac{\Lambda}{2}$ values are to the right of 0 and roughly $\frac{\Lambda}{2}$ are to the left. As $\Lambda \to \infty$ this tends to a continuous variable taking values in the range $\left[-\frac{1}{2}, \frac{1}{2}\right]$.

By identifying $\Lambda - 1$ with -1 and so on we can also enumerate the values of m as

$$m = \cdots - 2, -1, 0, 1, 2, \ldots$$

So, $\tilde{\Phi}(x)$ tends to an infinite sum

$$\tilde{\Phi}(x) = \sum_{m=-\infty}^{\infty} \phi(m) e^{-2\pi i x m}$$

This converges if $\mid \phi(m) \mid$ vanishes fast enough for large $|m|$. For example, each summand has magnitude $\mid \phi(m) \mid$, so the sum converges if $\sum_{m=-\infty}^{\infty} \mid \phi(m) \mid < \infty$.

OTOH, the sum $\frac{1}{\Lambda} \sum_{k=0}^{\Lambda-1} \tilde{\phi}(k) e^{\frac{2\pi i}{\Lambda} kn}$ tends to an integral:

$$\frac{1}{\Lambda} \sum_{k=0}^{\Lambda-1} \tilde{\phi}(k) e^{\frac{2\pi i}{\Lambda} kn} \to \int_{-\frac{1}{2}}^{\frac{1}{2}} \tilde{\Phi}(x) e^{2\pi i x m} dx$$

Thus there is a version of Fourier analysis for functions on an infinite lattice:

$$\phi(m) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \tilde{\Phi}(x) e^{2\pi i x m} dx$$

Although our arguments are not mathematically rigorous, this can be proved under some assumptions on the decay of ϕ at infinity. It is in fact closer to the original analysis of Fourier (who was an engineer solving problems in heat conduction.)

The range of values of x can be thought of as a circle: It is a periodic variable with period 1. So, the Fourier transform of a function on $\mathbb{Z}$ is a function on a circle. We will return to this theme in the next chapter.

12.1.2.1. *Convolutions*

The convolution of functions on $\mathbb{Z}$ is defined similarly:

$$\phi * \psi(m) = \sum_{n=-\infty}^{\infty} \phi(n)\psi(m-n)$$

This makes sense if the functions vanish at infinity fast enough (e.g., are square integrable). Again, the Fourier coefficients of a convolution is a product:

$$\sum_{m=-\infty}^{\infty} \phi * \psi(m) e^{-2\pi i x m} = \tilde{\Phi}(x)\tilde{\Psi}(x)$$

In the other direction, we have

$$\begin{aligned}\sum_{m=-\infty}^{\infty} \phi(m)\psi(m) e^{-2\pi i x m} &= \left\{\int_{-\frac{1}{2}}^{\frac{1}{2}} \tilde{\Phi}(y) e^{2\pi i y m} dy\right\}\left\{\int_{-\frac{1}{2}}^{\frac{1}{2}} \tilde{\Psi}(z) e^{2\pi i z m} dz\right\} \\ &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \tilde{\Phi}(y)\tilde{\Psi}(z)\left\{\sum_{m\in\mathbb{Z}} e^{2\pi i (y+z-x)m}\right\} dy dz\end{aligned}$$

Exercise 45. Prove an identity analogous to Lemma (12.1.2)

$$\sum_{m\in\mathbb{Z}} e^{2\pi i x m} = \delta(x)$$

where the Dirac delta function for periodic functions of period one satisfies $\int_{-\frac{1}{2}}^{\frac{1}{2}} \phi(y)\delta(x-y)dy = \phi(x)$.

Using this, we get

$$\sum_{m=-\infty}^{\infty} \phi(m)\psi(m)e^{-2\pi ixm} = \int_{-\frac{1}{2}}^{\frac{1}{2}} \tilde{\Phi}(y)\tilde{\Psi}(x-y)dy$$

The r.h.s. is a convolution on the circle, thought of as a group under the operation of addition modulo 1. Instead of a sum we have an integral.

12.2. Non-abelian Finite Groups

The remarkable fact is that there is a far reaching generalization of Fourier analysis to non-abelian groups. We will see that the "dual" of a group is the set of its irreducible representations. The exponentials of Fourier are replaced by the matrix elements of these representations. All representations in this section will be assumed to be finite dimensional complex vector spaces; and all groups are finite. Later we will extend to certain kinds of Lie groups (i.e., compact).

The main reference for this section is the excellent book [24].

12.2.1. *Finite dimensional representations of a finite group are unitarizable*

That is, each of them have an invariant inner product. This can be established by a simple trick. Pick some inner product $((,))$ in the vector space V carrying the representation ρ. Define a new inner product by "averaging" over the group:

$$(a,b) = \frac{1}{|G|}\sum_{g\in G} ((\rho(g)a, \rho(g)b)), \quad a,b \in V$$

The group being finite, we don't have to worry about the convergence of the sum; this could be a problem for infinite groups.

It is easy to see that this average is invariant:

$$\begin{aligned}(\rho(h)a, \rho(h)b) &= \frac{1}{|G|}\sum_{g\in G} ((\rho(g)\rho(h)a, \rho(g)\rho(h)b)) \\ &= \frac{1}{|G|}\sum_{g\in G} ((\rho(gh)a, \rho(gh)b)) = \frac{1}{|G|}\sum_{g\in G} ((\rho(g)a, \rho(g)b)).\end{aligned}$$

In the last step, we re-enumerate the group elements by $g \mapsto gh^{-1}$. So $(\rho(h)a, \rho(h)b) = (a,b)$ as needed.

12.2.2. *Finite dimensional reducible unitary representations are completely reducible*

If $W \subset V$ is an invariant subspace of the vector space carrying a unitary representation, then its orthogonal complement $W^\perp = \{v \mid (v, w) = 0 \forall w \in W\}$ is also invariant:

$\forall w \in W \& \forall v \in W^\perp,\ \rho(g)w \in W \implies (v, \rho(g)w) = 0 \implies (\rho(g^\dagger)v, w) = 0 \implies \rho(g)v \in W^\perp$.

Thus, the representation can be split as the direct sum of two representations

$$V = W \oplus W^\perp.$$

The representation matrices are then block diagonal[2] $\begin{pmatrix} \rho_W(g) & 0 \\ 0 & \rho_{W^\perp}(g) \end{pmatrix}$. Then we can repeat this argument on W and $W^\perp$ until V is expressed as a sum of irreducible representations (which don't have any proper invariant subspaces). Thus any finite dimensional unitary representation can be expressed as a direct sum[3]

$$\rho = \bigoplus_{r \in \tilde{G}} m_r r$$

Here, $\tilde{G}$ denotes the set of all equivalence classes of unitary irreducible representations. Also, $m_r = 0, 1, 2, \ldots$ (called the *multiplicity*) is the number of copies of eac h irreducible representation contained in ρ. This is similar to the decomposition of any number as a product of primes; the irreducible representations are like primes and the multiplicities are like the exponents of each prime.

To proceed further we will need several tools of representation theory. We start with

12.2.3. *Schur's lemma*

This is really a couple of results. The first of them is

Lemma 46. *Let r and s be two irreducible representations,on vector spaces V and W respectively, of a finite group G; and T is a linear map $T : V \to W$ such that*

$$T\left[r(g)v\right] = s(g)\left[Tv\right], \forall g \in G, v \in V \tag{12.2.1}$$

Then either T is an isomorphism or it is zero.

[2] If ρ is not unitary, there is the possibility that representation matrices are only block-triangular $\begin{pmatrix} \rho_W(g) & b(g) \\ 0 & \rho_{W^\perp}(g) \end{pmatrix}$; that is, the representation is not totally reducible.

[3] If the representation ρ is irreducible, there is just one term in this direct sum.

A map satisfying the condition (12.2.1)is called an *intertwiner*. First, we note that ker$T \equiv \{v \in V \mid Tv = 0\}$ is invariant under $r(g)$:

$$Tv = 0 \implies s(g)Tv = 0 \implies T\,[r(g)v] = 0.$$

Similarly, Image $T \equiv \{Tv \mid v \in V\}$ is an invariant under $s(g)$.

But, since r is an irreducible representation, an invariant subspace is either zero or all of V. Thus either ker $T = 0$ or ker $T = V$.

Since s is also irreducible, Image $T = 0$ or Image $T = W$. The only two possibilities are

$$\ker T = 0, \quad \text{Image } T = W$$

in which case T is an isomorphism; or

$$\ker T = V, \quad \text{Image } T = 0$$

in which case it is zero.This proves the Lemma.

Thus, if r is not equivalent to s the only intertwiner between them is zero. If $r \sim s$, we can identify V and W. An intertwiner of r to itself is simply a matrix that commutes with all of the representation matrices.

Corollary 47. *If an operator commutes with all the representation matrices of an irreducible representation, it is a multiple of the identity.*

For, every linear operator $T : V \to V$ has at least one eigenvalue (recall that V is a complex vector space) λ. If T is an intertwiner of an irerducible representation to itself, $T - \lambda 1$ is also an interwtiner. It cannot be an isomorphism because it cannot be invertible (the definition of an eigenvalue!). So, it must vanish; i.e., $T = \lambda 1$.

Corollary 48. *Any irreducible representation of an abelian group is one dimensional.*

The point is that the representation matrices commute with each other; so are multiples of the identity in an irreducible representation.

Another tool we need is that of a character

12.2.4. *The trace of the representation matrices associates a function on the group to every representation, called its character*

Let ρ be some representation. Define a function $\chi_\rho : G \to \mathbb{C}$ by

$$\chi_\rho(g) = \text{tr}\rho(g)$$

This functions is called the character. Note that this function is invariant under conjugation:

$$\chi_\rho(hgh^{-1}) = \chi_\rho(g), \quad g, h \in G.$$

Also its value at the identity is the dimension of the representation.

$$\chi_\rho(1) = \dim \rho.$$

Recall that if $\rho$and σ are two representations, their direct sum is another representation, whose matrices are block-diagonal:

$$\rho \oplus \sigma(g) = \begin{pmatrix} \rho(g) & 0 \\ 0 & \sigma(g) \end{pmatrix}$$

It takes but a moment to realize that

$$\chi_{\rho\oplus\sigma}(g) = \chi_\rho(g) + \chi_\sigma(g)$$

Iterating this, we see that if we decompose ρ into ireducible representations

$$\rho = \bigoplus_{r\in\tilde{G}} m_r r$$

we have a formula for its character:

$$\chi_\rho(g) = \sum_{r\in\tilde{G}} m_r \chi_r(g)$$

12.2.5. *An inner product on functions: $l^2(G)$*

The space of complex valued functions on G is a complex vector space whose dimension is equal to the cardinality $|G|$ of G: A function is completely determined by its values at each point of the group.[4]

We can establish an inner product on it.

$$\langle \phi, \psi \rangle = \sum_{g\in G} \phi^*(g)\psi(g).$$

We will call this vector space with inner product $l^2(G)$.

[4]Put another way, the functions δ_h, $h \in G$ equal to one at h and zero everywhere else is a basis for the space of functions on G. See below,

12.2.6. *The matrix elements of inequivalent irreducible representations are orthogonal in $l^2(G)$*

Suppose r and s are two irreducible representations of G which are not equivalent to each other. We have orthonormal bases $e_a, a = 1, \ldots \dim r$ and $f_i, i = 1, 2, \ldots \dim s$ in their vector spaces. Using them we have the matrix elements

$$r_{ab}(g) = (e_a, r(g)e_b), \quad s_{ij}(g) = \left(f_i, s(g)f_j\right)$$

which are functions on G for each ab and ij.

The $l^2(G)$ inner product of these functions is

$$\langle r_{ab}, s_{ij}\rangle \equiv \sum_{h\in G} r^*_{ab}(h)s_{ij}(h)$$

Using unitarity this can be also be written

$$\langle r_{ab}, s_{ij}\rangle \equiv \sum_{h\in G} r_{ba}(h^{-1})s_{ij}(h)$$

The trick is to construct an intertwiner between the two representations out of this inner product. Since r and s are inequivalent, such an intertwiner would have to be zero.

For each choice bj, define a matrix with elements.[5]

$$T_{ia} = \sum_{h\in G} r_{ba}(h^{-1})s_{ij}(h).$$

Recalling (12.2.1), we need to show that

$$\sum_{a=1}^{\dim r} T_{ia}r_{ac}(g) = \sum_{k=1}^{\dim s} s_{ik}(g)T_{kc}$$

Now

$$LHS \equiv \sum_{a=1}^{\dim r} T_{ia}r_{ac}(g) = \sum_{a=1}^{\dim r}\sum_{h\in G} r_{ba}(h^{-1})s_{ij}(h)r_{ac}(g)$$

Using the representation property we can combine the first and last factors:

$$LHS = \sum_{h\in G} s_{ij}(h)r_{bc}(h^{-1}g)$$

[5] We suppress the indices $b.j$ to simplify the notation; but beware that T_{ia} does depend on the choice of b, j.

OTOH,

$$RHS \equiv \sum_{k=1}^{\dim s} s_{ik}(g) T_{kc} = \sum_{h\in G} \sum_{k=1}^{\dim s} s_{ik}(g) r_{bc}(h^{-1}) s_{kj}(h)$$

Again combining the first and last factors,

$$RHS = \sum_{h\in G} s_{ij}(gh) r_{bc}(h^{-1})$$

Replacing $h \mapsto g^{-1}h$

$$RHS = \sum_{h\in G} s_{ij}(h) r_{bc}(h^{-1}g)$$

Thus, we get $LHS = RHS$. This proves T is an intertwiner; but then it has to be zero since r and s are inequivalent. This being true for each choice of b, j, we get $\langle r_{ab}, s_{ij}\rangle = 0$.

12.2.7. *The inner product between any pair of matrix elements*

Now we consider the situation when r is equivalent to s. By choosing appropriate bases, we can identify the matrix elements of r and s. We look at the inner product

$$\langle r_{ab}, r_{a'b'}\rangle \equiv \sum_{h\in G} r_{ba}(h^{-1}) r_{a'b'}(h)$$

Again, keeping b, b' fixed we define

$$T_{a'a} = \sum_{h\in G} r_{ba}(h^{-1}) r_{a'b'}(h)$$

The same argument as before shows that this is an intertwiner of r to itself:

$$\sum_{a=1}^{\dim r} T_{a'a} r_{ac}(g) = \sum_{a=1}^{\dim r} r_{a'a}(g) T_{ac}$$

But this time Schur's lemma tells us that $T_{a'a}$ is a multiple of the identity!

$$\sum_{h\in G} r^*_{ab}(h) r_{a'b'}(h) = \lambda^r_{bb'} \delta_{a'a}$$

The eigenvalue can depend on r, b and b', which we show explicitly. If we set $a = a'$ and sum over a, this gives (using the representation property on the lhs)

$$\sum_{h\in G} r_{bb'}(1) = \lambda^r_{bb'} \dim r$$

Since the summand on the lhs is independent of h, we get

$$|G| r_{bb'}(1) = \lambda^r_{bb'} \dim r$$

But $r_{bb'}(1) = \delta_{bb'}$.

$$|G|\delta_{bb'} = \lambda^r_{bb'} \dim r \implies \lambda^r_{bb'} = \frac{|G|}{\dim r}\delta_{bb'}$$

and

$$\langle r_{ab}, r_{a'b'}\rangle \equiv \frac{|G|}{\dim r}\delta_{aa'}\delta_{bb'}$$

12.2.8. *The component of a function along each irreducible representation*

Combining with the earlier result of orthogonality when $r \nsim s$, we have

$$\langle r_{ab}, s_{ij}\rangle \equiv \sum_{h\in G} r^*_{ab}(h)s_{ij}(h) = \frac{|G|}{\dim r}\delta_{rs}\delta_{bj}\delta_{ai}$$

Equivalently, $\sqrt{\frac{\dim r}{|G|}} r_{ab}$ is an orthonormal set of functions in $l^2(G)$ as r runs over the set of equivalence classes $\tilde{G}$ of representations of G, and a, b label the bases in each representations. These are analogous to the exponential functions of Fourier analysis.

So, the $(\dim r)^2$ numbers

$$\tilde{\phi}^r_{ab} = \sum_{h\in G} r^*_{ab}(h)\phi(h)$$

can be thought of as the components of the complex-valued function ϕ for each irreducible representation $r \in \tilde{G}$. Do these components completely determine a function? That is, can we reconstruct ϕ from its components $\tilde{\phi}^r_{ab}$?

To answer this we need to prove the completeness of this decomposition. The essential tool is the character of representations.

12.2.9. *Orthogonality of characters*

A particular case of orthogonality is of special interest. By setting $a = b$ and $i = j$ and summing over a and i we get from the above

$$\langle \chi_r, \chi_s \rangle \equiv \sum_{h \in G} \chi_r(h^{-1}) \chi_s(h) = |G| \delta_{rs}, \quad r, s \in \tilde{G}$$

Thus, the characters of inequivalent representations are orthogonal to each other. Moreover, the l^2 norm of the character of an irreducible representation is $\sqrt{|G|}$.

Suppose we have a representation

$$\rho = \bigoplus_{r \in \tilde{G}} m_r r$$

so that

$$\chi_\rho = \sum_{r \in \tilde{G}} m_r \chi_r .$$

If we know the character function of ρwe can determine the multiplicities:

$$\langle \chi_\rho, \chi_r \rangle = \sum_{s \in \tilde{G}} m_s \langle \chi_s, \chi_r \rangle = |G| m_r \implies m_r = \frac{1}{|G|} \langle \chi_\rho, \chi_r \rangle$$

This justifies the name character: It completely determines the representation!

12.2.10. *The left regular representation*

The vector space $l^2(G)$ carries a representation of the group, called its "left regular representation". That is, to every $h \in G$ we associate a linear operator $L(h) : l^2(G) \to l^2(G)$ by the formula

$$[L(h)\phi](g) = \phi(h^{-1} g), \quad h \in G.$$

We can check that this is indeed a representation[6]:

$$\begin{aligned} [L(h_1)L(h_2)\phi](g) &= [L(h_2)\phi](h_1^{-1} g) = \phi(h_2^{-1} h_1^{-1} g) \\ &= \phi((h_1 h_2)^{-1} g) = [L(h_1 h_2)\phi](g). \end{aligned}$$

[6] A point about notation: $L(h_1)L(h_2)\phi$ means that we act with $L(h_1)$ acts on the result of the action of with $L(h_2)$ on ϕ. This is the meaning of the first equality.

(What would have been wrong with defining $[L(h)\phi]\,(g) = \phi\,(hg)$?). Moreover, this is a unitary representation. That is

$$\langle L(h)\phi, L(h)\psi\rangle = \langle \phi, \psi\rangle, \quad h \in G.$$

For,

$$\langle L(h)\phi, L(h)\psi\rangle = \sum_{g\in G} \phi^*(h^{-1}g)\psi(h^{-1}g)$$

In the sum we can replace g by hg: This is a one-one correspondence of elements, just a different way to enumerate them. But then the r.h.s. just becomes $\langle \phi, \psi\rangle$, proving invariance.

Thus, the left regular representation is a unitary representation of dimension $|G|$ of the group.

Remark 49. The left regular representation of G is faithful. That is, $L(h)\phi = \phi, \forall \phi \in l^2(G) \implies h = 1$. (Just choose ϕ to be 1 at the identity and zero everywhere else).

Of course, there is a mirror image of this construction that gives the right regular representation:

$$[R(h')\phi](g) = \phi(gh'), \quad h' \in G$$

It is also a faithful unitary representation of G. But it contains the same information, so we just study the left regular.

12.2.11. *The character of the left regular representation*

Let us now calculate the character of the left regular representation. A basis for $l^2(G)$ is given by the functions δ_h defined by

$$\delta_h(g) = \begin{cases} 1 & g = h \\ 0 & g \neq h \end{cases}$$

Clearly any function can be expanded uniquely in terms of them:

$$\phi(g) = \sum_{h\in G} \phi(g)\delta_h(g).$$

(This is another way of seeing that $l^2(G)$ has dimension $|G|$.) Also, this basis is orthonormal:

$$\langle \delta_h, \delta_{h'} \rangle = \sum_{g \in G} \delta_h(g)\delta_{h'}(g) = \begin{cases} 1 & h = h' \\ 0 & h \neq h' \end{cases}$$

The representation matrices are

$$L_{h,h'}(g) = \sum_{g' \in G} \delta_h(g')\delta_{h'}(g^{-1}g') = \begin{cases} 1 & h = gh' \\ 0 & h \neq gh' \end{cases}$$

The diagonal entries are all equal to one if g is the identity and zero other wise:

$$L_{h,h}(g) = \begin{cases} 1 & 1 = g \\ 0 & 1 \neq g \end{cases}$$

Summing over h, there are $|G|$ terms equal to one when g is the identity:

$$\chi_L(g) = \begin{cases} |G| & g = 1 \\ 0 & g \neq 1 \end{cases}$$

12.2.12. *Left Regular is the mother of all representations*

See [25] for further study of the "mother of all representations".

Like any finite dimensional representation, the left regular representation can be decomposed into a direct sum of irreducibles.

$$L = \bigoplus_{r \in \tilde{G}} m_r r. \tag{12.1.2}$$

Since we know the character of L we can compute the multiplicities:

$$m_r = \frac{1}{|G|}\langle \chi_L, \chi_r \rangle = \chi_r(1) = \dim r$$

That is, left-regular representation contains *every* irreducible representation! In fact it contains $\dim r$ copies of each irreducible representation r. We have moreover

$$\chi_L(g) = \sum_{r \in \tilde{G}} \dim r \ \chi_r(g)$$

In particular, this gives for $g = 1$

$$\dim L = \sum_{r \in \hat{G}} (\dim r)^2$$

But the dimension of the left regular representation is $|G|$:

$$|G| = \sum_{r \in \hat{G}} (\dim r)^2 \tag{12.1.3}$$

This is a simple and beautiful relation between the irreducible representations and the size of the group.

Remark 50. Of course we could have also said all this about the right regular representation.In fact you see that the left-multiplicity of $r$being $\dim r$ simply has to do with the action of G on the right.This becomes especially clear if we look at how the components transform under the left and right regular actions. Perhaps Right Regular is the father of all representations!

12.2.13. *The transformations of the irreducible components of a function under the left and right regular actions*

Suppose we transform a function by the left regular action: $\phi \mapsto L_g\phi$ where

$$L_g\phi(h) = \phi(g^{-1}h).$$

How do its components change?

$$\begin{aligned}\widetilde{[L_g\phi]}^r_{ab} &= \sum_{h \in G} r^*_{ab}(h)\phi(g^{-1}h) \\ &= \sum_{h \in G} r^*_{ab}(gh)\phi(h) \\ &= \sum_{h \in G} r^*_{ac}(g) r^*_{cb}(h)\phi(h)\end{aligned}$$

so that

$$\widetilde{[L_g\phi]}^r_{ab} = r^*_{ac}(g)\tilde{\phi}^r_{cb}.$$

Similarly the right regular action

$$R_g\phi(h) = \phi(hg)$$

would give

$$\begin{aligned}\widetilde{[R_g\phi]}^r_{ab} &= \sum_{h\in G} r^*_{ab}(h)\phi(hg)\\ &= \sum_{h\in G} r^*_{ab}(hg^{-1})\phi(h)\\ &= \sum_{h\in G} r^*_{ac}(h)r^*_{cb}(g^{-1})\phi(h)\end{aligned}$$

so that

$$\widetilde{[R_g\phi]}^r_{ab} = \tilde{\phi}^r_{ac}r^*_{cb}(g^{-1})$$

So we see that the left and right regular actions act on the left and right indices of the components. As with Fourier analysis, there is a complex conjugation when we pass to the irreducible components.

12.2.14. *Peter–Weyl theorem for finite groups*

The decomposition (12.1.2) can be stated more explicitly: Every function on the group can be decomposed uniquely as a sum of matrix elements of irreducible representations. This is the completeness we seekd earlier.

Theorem 51. ***Peter-Weyl Theorem*** *Let $\phi : G \to \mathbb{C}$ be a function with Fourier components in each irreducible representation $r \in \tilde{G}$*

$$\tilde{\phi}^r_{ab} = \sum_{g\in G} r^*_{ab}(g)\phi(g)$$

The Fourier series synthesizes ϕ from its components:

$$\phi(g) = \sum_{r\in\tilde{G}}\sum_{ab}\frac{\dim r}{|G|}\tilde{\phi}^r_{ab}r_{ab}(g)$$

By choosing $\phi(g) = \delta_h(g)$, last identity can also be written as a completeness relation for the representation matrices:

$$\delta_h(g) = \sum_{r\in\tilde{G}}\sum_{ab}\frac{\dim r}{|G|}r^*_{ab}(h)r_{ab}(g)$$

12.2.15. *Convolution on groups*

Let G be a finite group: We no longer require it to be abelian. The space of functions $l^2(G)$ is a vector space of dimension equal to the number of elements $|G|$ of the group. This vector space becomes an algebra under the convolution

$$\phi * \psi(g) = \sum_{h \in G} \phi(h)\psi(h^{-1}g)$$

If G is not abelian, this convolution is not commutative: $\phi * \psi \neq \psi * \phi$ in general. But it is always associative!

Proposition 52. *The convolution product of functions on a group is associative*:

$$(\phi * \psi) * \eta = \phi * (\psi * \eta)$$

Proof. Let us write

$$\phi * \psi(g) = \sum_{h_1 \in G} \phi(h_1)\psi(h_1^{-1}g)$$

so that

$$[(\phi * \psi) * \eta](g) = \sum_{h_2 \in G} [\phi * \psi](h_2)\eta(h_2^{-1}g) = \sum_{h_1, h_2 \in G} \phi(h_1)\psi(h_1^{-1}h_2)\eta(h_2^{-1}g)$$

In the last sum replace $h_2 \mapsto h_1 h_2$ to get

$$[(\phi * \psi) * \eta](g) = \sum_{h_1, h_2 \in G} \phi(h_1)\psi(h_2)\eta([h_1 h_2]^{-1} g)$$

OTOH, we can write

$$\psi * \eta(g) = \sum_{h \in G} \psi(h_2)\eta(h_2^{-1}g)$$

and

$$\begin{aligned}[\phi * (\psi * \eta)](g) &= \sum_{h_1} \phi(h_1)\,[\psi * \eta](h_1^{-1}g) \\ &= \sum_{h_1, h_2 \in G} \phi(h_1)\psi(h_2)\eta\left(h_2^{-1}h_1^{-1}g\right)\end{aligned}$$

Since $[h_1 h_2]^{-1} = h_2^{-1}h_1^{-1}$ we have the equality we need. □

12.2.16. *Decomposition of the convolution algebra into matrix algebras*

Let us express the convolution in terms of the components of the functions.

$$\left[\widetilde{\phi * \psi}\right]^r_{ab} = \sum_{h\in G} r^*_{ab}(h)\phi * \psi(h)$$

$$= \sum_{h,h'\in G} r^*_{ab}(h)\phi(h')\psi(h'^{-1}h)$$

Replace $h \mapsto h'h$

$$\left[\widetilde{\phi * \psi}\right]^r_{ab} = \sum_{h,h'\in G} r^*_{ab}(h'h)\phi(h')\psi(h)$$

$$= \sum_{h,h'\in G} \sum_{c=1}^{\dim r} r^*_{ac}(h')r^*_{cb}(h)\phi(h')\psi(h)$$

The sum factors into two separate sums:

$$= \sum_{c=1}^{\dim r} \left\{ \sum_{h'\in G} r^*_{ac}(h')\phi(h') \right\} \left\{ \sum_{h\in G} r^*_{cb}(h)\psi(h) \right\}$$

Thus

$$\left[\widetilde{\phi * \psi}\right]^r_{ab} = \sum_{c=1}^{\dim r} \tilde{\phi}^r_{ac}\tilde{\psi}^r_{cb}$$

Thus within each irreducible representation, the convolution reduces to a matrix product. There is no mixing of different irreducible representations. This gives us a new perspective on the harmonic analysis of function on a group:

Theorem 53. *The convolution algebra on $l^2(G)$ decomposes into a direct sum of matrix algebras labelled by irreducible representations*

This explains in a natural why we have $(\dim r)^2$ components for each $r \in \tilde{G}$: that is the number of independent components of a matrix in the representation space. Thinking of the components as matrices also captures the non-commutativity of the convolution product.

12.3. Central Functions

Of special importance are functions satisfying the condition

$$\xi(g) = \xi(hgh^{-1}), \forall g, h \in G$$

These are called central functions; the space of such functions will be denoted by $\mathcal{Z}(G)$. The reason for the name is the following proposition

Proposition 54. *Central Functions commute with all functions under the convolution: $\xi * \phi = \phi * \xi$ for all $\xi \in \mathcal{Z}(G), \phi \in C(G)$*

Proof. By replacing $g \mapsto h^{-1}g$, the definition of central function can also be written as

$$\xi(h^{-1}g) = \xi(gh^{-1}), \forall g, h \in G$$

Now, setting $h = gh'^{-1}$

$$\xi * \phi(g) = \sum_{h \in G} \xi(h)\phi(h^{-1}g) = \sum_{h' \in G} \xi(gh'^{-1})\phi(h')$$

Using the centrality of ξ,

$$\xi * \phi(g) = \sum_{h' \in G} \xi(h'^{-1}g)\phi(h') = \phi * \xi(g)$$

The obvious examples of central functions are characters of representations. It should not be surprising that □

Proposition 55. *The functions $\frac{1}{\sqrt{|G|}}\chi_r$, $r \in \tilde{G}$ provide an orthonormal basis for the space of central functions*

This is a corrolary of the Peter–Weyl Theorem: Just replace ϕ by ξ, put $a = b$ and sum over a. Moreover,

Proposition 56. *The Fourier components of a central function are multiples of the identity in each irreducible representation*

This is a simple consequence of Schur's Lemma. The commutativity of central functions fits nicely with this.

12.4. An Example: The Finite Heisenberg Group

Let $\Lambda \geq 2$ be an integer and $Z_\Lambda = \mathbb{Z}/\Lambda\mathbb{Z}$ the additive group of integers modulo Λ. On the set $Z_\Lambda \times Z_\Lambda \times Z_\Lambda$ define the product

$$(m, n, c)(m', n', c') = (m + m', n + n', c + c' - nm') \mod \Lambda$$

This is not commutative, but is associative. We will call it the finite Heisenberg group $Heis(Z_\Lambda)$. (A similar group can be built out of any finite abelian group). If Z_Λ is replaced by the additive group of real numbers, we get the Heisenberg group of quantum mechanics. Such finite "approximations" to quantum mechanics have become of much interest in the context of quantum computing.

12.4.1. *Generators for $Heis(Z_\Lambda)$*

Define

$$X = (1,0,0), \quad Y = (0,1,0), \quad Z = (0,0,1)$$

Then

$$XY = (1,1,0), \quad YX = (1,1,-1)$$

$$ZYX = (0,0,1)\,(1,1,-1) = (1,1,0)$$

so that

$$XY = ZYX$$

With a little bit of work, we can see that the Heisenberg group is generated by X, Y, Z:

$$Heis(Z_\Lambda) = \langle X, Y, Z \mid XY = ZYX, ZX = XZ, ZY = YZ, X^\Lambda = 1 = Y^\Lambda = Z^\Lambda \rangle$$

Remark 57. A more economic choice of generators is just X and Y: We can identify $Z = XYX^{-1}Y^{-1}$. We just have to add the relations that $XYX^{-1}Y^{-1}$ commutes with X and Y. But we find it more convenient to retain Z as a generator.

12.4.2. *Some automorphisms*

Clearly

$$X \mapsto Z^a X, \quad Y \mapsto Z^b Y, \quad a, b \in Z_\Lambda$$

leaves the commutation relations unchanged. These are inner automorphisms as

$$XYX^{-1} = ZY \implies X^b Y X^{-b} = Z^b Y$$

$$Y^{-1}XY = ZX \implies Y^{-a} X Y^a = Z^a X$$

We now turn to studying representations of the group. We will denote the matrix representing X by $\hat{X}$, and similarly for Y and Z.

Since it commutes with everything, Z is a multiple of the identity in any irreducible representation:

$$\hat{Z} = \zeta 1, \quad \zeta \in Z_\Lambda$$

Here we are viewing Z_Λ as the *multiplicative*, group of Λth roots of unity. Hopefully it is clear from the context whether we mean addition or multiplication as the group operation.

Then, there is an equivalence transformation

$$\hat{X} \mapsto \zeta^a \hat{X}, \quad \hat{y} \mapsto \zeta^b \hat{Y}$$

We can use this to simplify the phases of the representation matrices (see below).

It is possible to generalize this a bit. Now, consider the more general transformation

$$X' = X^{a_1} Y^{b_1}, \quad Y' = X^{a_2} Y^{b_2}, \quad a_1, a_2, b_1, b_2 \in Z_\Lambda$$

Then

$$X'Y' = X^{a_1} Y^{b_1} X^{a_2} Y^{b_2} = Z^{-b_1 a_2} X^{a_1+a_2} Y^{b_1+b_2}$$
$$Y'X' = X^{a_2} Y^{b_2} X^{a_1} Y^{b_1} = Z^{-b_2 a_1} X^{a_1+a_2} Y^{b_1+b_2}$$

So,

$$X'Y' = ZY'X' \iff b_2 a_1 - b_1 a_2 = 1 \mod \Lambda$$

This is analogous to a symplectic transformation in mechanics.In particular $X \mapsto Y^{-1}$, $Y \mapsto X$ is an automorphism: Just choose $a_1 = 0, b_1 = -1, a_2 = 1$, $b_2 = 0$.

However, these are outer automorphisms: They cannot be written as conjugations $X \mapsto gXg^{-1}, Y \mapsto gYg^{-1}$ for some $g \in Heis(Z_\Lambda)$.

Our next task is to determine all the irreducible representations of $Heis(Z_\Lambda)$. But before we do that it is useful to work out a simpler special case:

Exercise 58. Determine all the irreducible representations (up to equivalence) of $Heis(Z_\Lambda)$ when Λ is a prime number.

Solution

The representations of a finite group are unitary. Since $\hat{X}, \hat{Z}$ commute they have a simultaneous eigenvector ψ_0.

$$\hat{X}\psi_0 = \xi\psi_0, \quad \hat{Z}\psi_0 = \zeta\psi_0$$

Choose ψ_0 to have length one.

Since $\hat{Z}^\Lambda = 1$, we have $\zeta^\Lambda = 1$. Let us assume for now that ζ a primitive root Λth of unity. (This means that no smaller power of ζ is equal to one; and hence that any other root can be written as a power of ζ).

Since Λ is a prime, any root which is not equal to one is primitive:

$$\zeta^{\Lambda} = 1, \quad \zeta \neq 1.$$

This is one of the simplifications when Λ is prime. Since $\hat{X}^{\Lambda} = 1$, ξ is also a root of unity. There must be an $a \in Z_{\Lambda}$ such that $\xi = \zeta^a$.

Now,

$$\hat{X}\hat{Y}^k\psi_0 = \hat{Z}^k\hat{Y}^k\hat{X}\psi_0 = \xi\zeta^k\hat{Y}^k\psi_0$$

Thus $\hat{Y}^k\psi_0$ are also eigenvectors of $\hat{X}$ and $\hat{Z}$. The eigenvalues $\xi\zeta^k, k = 0, 1, \cdots \Lambda - 1$ are distinct, since ζ is primitive and Λ is prime. So, $\hat{Y}^k\psi_0$ are orthogonal to each other and each of length one. (We also use the fact that $\hat{Y}$ is unitary).

Thus, in the Λ dimensional space spanned by $Y^k\psi_0, k = 0, \ldots, \Lambda - 1$, $\hat{X}$ is a diagonal matrix:

$$\hat{X} = \xi\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & \zeta & 0 & \cdots & 0 \\ 0 & 0 & \zeta^2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \zeta^{\Lambda-1} \end{pmatrix}, \quad Z = \zeta 1_{\Lambda}$$

Also,

$$\hat{Y}\hat{Y}^k\psi_0 = \hat{Y}^{k+1}\psi_0, \quad k = 0, 1, \ldots \Lambda - 2$$
$$\hat{Y}\hat{Y}^{\Lambda-1}\psi_0 = \psi_0$$

So $\hat{Y}$ is a cyclic permutation:

$$\hat{Y} = \begin{pmatrix} 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}$$

We can now make the equivalence transformation

$$\hat{X} \mapsto \hat{Y}^{-a}\hat{X}\hat{Y}^a, \quad \xi = \zeta^a$$

to set $\xi = 1$. This simplifies our analysis: ξ is not a relevant parameter; it can be removed by an equivalence transformation.

To summarize, for each choice of a primitive root ζ we have a representation of dimension Λ:

$$\hat{X} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & \zeta & 0 & \cdots & 0 \\ 0 & 0 & \zeta^2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \zeta^{\Lambda-1} \end{pmatrix}, \quad \hat{Y} = \begin{pmatrix} 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix},$$

$$\hat{Z} = \zeta 1_\Lambda, \quad \zeta^\Lambda = 1, \zeta \neq 1$$

There are $\Lambda - 1$ such choices of ζ.

It remains to consider the case where $\zeta = 1$; i.e.,when Z is in the kernel.So these representations factor through the commutative group $Z_\Lambda \times Z_\Lambda = Heis(Z_\Lambda)/Z_\Lambda$ where we quotient out the center. The equivalence classes of irreducible representations of this group (its dual) is again $Z_\Lambda \times Z_\Lambda$: there are Λ^2 such representations, all of which are one dimensional:

$$\hat{X} = \xi, \quad \hat{Y} = \eta \quad \xi^\Lambda = 1 = \eta^\Lambda$$

This time we cannot remove ξ or η by equivalence transformations: All the representations are one dimensional and so conjugations have no effect.

Recall that if $\tilde{G}$ is the set of equivalence classes of irreducible representations of a group G,

$$\sum_{r \in \tilde{G}} (\dim r)^2 = |G|.$$

There are $\Lambda - 1$ choices of a primitive root ζ giving dimension Λ representations, and Λ^2 one dimensional representations. Summing over all the representations we have so far gets us

$$(\Lambda - 1)\Lambda^2 + \Lambda^2 = \Lambda^3.$$

This proves that, we have found all the irreducible representations of $Heis(Z_\Lambda)$ when Λ is prime.

Representations of this type are studied further in the delightful book by Terras [25].

12.4.3. *Irreducible representations of $Heis(Z_\Lambda)$*

Since Z is central, in any irreducible representation it will be a multiple of the identity:

$$\hat{Z} = \zeta 1_D$$

where D is the dimension of the representation. Since $Z^\Lambda = 1$ we must have $\zeta^\Lambda = 1$. Let d be the smallest number such that $\zeta^d = 1$; i.e., ζ is a primitive root of unity, of order d. Then d is a divisor of Λ. Now,

$$XY = ZYX \implies XY^d = Z^d Y^d X$$

Thus, $\hat{X}$ and $\hat{Y}^d$ commute with each other. So, $\hat{Y}^d$ commutes with all the representation matrices, and must be a multiple of the identity:

$$\hat{Y}^d = \eta 1_D$$

Since $\left(Y^d\right)^{\frac{\Lambda}{d}} = 1$ we must have

$$\eta^{\frac{\Lambda}{d}} = 1.$$

Let ψ_0 be an eigenvector of $\hat{X}$:

$$\hat{X}\psi_0 = \xi_1 \psi_0$$

Choose ψ_0 to be of unit length. Then $\hat{Y}^k\psi_0$ is also an eigenvector of $\hat{X}$:

$$\hat{X}(\hat{Y}^k\psi_0) = \zeta^k \hat{Y}^k \hat{X}\psi_0 = \xi_1 \zeta^k (\hat{Y}^k \psi_0)$$

The eigenvalues $\xi_1\zeta^k$ are distinct from each other for $k = 0, 1, \ldots, d-1$ (since ζ is primitive root of order d). So $\psi_k \equiv \hat{Y}^k\psi_0$ are mutually orthogonal for $k = 0, 1, \ldots, d-1$. They are all also of length one as $\hat{Y}$ is unitary. Thus there is a vector space V of dimension d for with $\hat{Y}^k\psi$ is an orthonormal basis.

In this basis

$$\hat{Y}\psi_k = \psi_{k+1} \quad k = 0, 1, \ldots, d-2$$

and

$$\hat{Y}\psi_{d-1} = \eta\psi_0$$

since $\hat{Y}\psi_{d-1} = \hat{Y}^d\psi_0$. As matrices

$$\hat{X} = \xi_1 \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & \zeta & 0 & \cdots & 0 \\ 0 & 0 & \zeta^2 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdots & 0 \\ 0 & 0 & 0 & \cdots & \zeta^{d-1} \end{pmatrix}, \quad \hat{Y} = \begin{pmatrix} 0 & 0 & 0 & \cdots & \eta \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix},$$

$$\hat{Z} = \zeta \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

Since only multiples of identity commute with $\hat{X}$ and $\hat{Y}$ it is an irreducible representation of dimension d, the chosen divisor of Λ.

Since

$$\hat{Y}^{-a}\hat{X}\hat{Y}^{a} = \zeta^{a}\hat{X},$$

we can transform $\xi \mapsto \xi\zeta^{a}$ by an equivalence transformation. Two different representations with parameters ξ_1 and ξ_1' which are related by $\xi_1' = \zeta^{a}\xi_1$ are equivalent for $a \in Z_d$. The equivalence class is labelled by $Z_\Lambda/Z_d = Z_{\frac{\Lambda}{d}}$. The projection $Z_\Lambda \to Z_{\frac{\Lambda}{d}}$ is given by ξ_1^d. In other words, two representations are equivalent if $\xi_1^d = \xi_1'^d$.Let us denote $\xi = \xi_1^d$.

To summarize, the representation we found is determined by three parameters[7] (ξ, η, ζ) with[8]

$$\xi, \eta \in Z_{\frac{\Lambda}{d}}, \quad d \mid \Lambda$$

and ζ is a primitive root of order d.

Recall that for any finite group G

$$|G| = \sum_{r \in \tilde{G}} (\dim\ r)^2$$

where $\tilde{G}$ is the set of equivalence classes of irreducible representations. If we form this sum for the representations we have found,

$$\sum_{d|\Lambda} \sum_{\xi,\eta \in Z_{\frac{\Lambda}{d}}} \sum_{\zeta} d^2 = \sum_{d|\Lambda} \frac{\Lambda}{d} \times \frac{\Lambda}{d} \times \tau(d) \times d^2$$

The factors of $\frac{\Lambda}{d}$ account for the choices of ξ and η. Also, $\tau(d)$ is the number of primitive roots ζ of unity of order d. This is equal to the number $\tau(d)$ of integers co-prime to d, called the Euler totient function. So,

$$\sum_{d|\Lambda} \sum_{\xi,\eta \in Z_{\frac{\Lambda}{d}}} \sum_{\zeta} d^2 = \Lambda^2 \sum_{d|\Lambda} \tau(d)$$

Now, it is a well-known identity[9] that

$$\sum_{d|\Lambda} \tau(d) = \Lambda$$

[7] No need to list d separately, as it is determined by ζ: it is the smallest number such that $\zeta^{d} = 1$.

[8] We use a notation common in number theory: $d \mid \Lambda$ means that d is a divisor of Λ. That is, $\frac{\Lambda}{d}$ is an integer.

[9] The proof is elementary. The rhs is the number of roots of unity of order Λ; we can partition them into sets of primitive roots of order d, for each divisor of Λ. The number of elements in each set is $\tau(d)$.

Thus,

$$\sum_{\xi,\eta\in Z_{\frac{\Lambda}{d}}}\sum_{\zeta} d^2 = \Lambda^3$$

Thus, we have found all the irreducible representations of $Heis(Z_\Lambda)$.

Remark 59. In particular, only the representations where ζ is a primitive Λth root of unity are **faithful**. There are $\tau(\Lambda)$ such inequivalent representations, each of dimension Λ. In this case, ξ and η are both equal to one: The choice of ζ determines the representation up to equivalence. This is the finite analogue of the Stone–von Neumann theorem for the continuous Heisenberg group of quantum mechanics: The analogue of ζ is the choice of a value of Planck's constant. Given this, there is a unique equivalence class of faithful irreducible representations for that Heisenberg group. The discrete Heisenberg group we studied arises in approximating quantum systems using "qudit" systems; there are hopes that quantum computers can be built out of these.

12.4.4. *Characters*

Let us compute the character of the representation labelled by (ξ,η,ζ) .

We need the sum of diagonal elements of $\hat{X}^m\hat{Y}^n$. But $\hat{Y}^n$ has diagonal elements only if $d \mid n$. Then

$$\hat{X}^m\hat{Y}^n\hat{Z}^c = \zeta^c\xi_1^m\eta^{\frac{n}{d}}\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & \zeta^m & 0 & \cdots & 0 \\ 0 & 0 & \zeta^{2m} & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdots & 0 \\ 0 & 0 & 0 & \cdots & \zeta^{m(d-1)} \end{pmatrix}$$

Recalling that $\zeta^d = 1$, the geometric series

$$1 + \zeta^m + \zeta^{2m} + \cdots \zeta^{m(d-1)} = \frac{1-\zeta^{md}}{1-\zeta^m}$$

vanishes if $d \nmid m$; if $d \mid m$ each terms is equal to one so it sums to d. So, the character is

$$\chi_{(\xi,\eta,\zeta)}(m,n,c) = \mathrm{tr}\hat{X}^m\hat{Y}^n\hat{Z}^c = d\zeta^c\xi_1^m\eta^{\frac{n}{d}}\delta(d \mid m)\delta(d \mid n)$$

Since $\xi_1^d = \xi$ we can write this as

$$\chi_{(\xi,\eta,\zeta)}(m,n,c) = d\zeta^c\xi^{\frac{m}{d}}\eta^{\frac{n}{d}}\delta(d \mid m)\delta(d \mid n)$$

As expected the answer is symmetric under the interchange $m \leftrightarrow n, \xi \leftrightarrow \eta$.

12.4.4.1. *Orthonormality of characters*

Recall the general formula

$$\langle \chi_r, \chi_s \rangle \equiv \sum_{h \in G} \chi_r^*(h)\chi_s(h) = |G|\delta_{rs}, \quad r, s \in \tilde{G}$$

Let us verify this by calculating

$$\langle \mathcal{X}^*_{(\xi,\eta,\zeta)}, \mathcal{X}_{(\xi,'\eta',\zeta')} \rangle = \sum_{m,n,c \in Z_\Lambda} d\zeta^{*c} \xi^{*\frac{m}{d}} \eta^{*\frac{n}{d}} \delta(d \mid m)\delta(d \mid n) d'\zeta'^{c}$$

$$\times \xi'^{\frac{m}{d'}} \eta'^{\frac{n}{d'}} \delta(d' \mid m)\delta(d' \mid n)$$

The sums factorize

$$\langle \mathcal{X}^*_{(\xi,\eta,\zeta)}, \mathcal{X}_{(\xi,'\eta',\zeta')} \rangle = dd' \sum_{m \in Z_\Lambda} \xi^{*\frac{m}{d}} \xi'^{\frac{m}{d'}} \delta(d \mid m)\delta(d' \mid m)$$

$$\times \sum_{n \in Z_\Lambda} \eta^{*\frac{n}{d}} \eta'^{\frac{n}{d'}} \delta(d \mid n)\delta(d' \mid n) \sum_{c \in Z_\Lambda} \zeta^{*c}\zeta'^{c}$$

Since ζ, ζ' are Λthroots of unity we know that

$$\sum_{c \in Z_\Lambda} \zeta^{*c}\zeta'^{c} = \Lambda\delta_{\zeta\zeta'}$$

Since d is the smallest number such that $\zeta^d = 1$ and similarly d' is for ζ', it follows that when $\zeta = \zeta'$, we have also $d = d'$. Then we have

$$\delta(d \mid m)\delta(d' \mid m = \delta(d \mid m)$$

The sum over m becomes, after the change of variables $\mu = \frac{m}{d}$, (remembering that ξ, ξ' are roots of unity of order $\frac{\Lambda}{d}$)

$$\sum_{m \in Z_\Lambda} \xi^{*\frac{m}{d}} \xi'^{\frac{m}{d}} \delta(d \mid m) = \sum_{\mu \in Z_{\frac{\Lambda}{d}}} \xi^{*\mu}\xi'^{\mu} = \frac{\Lambda}{d}\delta_{\xi,\xi'}$$

and similarly for the sum over n. Thus,

$$\langle \mathcal{X}^*_{(\xi,\eta,\zeta)}, \mathcal{X}_{(\xi,'\eta',\zeta')} \rangle = d^2 \frac{\Lambda}{d}\delta_{\xi,\xi'} \frac{\Lambda}{d}\delta_{\eta,\eta'}\Lambda\delta_{\zeta\zeta'} = \Lambda^3\delta_{\xi,\xi'}\delta_{\eta,\eta'}\delta_{\zeta\zeta'}.$$

as was to be proved (since $|G| = \Lambda^3$).

For a beautiful application of this representation theory to understand random walks on the Heisenberg group, see [26].

This page intentionally left blank

Chapter 13

HARMONIC ANALYSIS ON COMPACT LIE GROUPS

13.1. Compact?

The abstract definition of compactness is somewhat obscure and not of direct utility to us.

Definition 60. An open cover of a topological space X is a family Φ of open subsets such that $X = \bigcup_{S \in \Phi} S$. If every open cover has a finite subcover $F \subseteq \Phi$ we say that X is compact.

There are many familiar examples, though:

- Any finite set is compact
- The real line is not compact. But a subset of it is compact iff it is closed bounded
- Vector spaces are not compact. But for *finite dimensional* vector spaces, a subset is compact iff it is closed bounded
- In infinite dimensional vector spaces, there are closed bounded subsets that are not compact. For example the unit ball in a Hilbert space is not compact
- An ellipsoid whose principal axes is a sequence that converges to zero is a compact subset of an infinite dimensional Hilbert space
- A closed subset of a compact set is again compact
- Many familiar shapes such as a circle or a sphere are therefore compact spaces

It is useful to understand compactness in terms of its implication for continuous functions. A continuous function on a compact space is bounded; moreover it achieves its lowest upper bound.In many situations, compactness is a substitute for finiteness. This is true of groups: Much of the theory of finite groups can be generalized to compact Lie groups.

13.1.1. *Compact Lie groups?*

When it comes to groups,

$$U(n), SU(n), O(n)$$

are compact for any n. Examples of non-compact Lie groups are

- The real line
- The Lorentz group $O(1,3)$
- $GL(n), SL_n(R), SL_n(C)$

The point is that in these groups we can "off to infinity" along a one-parameter subgroup.

13.2. Non-Compact Lie Groups

Before we get going, let us note an important fact about non-compact groups:

Theorem 61. *A non-compact Lie group has no faithful*[1] *finite dimensional unitary representation*

The point is that a continuous map cannot take a non-compact set to a compact set. The finite dimensional unitary group $U(n)$ is compact. So it cannot contain the image of a faithful representation of a non-compact Lie group. So, unitary representations of non-compact Lie groups are infinite dimensional and so are much more subtle. The mathematics, pioneered by Gelfand and Bargmann and developed fully by Harish-Chandra, is exquisite. But has not found much application in physics as yet, apart from the work of Wigner on the Poincare' group. There are many excellent discussions of this, including Wigner's original papers.

13.3. A Tale of Two Hilbert Spaces: $l^2(\mathbb{Z})$ and $L^2(U(1))$

We begin with the simplest case of a compact Lie group, $U(1)$.

A Hilbert space is a vector space with an inner product (hence a norm) that satisfies two additional conditions:

- It is complete; i.e., every Cauchy sequence of vectors $v_k \in \mathcal{H}$ has a limit which is also in $\mathcal{H}$
- It has a countable orthonormal basis e_m

[1] Faithful means that the kernel is trivial: the representation matrix $R(g)$ determines g.

Obviously any finite dimensional vector space with positive inner product satisfies these conditions. So the interesting cases are infinite dimensional.

13.3.1. $l^2(\mathbb{Z})$

The additive group of integers yields our first example of an infinite dimensional Hilbert space. A function $\tilde{\phi} : \mathbb{Z} \to \mathbb{C}$ is simply a sequence $\tilde{\phi}_m$ of numbers indexed by the integers. Define the space $l^2(\mathbb{Z})$ to be the set of all square-summable sequences:

$$l^2(\mathbb{Z}) = \left\{ \tilde{\phi} : \mathbb{Z} \to \mathbb{C} \middle| \sum_{m\in\mathbb{Z}} \middle| \tilde{\phi}_m|^2 < \infty \right\}$$

If the function vanish fast enough for large $|m|$ this sum will converge. We define the norm $||\tilde{\phi}||$ of $\tilde{\phi} \in l^2(\mathbb{Z})$ by

$$||\tilde{\phi}||^2 = \sum_{m\in\mathbb{Z}} |\tilde{\phi}_m|^2$$

For two functions in $l^2(\mathbb{Z})$ we can define an inner product

$$\langle \tilde{\phi}, \tilde{\psi} \rangle = \sum_{m\in\mathbb{Z}} \tilde{\phi}^*_m \tilde{\psi}_m$$

This will converge because of the Schwarz inequality:

$$| \langle \tilde{\phi}, \tilde{\psi} \rangle | \leq ||\tilde{\phi}|| \, ||\tilde{\psi}||$$

It is not hard to show that $l^2(\mathbb{Z})$ is indeed complete: That is, Cauchy sequences converge to a limit in $l^2(\mathbb{Z})$. And that there is a countable orthonormal basis. For example the functions δ_m concentrated at each point of $\mathbb{Z}$:

$$\delta_m(n) = \begin{cases} 1 & m = n \\ 0 & m \neq n \end{cases}$$

is such a basis.

13.3.2. $L^2(U(1))$

The simplest compact Lie group is $U(1)$; also called the circle group.Functions $\phi : U(1) \to \mathbb{R}$ are simply periodic functions of period 2π of the angular coordinate θ.

The space of continuous periodic functions admit an inner product

$$\langle \phi, \psi \rangle = \int_{-\pi}^{\pi} \phi^*(\theta)\psi(\theta)\frac{d\theta}{2\pi}$$

This inner product is much like the dot product in Euclidean space; instead of a sum over products of components, we have an integral. This allows us to define a metric; i.e., a notion of distance between two continuous functions.

$$||\phi - \psi|| = \sqrt{\int_{-\pi}^{\pi} \mid \phi(\theta) - \psi(\theta) \mid^2 \frac{d\theta}{2\pi}}$$

Alas, the space of continuous functions is **not complete** in this metric. Here is an example:

Exercise 62. Define the "pyramid" function

$$f(\theta) = \begin{cases} \sqrt{3\pi}\,(1 - |\theta|) & |\theta| < 1 \\ 0 & \text{otherwise} \end{cases}$$

It is continuous on the interval $[-\pi, \pi]$. (It can be extended to a continuous periodic function of period 2π on the real line.)Show that the sequence $f_k(\theta) = \sqrt{k} f(k\theta), \quad k = 1, 2, \ldots$ is a Cauchy sequence; i.e., that $||f_{k+1} - f_k|| \to 0$ as $k \to \infty$. But that the limit is not itself a continuous function.

Solution

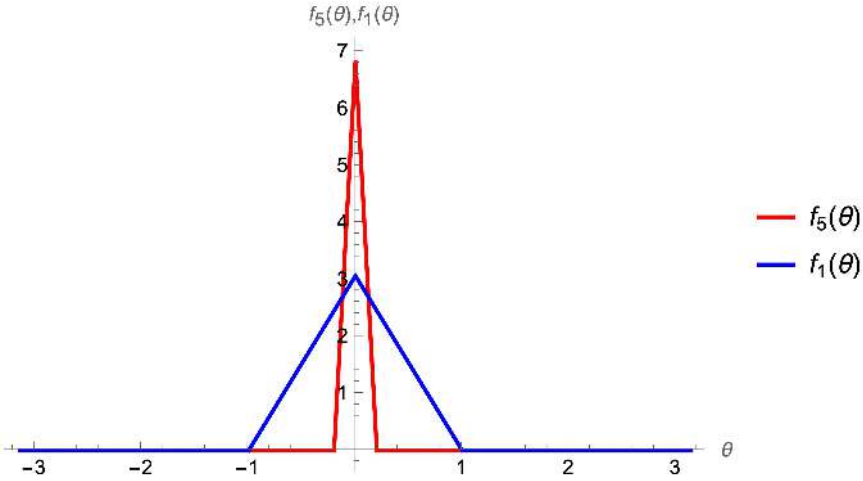
The constant $\sqrt{3\pi}$ and the factor $\sqrt{k}$ are chosen so that $||f_k|| = 1$. (Verify by calculating the integral using Mathematica).

The integral $\int_{-\pi}^{\pi} (f_{k+1}(\theta) - f_k(\theta))^2 \frac{d\theta}{2\pi}$ can be evaluated using Mathematica to get a complicated formula for $||f_{k+1} - f_k||$. We need only the limiting behavior

$$||f_{k+1} - f_k|| = \frac{\sqrt{5}}{2\sqrt[4]{3\pi}} k^{-\frac{3}{4}} + \mathrm{O}(k^{-1})$$

which tends to zero. The limit of $f_k(\theta)$ is not a continuous function. For example, $f_k(0) = \sqrt{3\pi k}$ tends to infinity. In fact the functions become more peaked at the origin and the support tends to a vanishingly small interval of length $\frac{2}{k}$ for large k. See the figure.

This leads to some technical complications which have been largely resolved by mathematicians. The field of functional analysis is about this and related matters. We will not delve into these matters much. We will just mention (without proof



or even a precise statement sometimes) some of the ideas involved; and give some examples.

13.3.2.1. $L^2(U(1))$ *as a completion*

Recall that a Cauchy sequence in a metric space is a sequence which approach each other in distance: $||\phi_{k+1} - \phi_k||$ is as small as you want for large enough k. In this language, what we are saying is that the space of continuous functions is not complete. This phenomenon also occurs in the space of rational numbers $\mathbb{Q}$: There are Cauchy sequences of rational numbers whose limit is not rational. An example is the sequence of decimal approximations for $\sqrt{2}$:

$$1, 1.4, 1.41, 1.414, \ldots$$

We can remedy this situation by enlarging $\mathbb{Q}$ by passing to space of equivalence classes of Cauchy sequences of rationals. The notion of equivalence is this: Given two sequences ϕ_k and ψ_k, we can make a new sequence by interlacing them:

$$\phi_1, \psi_1, \phi_2, \psi_2, \ldots$$

If this interlaced sequence is also Cauchy, we say they are equivalent. The idea is that equivalent sequences "tend to the same limit". The set of such equivalence classes is again a metric space; but one that is complete. This is one way to construct real numbers. It is just a precise way of saying that real numbers are those that can be approximated as close as you wish by sequences of rational numbers.

A similar process can be put through for smooth functions on the circle with the above notion of metric. $L^2(U(1))$ is the completion of the space of continuous

functions by in the metric $||\phi - \psi||$. A typical element of $L^2(U(1))$ is **not** a continuous function on $U(1)$: But it can be approximated by continuous functions as closely as we want.

13.3.3. *An orthonormal set in $L^2(U(1))$*

Examples of continuous functions on the circle are exponentials such as

$$\ldots, e^{-2i\theta}, e^{-i\theta}, 1, e^{i\theta}, e^{2i\theta}, \ldots$$

We will denote them by e_m:

$$e_m(\theta) = e^{im\theta}, \quad m \in \mathbb{Z}.$$

Proposition 63. *The exponential functions form an orthonormal set*:

$$\int_{-\pi}^{\pi} e_m^*(\theta) e_n(\theta) \frac{d\theta}{2\pi} = \delta_{m,n}, \quad m, n \in \mathbb{Z}.$$

Proof. When $m = n$ this is obvious, as the exponentials in the integrand cancel and the statement reduces to $\int_{-\pi}^{\pi} \frac{d\theta}{2\pi} = 1$. When $m \neq n$ the lhs is equal to

$$\int_{-\pi}^{\pi} e^{i(n-m)\theta} \frac{d\theta}{2\pi} = I_{n-m} = \left[\frac{e^{i(n-m)\theta}}{i(n-m)} \right]_{\theta=-\pi}^{\theta=\pi}.$$

This is zero since the exponential has the same value at $\theta = \pi$ and $\theta = -\pi$. □

13.3.4. *The duality of $L^2(U(1))$ and $l^2(\mathbb{Z})$*

Given any element of $L^2(U(1))$ we can find its components along the exponentials:

$$\tilde{\phi}_m = \langle e_m, \phi \rangle = \int e_m^*(\theta) \phi(\theta) \frac{d\theta}{2\pi}$$

The integral converges because of the Schwarz inequality.

It is fair to ask if these components can be used to reconstruct the original function from these components. We now state a few results which allow this reconstruction. The proofs can be skipped in a first reading. The proofs of these results (under slightly weaker assumptions) are detailed in the classic book [27]. They are summarized in subsections below.

From orthonormality we can conclude already the following:

Proposition 64. ***Bessel's inequality***

$$\sum_{m \in \mathbb{Z}} |\tilde{\phi}_m|^2 \leq \int_{U(1)} |\phi(\theta)|^2 \frac{d\theta}{2\pi}$$

The remaining question is whether there is some information about ϕ missing in the components $\tilde{\phi}_m$. Can we prove

$\phi(\theta) = \sum_{m\in\mathbb{Z}} \tilde{\phi}_m e^{im\theta}$? A systematic approach to studying this infinite sum is to start with the partial sums

$$S_\Lambda = \sum_{|m|\leq\Lambda} \tilde{\phi}_m e_m$$

and study the limit $\Lambda \to \infty$. It turns out that the sequence $S_\Lambda(\theta)$ may **not** converge to $\phi(\theta)$ for individual values of θ, for $\phi \in L^2(U(1))$. But we don't need that. We do have convergence in the norm of $L^2(U(1))$:

Proposition 65. ***Convergence in*** $L^2(U(1))$. *If ϕ is a continuous function on the circle,* $\lim_{\Lambda\to\infty} ||\phi - S_\Lambda|| = 0$

This guarantees that $e_m, m \in \mathbb{Z}$ is a basis for $L^2(U(1))$; the components $\tilde{\phi}_m$ contain all the information contained in the element ϕ of $L^2(U(1))$ at least for continuous functions. And since any element of $L^2(U(1))$ can be approximated by continuous functions, this extends to all of $L^2(U(1))$. So, $L^2(U(1))$ is a Hilbert space.

Theorem 66. *The set $e_m, m \in \mathbb{Z}$ is an orthonormal basis in $L^2(U(1))$*

Moreover,

Theorem 67. ***Plancherel Formula a.k.a. Parseval identity***

$$\sum_{m\in\mathbb{Z}} |\ \tilde{\phi}_m|^2 = \int_{U(1)} |\ \phi(\theta)|^2 \frac{d\theta}{2\pi}$$

This says that Fourier transform is a unitary map from $L^2(U(1))$ to $l^2(\mathbb{Z})$. In some applications (signal processing) the rhs has the meaning of energy; then this says that all the energy can be accounted for by adding up the energies of the Fourier components. In quantum mechanics, probability-rather than energy-is the meaning of the rhs.

13.3.5. *Proof of Bessel's inequality*

Given any orthonormal set (not necessarily a basis) e_m we can find the components $\tilde{\phi}_m = \langle e_m, \phi\rangle = \int e_m^*(\theta)\phi(\theta)\frac{d\theta}{2\pi}$ and form the series $S = \sum_{m\in\mathbb{Z}} \tilde{\phi}_m e_m$. It follows

that

$$\langle \phi, S\rangle = \sum_{m\in\mathbb{Z}} \mid \tilde{\phi}_m|^2$$

Also,

$$||S||^2 = \sum_{m,n} \tilde{\phi}^*_m \tilde{\phi}_n \langle e_m, e_n\rangle = \sum_m \mid \tilde{\phi}_m|^2$$

which is a version of Pythagoras' theorem. Therefore

$$\langle S, \phi - S\rangle = 0.$$

Thus

$$\phi = S + (\phi - S)$$

is an orthogonal decomposition. By Pythagoras,

$$||\phi||^2 = ||S||^2 + ||\phi - S||^2 \geq ||S||^2$$

This is Bessel's inequality.

13.3.6. *Proof of Convergence in $L^2(U(1))$*

This relies on a basic result of analysis (a version of the Weierstrass approximation theorem) that continuous functions on a circle can be approximated by trigonometric polynomials (i.e., finite series of the form $P(\theta) = \sum_m \tilde{P}_m e^{im\theta}$) as closely as desired:

For any $\epsilon > 0$ there is a trig polynomial $P(\theta)$ such that

$$|\phi(\theta) - P(\theta)| < \epsilon$$

for all θ. Let M be the degree (the largest value of $|m|$ for which $\tilde{P}_m \neq 0$) of this polynomial. Integrating the square of the above inequality, and then taking square root, we get

$$||\phi - P|| < \epsilon.$$

Another fact we need is that

$$\langle e_m, \phi - S_\Lambda\rangle = 0, \quad |m| \leq \Lambda$$

(It follows from $\langle e_m, \phi\rangle = \tilde{\phi}_m$ and $S_\Lambda = \sum_{|m|\leq\Lambda} \tilde{\phi}_m e_m$). In other words, $\phi - S_\Lambda$ is orthogonal to the subspace of trig polynomials of degree $\leq \Lambda$. If we choose

$\Lambda \geq M$, the approximation P, will lie in this subspace. Since $S_\Lambda - P$ is orthogonal to $\phi - S_\Lambda$, and $\phi - P = (\phi - S_\Lambda) + (S_\Lambda - P)$, Pythagoras says

$$||\phi - P||^2 = ||\phi - S_\Lambda||^2 + ||S_\Lambda - P||^2$$

Therefore,

$$||\phi - S_\Lambda|| \leq ||\phi - P|| < \epsilon$$

Thus, given any $\epsilon > 0$, we can find a Λ such that $||\phi - S_\Lambda|| < \epsilon$. This is the convergence we seek.

13.3.7. *Proof of Parseval's identity*

This is a consequence of the above convergence. The orthogonal decomposition

$$\phi = S_\Lambda + (\phi - S_\Lambda)$$

gives, by Pythagoras,

$$||\phi||^2 = ||S_\Lambda||^2 + ||\phi - S_\Lambda||^2$$

As $\Lambda \longrightarrow \infty$ the second term tends to zero; and the first term tends to $\sum_m |\phi_m|^2$.

Exercise 68. For the "pyramid" function

$$f(\theta) = \begin{cases} \sqrt{3\pi}\,(1 - |\theta|) & |\theta| < 1 \\ 0 & \text{otherwise} \end{cases}$$

show that

$$\langle e_m, f \rangle = 2\sqrt{\frac{3}{\pi}} \left(\frac{\sin \frac{m}{2}}{m} \right)^2 .$$

13.3.8. *Convolution*

Define a multiplication operation on continuous functions

$$\phi * \psi(\theta) = \int \phi(\theta - \theta')\psi(\theta') \frac{d\theta'}{2\pi}$$

called **convolution**. This is commutative:

$$\psi * \phi(\theta) = \int \psi(\theta - \theta')\phi(\theta') \frac{d\theta'}{2\pi}$$

Making the change of variables $\theta' \mapsto \theta - \theta'$ we see that it is equal to $\phi * \psi(\theta)$.

Proposition 69. *The Fourier components of a convolution is the product of Fourier components. That is,*

$$\int e^{-im\theta}\phi * \psi(\theta)\frac{d\theta}{2\pi} = \tilde{\phi}_m\tilde{\psi}_m$$

The proof is simply to evaluate the double integral:

$$\int e^{-im\theta}\phi * \psi(\theta)\frac{d\theta}{2\pi} = \int e^{-im\theta}\phi(\theta-\theta')\psi(\theta')\frac{d\theta}{2\pi}\frac{d\theta'}{2\pi}$$

We change variables $\theta \mapsto \theta + \theta'$:

$$= \int e^{-im[\theta+\theta']}\phi(\theta)\psi(\theta')\frac{d\theta}{2\pi}\frac{d\theta'}{2\pi}$$

and notice that the integral factorizes as the product of two integrals:

$$= \int e^{-im\theta}\phi(\theta)\frac{d\theta}{2\pi}\int e^{-im\theta'}\psi(\theta')\frac{d\theta'}{2\pi} = \tilde{\phi}_m\tilde{\psi}_m$$

Fourier developed his analysis to solve partial differential equations arising from engineering. In this context, the following exercise is interesting.

Exercise 70. Solve the Laplace equation in the unit disc, given the boundary value at the unit circle: In polar co-ordinates

$$\frac{\partial^2}{\partial r^2}\Phi(r,\theta) + \frac{1}{r}\frac{\partial}{\partial r}\Phi(r,\theta) + \frac{1}{r^2}\frac{\partial^2}{\partial\theta^2}\Phi(r,\theta) = 0, \quad \lim_{r\to 1^-}\Phi(r,\theta) = \phi(\theta)$$

Solution

If the solution is continuous inside the disc, $\lim_{r\to 0^+}\Phi(r,\theta)$ must be independent of θ(since all values of θ describe the origin when $r = 0$). We can choose this value to be zero without any loss of generality (a constant can always be added to the electrostatic potential).

Fourier analysis

$$\Phi(r,\theta) = \sum_{n\in\mathbb{Z}}\tilde{\Phi}_n(r)e^{in\theta}, \quad \phi(\theta) = \sum_{n\in\mathbb{Z}}\tilde{\phi}_n e^{in\theta},$$

gives

$$\frac{d^2}{dr^2}\tilde{\Phi}_n(r) + \frac{1}{r}\frac{d}{dr}\tilde{\Phi}_n(r) - \frac{n^2}{r^2}\tilde{\Phi}_n(r) = 0, \quad \lim_{r\to 1^-}\tilde{\Phi}_n(r) = \tilde{\phi}_n, \quad \lim_{r\to 0^+}\tilde{\Phi}_n(r) = 0.$$

The solution is $\tilde{\Phi}_n(r) = r^{|n|}\tilde{\phi}_n$. Thus

$$\Phi(r,\theta) = \sum_{n\in\mathbb{Z}} r^{|n|}\tilde{\phi}_n e^{in\theta}$$

Equivalently we have the convolution

$$\Phi(r,\theta) = \int P_r(\theta - \theta')\phi(\theta')\frac{d\theta'}{2\pi}$$

where the Poisson kernel P_r is defined by

$$P_r(\theta) = \sum_{m\in\mathbb{Z}} r^{|m|}e^{im\theta}$$

Summing the geometric series for $m \geq 0$ and $m < 0$ and recombining we get the explicit formula

$$P_r(\theta) = \frac{1-r^2}{1-2r\cos\theta + r^2}$$

Note that in this case the solution in the interior is smoother than the boundary data on the unit circle: The factor $r^{|n|}$ in the sum $\sum_{n\in\mathbb{Z}} r^{|n|}\tilde{\phi}_n e^{in\theta}$ provides a nice convergence factor. Incidentally, we can deduce the mean value property of solutions of the Laplace equation:

$$\int \Phi(r,\theta)\frac{d\theta}{2\pi}$$

is independent of r. If we insert the series $\sum_{n\in\mathbb{Z}} r^{|n|}\tilde{\phi}_n e^{in\theta}$ into the integral, only the term $n = 0$ survives.

Fourier Analysis also gives a formula with far reaching applications in number theory and physics:

Exercise 71. Poisson Sum Formula Let $f : \mathbb{R} \to \mathbb{C}$ be a (necessarily not periodic) function on the real line that vanishes at infinity along with all of its derivatives. Define $\tilde{f}(\xi) = \int f(x)e^{-2\pi i\xi x}d\xi$. Then

$$\sum_{n\in\mathbb{Z}} f(n) = \sum_{n\in\mathbb{Z}} \tilde{f}(n).$$

Apply to the special case $f(x) = e^{-tx^2}$ with $t > 0$.

Solution

The trick is to build a periodic function by averaging f:

$$\phi(\theta) = \sum_{n\in\mathbb{Z}} f\left(\frac{\theta}{2\pi} + n\right)$$

The sum converges because f and all its derivatives vanish at infinity.Clearly

$$\phi(0) = \sum_n f(n)$$

On the other hand we know from Fourier synthesis that

$$\phi(0) = \sum_{m\in\mathbb{Z}} \tilde{\phi}_m$$

where

$$\tilde{\phi}_m = \int e^{-i\theta m}\phi(\theta)\frac{d\theta}{2\pi}$$

The Poisson sum formula is proved if we can show that $\tilde{f}(m) = \tilde{\phi}(m)$.

We can calculate

$$\tilde{\phi}_m = \int_{-\pi}^{\pi} e^{-i\theta m} \sum_{n\in\mathbb{Z}} f\left(\frac{\theta}{2\pi} + n\right)\frac{d\theta}{2\pi}$$

make the change of variables $x = \frac{\theta}{2\pi} + n$

$$\tilde{\phi}_m = \sum_{n\in\mathbb{Z}} \int_{-\frac{1}{2}+n}^{\frac{1}{2}+n} e^{-i2\pi[x-n]m} f\,(x)\,dx$$

Since $e^{2\pi i[nm]} = 1$

$$\begin{aligned}\tilde{\phi}_m &= \sum_{n\in\mathbb{Z}} \int_{-\frac{1}{2}+n}^{\frac{1}{2}+n} e^{-2\pi imx} f\,(x)\,dx \\ &= \left\{\cdots + \int_{-\frac{1}{2}-1}^{\frac{1}{2}-1} + \int_{-\frac{1}{2}}^{\frac{1}{2}} + \int_{-\frac{1}{2}+1}^{\frac{1}{2}+1} + \cdots\right\} e^{-2\pi imx} f\,(x)\,dx \\ &= \int_{-\infty}^{\infty} e^{-2\pi imx} f\,(x)\,dx = \tilde{f}(m)\end{aligned}$$

as needed.

If $f(x) = e^{-tx^2}$ with $t > 0$, the Fourier transform can be calculated using simple methods (completing the square in the exponent):

$$\tilde{f}(\xi) = \frac{\sqrt{\pi}}{\sqrt{t}} e^{-\frac{\pi^2\xi^2}{t}}.$$

Thus

$$\sum_{n\in\mathbb{Z}} e^{-tn^2} = \frac{\sqrt{\pi}}{\sqrt{t}} \sum_{n\in\mathbb{Z}} e^{-\frac{\pi^2 n^2}{t}}$$

The point of this formula is that the lhs converges fast for large t while the rhs does so for small t. This leads to an inversion symmetry for the elliptic modular function $\vartheta(t) = \sum_{n\in\mathbb{Z}} e^{-tn^2}$:

$$\vartheta(t) = \sqrt{\frac{\pi}{t}}\vartheta\left(\frac{\pi^2}{t}\right)$$

This function arises in the study of lattices and in string theory.

13.4. Invariant Integrals on Lie Groups

Our task is to generalize Fourier analysis to compact Lie groups. In other words, replace $U(1)$ by a possibly non-abelian Lie group in the theory of the last section. Alternatively, we want to generalize the harmonic analysis on finite non-abelian groups (the last chapter) to compact Lie groups. An essential tool of that chapter were sums of the type

$$\sum_{g\in G} \phi(g)$$

These sums are invariant under the left or right action of the group:

$$\sum_{g\in G} L_h\phi(g) = \sum_{h\in G} \phi\left(h^{-1}g\right) = \sum_g \phi(g)$$

$$\sum_{g\in G} R_h\phi(g) = \sum_{h\in G} \phi\left(gh\right) = \sum_g \phi(g)$$

It turns out that there is an analogue of this on connected compact Lie groups; there is a volume form dg such that

$$\int L_h\phi(g)dg = \int \phi(g)dg = \int R_h\phi(g)dg$$

Moreover, this volume form is unique up to an overall constant.

If the Lie group is not compact, but still finite dimensional, there are two different volume forms d_Lg and d_Rg which are left and right invariant respectively; they might differ by a non-trivial function.

We will construct left and right volume elements d_Lg and d_Rg in the general case of non-compact (finite dimensional) Lie groups. Then we will see why they are proportional to each other in the case of compact groups.

Remark 72. Haar constructed a left-invariant measure on topological groups (i.e., without assuming they are differential manifolds), using only local

compactness. We do not need this more general and much more subtle concept. Even so, the invariant volume element is still often called "the Haar measure".

As always, the best practice is to work out a couple of examples before we work out the general case.

13.4.1. *Example: $GL(n, R)$*

In some ways the general linear group is the easiest case. So let us start with that. The matrix elements themselves provide a natural co-ordinate system on $GL(n, R)$; the only condition is that the determinant be non-zero. This cuts out a one-dimensional sub-set out of $\mathbb{R}^{n^2}$. The co-ordinates are transformed by left multiplication $g \mapsto h^{-1}g$

$$g_{ij} \mapsto \sum_k [h^{-1}]_{ik} g_{kj}$$

The volume element $dg_{11}dg_{21}\cdots dg_{n1}\ dg_{12}dg_{22}\cdots dg_{n2}\cdots dg_{1n}\cdots dg_{nn}$ is transformed by the magnitude of the Jacobian determinant of this transformation. The first column of co-ordinates

$$g_{11}, g_{21}, \ldots, g_{n1}$$

are transformed linearly by the matrix h^{-1}. This means that

$$dg_{11}dg_{21}\cdots dg_{n1} \mapsto \left|\det h^{-1}\right| dg_{11}dg_{21}\cdots dg_{n1}$$

The next column similarly changes by

$$dg_{12}dg_{22}\cdots dg_{n2} \mapsto \left|\det h^{-1}\right| dg_{12}dg_{22}\cdots dg_{n2}$$

and so on. Putting all the changes from the columns together,

$$dg_{11}dg_{21}\cdots dg_{n1}\ dg_{12}dg_{22}\cdots dg_{n2}\cdots$$
$$dg_{1n}\cdots dg_{nn} \to \left|\det h^{-1}\right|^n dg_{11}dg_{21}\cdots dg_{n1}\ dg_{12}dg_{22}\cdots dg_{n2}\cdots dg_{1n}\cdots dg_{nn}$$

Now, the determinant of g transforms as

$$\det g \mapsto (\det h^{-1}) \det g$$

so that its magnitude transforms as

$$|\det g| \mapsto \left|\det h^{-1}\right| |\det g|$$

Thus the combination

$$d_L g = \frac{dg_{11}dg_{21}\cdots dg_{n1}\ dg_{12}dg_{22}\cdots dg_{n2}\cdots dg_{1n}\cdots dg_{nn}}{|\det g|^n}$$

is invariant under left translations. Exactly the same argument (applied to rows instead of columns) shows that the same expression is also invariant under right translations. So we have the left and right invariant volume element

$$dg = \frac{dg_{11}dg_{21}\cdots dg_{n1}\ dg_{12}dg_{22}\cdots dg_{n2}\cdots dg_{1n}\cdots dg_{nn}}{|\det g|^n}$$

on $GL(n,R)$. This group is not compact; so the volume of the group $\int dg$ is infinite. So even for some non-compact groups the left and right invariant volume elements coincide.

We now turn to another example that illustrates phenomena of a different sort.

13.4.2. *Example: The upper triangular group*

Consider the group of matrices of the form $g = \begin{pmatrix} a_0 & a_1 \\ 0 & 1 \end{pmatrix}$ with $a_0 > 0$ and $a_1 \in \mathbb{R}$. In this case a_0, a_1 are themselves co-ordinates on the group. The Lie algebra has basis

$$t_0 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad t_1 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Consider the combinations

$$g^{-1}\frac{\partial g}{\partial a^i}da^i = \begin{pmatrix} a_0^{-1} & -\frac{a_1}{a_0} \\ 0 & 1 \end{pmatrix}\begin{pmatrix} da_0 & da_1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} \frac{da_0}{a_0} & \frac{da_1}{a_0} \\ 0 & 0 \end{pmatrix} \equiv \lambda_0 t_0 + \lambda_1 t_1$$

$$\lambda_0 = \frac{da_0}{a_0}, \quad \lambda_1 = \frac{da_1}{a_0}$$

Under a left translation by $h = \begin{pmatrix} b_0 & b_1 \\ 0 & 1 \end{pmatrix}$ we have

$$h^{-1}g = \begin{pmatrix} \frac{a_0}{b_0} & \frac{a_1 - b_1}{b_0} \\ 0 & 1 \end{pmatrix}$$

Treating h, b_0, b_1 as constants, we can vary g, a_0, a_1:

$$\lambda_0 \mapsto \frac{d\left(\frac{a_0}{b_0}\right)}{\frac{a_0}{b_0}} = \lambda_0, \quad \lambda_1 \mapsto \frac{d\left[\frac{a_1 - b_1}{b_0}\right]}{\frac{a_0}{b_0}} = \lambda_1$$

Thus, λ_0, λ_1 are left-invariant differentials.

Similarly,

$$\frac{\partial g}{\partial a^i} da^i g^{-1} = \begin{pmatrix} da_0 & da_1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} a_0^{-1} & -\frac{a_1}{a_0} \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{da_0}{a_0} & da_1 - \frac{a_1}{a_0} da_0 \\ 0 & 0 \end{pmatrix} \equiv \rho_0 t_0 + \rho_1 t_1$$

$$\rho_0 = \frac{da_0}{a_0}, \quad \rho_1 = da_1 - \frac{a_1}{a_0} da_0$$

Under a right translation

$$gh = \begin{pmatrix} a_0 b_0 & a_1 + a_0 b_1 \\ 0 & 1 \end{pmatrix}$$

$$\rho_0 \mapsto \frac{d[a_0 b_0]}{a_0 b_0} = \rho_0,$$

$$\rho_1 \mapsto d\left[a_1 + a_0 b_1\right] - \frac{a_1 + a_0 b_1}{a_0 b_0} d[a_0 b_0] = da_1 + b_1 da_0 - \frac{a_1}{a_0} da_0 - b_1 da_0 = \rho_1$$

We see that ρ_0 and ρ_1 are right invariant differentials. Thus,[2]

$$d_L g = \frac{da_0 da_1}{a_0^2}, \quad d_R g = \frac{da_0 da_1}{a_0}$$

are left and right invariant volume elements respectively. Note that they are not the same:

$$d_L g = \Delta(g) d_R g, \quad \Delta(g) = \frac{1}{a_0}$$

The ratio $\frac{d_L g}{d_R g} \equiv \Delta : G \to \mathbb{R}^+$ is a group homomorphism.

13.4.3. *Invariant differentials*

In a more complicated matrix group (such as $SO(n)$) we cannot argue as for $GL(n, R)$, as the matrix elements are not independent of each other. The second example suggests that we must use invariant differential 1−forms; and take their wedge product (a kind of determinant) to get the invariant volume element.

The matrix elements are functions of some independent co-ordinates $\xi^1, \ldots, \xi^D$ where D is the dimension of the group. Under an infinitesimal change of the co-ordinates, the matrix elements will change by $\sum_k \frac{\partial g}{\partial \xi^k} d\xi^k$. It is more

[2]For brevity, we are using the idea of a wedge product in differential geometry. The same result can be obtained by taking the determinant of the components of λ and ρ.

natural to consider the products with g^{-1}

$$g^{-1}\frac{\partial g}{\partial \xi^k}$$

as they belong to the Lie algebra of the group. (Recall our discussion of the Baker–Campbell–Hausdorff lemma where we saw that derivatives such as $g^{-1}(t)\frac{dg(t)}{dt}$ belong to the Lie algebra). Under a transformation $g \mapsto h^{-1}g$ (where h is independent of g) this is invariant.

Choosing a basis for the Lie algebra $t_a,\ a = 1, \dots D$ we can define a $D \times D$ matrix λ whose components are λ_{ak}:

$$\sum_a \lambda_{ak}(g)t_a = g^{-1}\frac{\partial g}{\partial \xi^k}$$

By using the chain rule of differentiation, under a change of coordinates $\xi^k \mapsto \eta^k$ the matrix λ transforms as

$$\lambda_{ak} \mapsto \sum_l \lambda_{al}\frac{\partial \eta^l}{\partial \xi^k}.$$

Thus its determinant transforms as

$$\det\lambda \mapsto \det\lambda \det\frac{\partial \eta}{\partial \xi}$$

So the combination

$$\det\lambda d\xi^1 \cdots d\xi^D$$

transforms as a volume element invariant under these co-ordinate transformations. A change of basis in the Lie algebra

$$t_a \mapsto S_{ab}t_b$$

will only change $\det\lambda d\xi^1 \cdots d\xi^D$ by a constant multiple:

$$\det\lambda d\xi^1 \cdots d\xi^D \mapsto \det S \det\lambda d\xi^1 \cdots d\xi^D.$$

Thus we see that:

$$d_L g = |\det\lambda|\ d\xi^1 \cdots d\xi^d$$

is invariant under left translations, as well as changes of basis and co-ordinates (up to a constant multiple).

In the same spirit, we can get a right-invariant volume element from the matrix ρ:

$$\sum_a \rho_{ai} t_a = \frac{\partial g}{\partial \xi^i} g^{-1}$$

$$d_R g = |\det \rho|\, d\xi^1 \cdots d\xi^d$$

These $d_L g$ and $d_R g$ may not coincide.

Exercise 73. Show that inversion $g \mapsto g^{-1}$ takes the left invariant volume element to the right invariant one:

$$d_L g^{-1} = d_R g$$

Solution Note that $\frac{\partial g^{-1}}{\partial \xi^i} = -g^{-1} \frac{\partial g}{\partial \xi^i} g^{-1}$ so that

$$g \frac{\partial g^{-1}}{\partial \xi^i} = -\frac{\partial g}{\partial \xi^i} g^{-1} \implies$$

$$\lambda_{ai}(g^{-1}) t_a = -\rho_{ai}(g) t_a$$

and

$$|\det \lambda(g^{-1})| = |\det \rho(g)|\,.$$

So, if the left and right invariant volume elements coincide, they are also invariant under inversion.

13.4.4. *The modular homomorphism*

Since

$$\frac{\partial g}{\partial \xi^k} g^{-1} = g \left(g^{-1} \frac{\partial g}{\partial \xi^k} \right) g^{-1}$$

we have a relation between λ and ρ:

$$\sum_{ak} \rho_{ak} t_a = g \left(\sum_{ak} \lambda_{ak} t_a \right) g^{-1}$$

Since

$$g t_a g^{-1} = \hat{g}_{ab} t_b$$

where $\hat{g}_{ab}$ is the matrix of g in the adjoint representation, we have

$$\rho_{ak} = \hat{g}_{ba} \lambda_{bk}$$

Thus, the ratio

$$\Delta(g) = \frac{d_R g}{d_L g} = \frac{|\det \rho(g)|}{|\det \lambda(g)|}$$

is simply the magnitude of the determinant of the matrix representing g in the adjoint representation:

$$\Delta(g) = |\det \hat{g}|$$

In particular, it is a continuous homomorphism $\Delta : G \to \mathbb{R}^+$:

$$\Delta(gh) = \Delta(g)\Delta(h)$$

This is called the modular homomorphism. If $\Delta(g) = 1$ the left and right invariant volume elements are equal. Such groups are said to be unimodular.

13.4.4.1. *Unimodular groups*

There are many examples of unimodular groups:

- For compact groups $\Delta(g) = 1$: The left and right invariant measures are the same.

For, suppose there were a $g \in G$ with $\Delta(g) \neq 1$. Then either $\Delta(g^r)$ or $\Delta(g^{-r})$ would tend to infinity for large r. But continuous functions on a compact group are bounded. So this can't happen.

In particular, groups such as $SU(n)$ or $SO(n)$ are unimodular, being compact.

- If G is a simple Lie group (i.e.,the only normal Lie subgroup is the trivial one) again $\Delta(g) = 1$

Even if there are normal subgroups that are discrete (i.e., not Lie subgroups) unimodularity holds. For, the kernel of Δ would be a normal *Lie* subgroup. For example, $SL_2(R)$ is unimodular; because its normal subgroups are discrete.

We saw earlier that

- $GL(n, R)$ is unimodular; although it is neither simple nor compact.

Obviously,

- Abelian groups (compact or not) are unimodular: There is no difference between left and right translations

In these cases we will denote by $d_{LR}g$ a volume element that is both left and right invariant. Note that our construction only determines it up to an overall constant: a change of basis in the Lie algebra can change it by a constant multiple.

13.4.5. *Example: $SU(2)$ in exponential co-ordinates*

Let us work out an example in detail. Recall that the group $SU(2)$ can be described by exponential co-ordinates $g = e^a$ with a being a trace-less anti-hermitian 2×2 matrix. Explicitly,

$$g = \cos\left(\frac{|\mathbf{a}|}{2}\right) + \mathbf{a}\cdot\mathbf{s}\frac{\sin\left(\frac{|\mathbf{a}|}{2}\right)}{\frac{|\mathbf{a}|}{2}}, \quad \mathbf{a} \in \mathbb{R}^3, \quad |\mathbf{a}| < 2\pi$$

where s_1, s_2, s_3 is a specific basis in the Lie algebra. The co-ordinate system breaks down at $|\mathbf{a}| = 2\pi$; all the points with $|\mathbf{a}| = 2\pi$ correspond to $g = -1$.

Using

$$df(|\mathbf{a}|) = f'(|\mathbf{a}|)\,\frac{\mathbf{a}\cdot d\mathbf{a}}{|\mathbf{a}|}$$

we get

$$dg = \left[-\frac{1}{2}\sin\frac{|\mathbf{a}|}{2} + \mathbf{a}\cdot\mathbf{s}\left\{\frac{|\mathbf{a}|\cos\left(\frac{|\mathbf{a}|}{2}\right) - 2\sin\left(\frac{|\mathbf{a}|}{2}\right)}{|\mathbf{a}|^2}\right\}\right]\frac{\mathbf{a}\cdot d\mathbf{a}}{|\mathbf{a}|} + d\mathbf{a}\cdot\mathbf{s}\frac{\sin\left(\frac{|\mathbf{a}|}{2}\right)}{\frac{|\mathbf{a}|}{2}}$$

Since

$$g^{-1} = \cos\left(\frac{|\mathbf{a}|}{2}\right) - \mathbf{a}\cdot\mathbf{s}\frac{\sin\left(\frac{|\mathbf{a}|}{2}\right)}{\frac{|\mathbf{a}|}{2}}$$

we get

$$\begin{aligned} g^{-1}dg = {} & \cos\left(\frac{|\mathbf{a}|}{2}\right)\left[-\frac{1}{2}\sin\frac{|\mathbf{a}|}{2} + \mathbf{a}\cdot\mathbf{s}\left\{\frac{|\mathbf{a}|\cos\left(\frac{|\mathbf{a}|}{2}\right) - 2\sin\left(\frac{|\mathbf{a}|}{2}\right)}{|\mathbf{a}|^2}\right\}\right]\frac{\mathbf{a}\cdot d\mathbf{a}}{|\mathbf{a}|} \\ & + d\mathbf{a}\cdot\mathbf{s}\frac{\sin\left(\frac{|\mathbf{a}|}{2}\right)\cos\left(\frac{|\mathbf{a}|}{2}\right)}{\frac{|\mathbf{a}|}{2}} \\ & + \left[\frac{1}{2}\mathbf{a}\cdot\mathbf{s}\frac{\sin\left(\frac{|\mathbf{a}|}{2}\right)}{\frac{|\mathbf{a}|}{2}}\sin\frac{|\mathbf{a}|}{2} - (\mathbf{a}\cdot\mathbf{s})^2\frac{\sin\left(\frac{|\mathbf{a}|}{2}\right)}{\frac{|\mathbf{a}|}{2}}\right. \\ & \left.\times\left\{\frac{|\mathbf{a}|\cos\left(\frac{|\mathbf{a}|}{2}\right) - 2\sin\left(\frac{|\mathbf{a}|}{2}\right)}{|\mathbf{a}|^2}\right\}\right]\frac{\mathbf{a}\cdot d\mathbf{a}}{|\mathbf{a}|} - \mathbf{a}\cdot\mathbf{s}d\mathbf{a}\cdot\mathbf{s}\left[\frac{\sin\left(\frac{|\mathbf{a}|}{2}\right)}{\frac{|\mathbf{a}|}{2}}\right]^2 \end{aligned}$$

Using (4.3.1) and some trig identities we can simplify this to

$$g^{-1}dg = \left[\frac{|\mathbf{a}| - \sin|\mathbf{a}|}{|\mathbf{a}|^3}\right]\mathbf{a}\cdot\mathbf{s}\,\mathbf{a}\cdot d\mathbf{a} + d\mathbf{a}\cdot\mathbf{s}\frac{\sin|\mathbf{a}|}{|\mathbf{a}|} + (\mathbf{a}\times d\mathbf{a})\cdot\mathbf{s}\left[\frac{1-\cos|\mathbf{a}|}{|\mathbf{a}|^2}\right]$$

Or,

$$g^{-1}dg = \mathbf{s}^T\left\{\frac{|\mathbf{a}| - \sin|\mathbf{a}|}{|\mathbf{a}|^3}\mathbf{a}\otimes\mathbf{a} + 1_3\frac{\sin|\mathbf{a}|}{|\mathbf{a}|} + \hat{\mathbf{a}}\frac{1-\cos|\mathbf{a}|}{|\mathbf{a}|^2}\right\}d\mathbf{a}$$

where $\hat{\mathbf{a}} = \begin{pmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{pmatrix}$ and $\mathbf{a}\otimes\mathbf{a} = \begin{pmatrix} a_1^2 & a_1a_2 & a_1a_3 \\ a_1a_2 & a_2^2 & a_2a_3 \\ a_1a_3 & a_2a_3 & a_3^2 \end{pmatrix}$ as before. The matrix in the curly brackets is

$$\lambda(g) = \frac{|\mathbf{a}| - \sin|\mathbf{a}|}{|\mathbf{a}|^3}\mathbf{a}\otimes\mathbf{a} + 1_3\frac{\sin|\mathbf{a}|}{|\mathbf{a}|} + \hat{\mathbf{a}}\frac{1-\cos|\mathbf{a}|}{|\mathbf{a}|^2}$$

Then,

$$\det\lambda(g) = \frac{4\sin^2\frac{|\mathbf{a}|}{2}}{|\mathbf{a}|^2}$$

The left-invariant volume element in exponential co-ordinates is thus,

$$d_Lg = \frac{4\sin^2\frac{|\mathbf{a}|}{2}}{|\mathbf{a}|^2}da_1da_2da_3$$

The volume of the group $SU(2)$ in these conventions is

$$4\pi\int_0^{2\pi}4\sin^2\frac{|\mathbf{a}|}{2}\,d|\mathbf{a}| = 16\pi^2$$

The 4π comes from the "angular co-ordinates", leaving an integral over the "radial" co-ordinate $|\mathbf{a}|$.

13.5. Representations of a Compact Lie Group

The fundamental work of Peter and Weyl developed the harmonic analysis on compact Lie groups. They showed that $L^2(G)$ can be decomposed into direct sums of finite dimensional unitary representations. The whole theory is remarkably similar to that of finite groups: Compactness is a good substitute for finiteness. The basic reason is that the integral of any continuous function on a compact group

is finite. In particular, the integral $\int dg = \mathrm{vol}(G)$ is finite. This volume plays a role analogous to the number of elements $|G|$ in the case of finite groups.

One important difference with the case finite groups is that the left regular representation is infinite dimensional. In particular, it does not have a character: The trace diverges. Peter and Weyl found a way around using this character.

13.5.1. *Finite dimensional representations of a compact group are unitary*

The trick is again, "averaging" over the group. Let $\rho : G \to GL(V)$ be a continuous finite dimensional representation on some vector space V. Pick some inner product $((,))$ on this vector space. Then define

$$(a, b) = \int ((\rho(g)a, \rho(g)b))\, d(g), \quad a, b \in V$$

Now,

$$\begin{aligned}(\rho(h)a, \rho(h)b) &= \int ((\rho(g)\rho(h)a, \rho(g)\rho(h)b))dg \\ &= \int ((\rho(gh)a, \rho(gh)b))dg\end{aligned}$$

using the invariance under $g \mapsto gh^{-1}$ we get

$$(\rho(h)a, \rho(h)b) = \int ((\rho(g)a, \rho(g)b))\, d(g) = (a, b)$$

proving unitarity.

Now recall that we already proved in the earlier chapter that

13.5.2. *Finite dimensional unitary representations are completely reducible*

This means that any finite dimensional representation can be decomposed as direct sum

$$\rho = \bigoplus_{r \in \tilde{G}} m_r r$$

where $\tilde{G}$ is the set of equivalence classes of irreducible representations; and m_r is the multiplicity (i.e., the number of copies of r contained in ρ).

The proof of Schur's Lemma and its corollaries are also essentially the same:

13.5.3. *Schur's lemma*

Lemma. *Let r and s be two irreducible representations, on vector spaces V and W respectively, of a compact group G; and T is a linear map $T : V \to W$ such that*

$$T[r(g)v] = s(g)[Tv], \forall g \in G, v \in V \tag{13.5.1}$$

Then either T is an isomorphism or it is zero.

Corollary. *If an operator commutes with all the representation matrices of an irreducible representation, it is a multiple of the identity.*

Also,

Corollary. *Any irreducible representation of an abelian group is one dimensional.*

13.5.4. *The character of a finite dimensional representation*

The character of a representation remains a powerful tool in the theory of compact groups. If $\rho : G \to GL(V)$ is a finite dimensional representation of a compact group, its character is the trace:

$$\chi_\rho(g) = \mathrm{tr}\rho(g)$$

ρ being finite dimensional, the trace converges for any g. In particular

$$\chi_{\rho(1)} = \dim \rho$$

We also have the notion of direct sum of representations

$$\rho \oplus \sigma(g) = \begin{pmatrix} \rho(g) & 0 \\ 0 & \sigma(g) \end{pmatrix}$$

with

$$\chi_{\rho\oplus\sigma}(g) = \chi_\rho(g) + \chi_\sigma(g)$$

More generally, if

$$\rho = \bigoplus_{r\in\tilde{G}} m_r r$$

we have

$$\chi_\rho(g) = \sum_{r\in\tilde{G}} m_r \chi_r(g)$$

We will have to work a bit harder to understand the characters of infinite dimensional unitary representations.

13.5.5. *Inner product space $C(G)$*

We can define an inner product in the vector space of continuous functions

$$\langle \phi, \psi \rangle = \int \phi^*(g)\psi(g)dg$$

This is not quite a Hilbert space: We cannot yet prove that it has a countable basis. Indeed, this completeness is a direct consequence of the Peter–Weyl theorem which we have not proved yet. Still this inner product is a powerful tool when combined with Schur's lemma.

13.5.6. *Orthogonality of representation matrix elements*

Exactly as in the case of finite groups, we can define the quantities

$$T_{ia}^{(bj)} = \int r_{ba}(h^{-1})s_{ij}(h)dh$$

Here, r and s are two inequivalent unitary representations and r_{ab} and s_{ij} are their matrix elements in orthogonal bases.

The invariance of the volume element can be used to prove that, for each choice of bj, this is an intertwining operator between the representations r and s. Since they are not equivalent, Schur's lemma says this is zero:

Proposition 74. *Matrix elements of inequivalent irreducible representations are orthogonal*

Also, defining

$$T_{a'a}^{(bb')} = \int r_{ba}(h^{-1})r_{a'b'}(h)dh$$

we again have (for each choice of bb') an intertwining operator of r to itself. This time Schur's Lemma gives,

$$\langle r_{ab}, r_{a'b'} \rangle = C_r \delta_{aa'}\delta_{bb'} \tag{13.5.2}$$

for some C_r. We can determine it by putting $a = a'$ and summing over a. On the lhs we get

$$\sum_a \int r_{ab}^*(g)r_{ab'}(g)dg = \int \left[r^\dagger(g)r(g)\right]_{bb'} dg$$

Using unitarity of r

$$\sum_a \langle r_{ab}, r_{a'b'} \rangle = \mathrm{vol}(G)\delta_{bb'}.$$

Here $\mathrm{vol}(G) = \int dg$ is the volume of the group. It depends on the choice of invariant volume element. The rhs of (13.5.2) is, upon putting $a = a'$ and summing over a, equal to $C_r \ \dim r \ \delta_{bb'}$. Thus $C_r = \frac{\mathrm{vol}(G)}{\dim r}$.

Proposition 75. *Matrix elements of an irreducible representation satisfy*

$$\langle r_{ab}, r_{a'b'} \rangle = \frac{\mathrm{vol}(G)}{\dim r}\delta_{aa'}\delta_{bb'}$$

We can combine these two statements to

Proposition 76. *The functions* $\sqrt{\frac{\dim r}{\mathrm{vol}(G)}} r_{ab}$ *form an orthogonal family when* r *runs over the set of equivalence classes of unitary irreducible representations and* a, b *label an orthonormal basis in the vector space of each such* r

What we do not have as yet is the completeness of this family: That any continuous can be expanded as a linear combination of functions in this family.This is the content of the Peter–Weyl theorem (see below).

13.5.7. *Fourier components of a function*

We can, as in the case of finite groups, define the components of a function $\phi : G \to \mathbb{C}$ along this orthogonal family:

$$\tilde{\phi}^r_{ab} = \int r^*_{ab}(h)\phi(h)dh$$

Just from the orthonormality of the family, we have[3]

Proposition 77. *Bessel's inequality*

$$\sum_{r \in \tilde{G}} \sum_{ab=1}^{\dim r} \frac{\dim r}{\mathrm{vol}(G)} \mid \tilde{\phi}^r_{ab} \mid^2 \leq \int \phi^*(h)\phi(h)dh$$

Once we have completeness (that any continuous function can be expanded in the above basis with these components) we can get the Plancherel theorem/Parseval's formula.

[3]The factor $\frac{\dim r}{\mathrm{vol}(G)}$ comes from the fact that its is $\sqrt{\frac{\dim r}{\mathrm{vol}(G)}} r_{ab}$ that is the orthonormal family; not r_{ab} which appears in the definition of the components.If we had defined components w.r.t. the orthonormal family, there would be some factors of $\frac{\dim r}{\mathrm{vol}(G)}$ in other places.

For each $r \in \tilde{G}$ the components form a matrix $\tilde{\phi}^r_{ab}$. The sum $\sum_{ab} \mid \tilde{\phi}^r_{ab} \mid^2$ for each fixed r is the trace of the matrix $\tilde{\phi}^{r\dagger}\tilde{\phi}^r$. (This trace is also called the Hilbert-Schmidt norm of the matrix $\tilde{\phi}^r$). The matrix $\tilde{\phi}^{r\dagger}\tilde{\phi}^r$ is therefore of some interest.

13.5.8. *The convolution algebra*

Given two continuous functions, define the convolution

$$\phi \circ \psi(g) = \int \phi(gh^{-1})\psi(h)dh$$

Using the invariance of the volume element we can see that this is an associative multiplication; although not commutative if G is not abelian.

Proposition 78. *Fourier components of a convolution are the matrix products*[4] *of the Fourier components of each function*:

$$\left(\widetilde{\phi \circ \psi}\right)^r_{ab} = \sum_c \tilde{\phi}^r_{ac}\tilde{\psi}^r_{cb}$$

Proof. For,

$$\left(\widetilde{\phi \circ \psi}\right)^r_{ab} = \int r^*_{ab}(g)dg \int \phi(gh^{-1})\psi(h)dh = \int r^*_{ab}(g)\phi(gh^{-1})\psi(h)dgdh$$

Using the invariance of the volume element under the change $g \mapsto gh$

$$\left(\widetilde{\phi \circ \psi}\right)^r_{ab} = \int r^*_{ab}(gh)\phi(g)\psi(h)dgdh$$

Using the representation property of r

$$\left(\widetilde{\phi \circ \psi}\right)^r_{ab} = \sum_{c=1}^{\dim r} \int r^*_{ac}(g)r^*_{cb}(h)\phi(g)\psi(h)dgdh$$

The integral factorizes into one over g and another over h:

$$\left(\widetilde{\phi \circ \psi}\right)^r_{ab} = \sum_c \tilde{\phi}^r_{ac}\tilde{\psi}^r_{cb}$$

□

[4] This accounts for the non-commutativity of the convolution.

13.6. The Peter–Weyl Theorem

A function of the form

$$\sum_{r\in\tilde{G}}\sum_{ab} c^r_{ab} r_{ab}(g)$$

with only a *finite number of non-zero terms* is the non-commutative analogue of a trigonometric polynomial: The matrix elements $r_{ab}(g)$ generalize the exponential functions on $U(1)$. Let $L^2_{\text{alg}}(G)$ be the space of these "algebraic" functions. Then the abstract version of the Peter–Weyl theorem is that $L^2(G)$ is the completion $\overline{L^2_{\text{alg}}(G)}$ of this space of algebraic functions. But we can be more concrete: Determine the coefficients c^r_{ab}.

Exercise 79. Show that of all linear combinations of the form $\sum_{r,ab} c^r_{ab} r_{ab}(g)$ the one that minimizes the distance $||\phi - \sum_{r,ab} c^r_{ab} r_{ab}||$ has

$$c^r_{ab} = \frac{\dim r}{\text{vol}(G)} \tilde{\phi}^r_{ab}$$

Solution Minimize $\int \mid \phi(g) - \sum_{r,ab} c^r_{ab} r_{ab}(g) \mid^2 dg$ with respect to c^r_{ab} and use orthonormality of $\sqrt{\frac{\dim r}{\text{vol}(G)}} r_{ab}(g)$.

Theorem 80. ***Peter–Weyl*** *Let $\phi : G \to \mathbb{C}$ be a continuous function with Fourier components in each irreducible representation $r \in \tilde{G}$*

$$\tilde{\phi}^r_{ab} = \int r^*_{ab}(g)\phi(g)dg.$$

Then the Fourier series converges in the L^2- norm, synthesizing ϕ from its components:

$$\phi(g) = \sum_{r\in\tilde{G}}\sum_{ab} \frac{\dim r}{\text{vol}(G)} \tilde{\phi}^r_{ab} r_{ab}(g)$$

Moreover, this allows us to promote the Bessel inequality to an equality:

Theorem. ***(Plancherel)*** *We have*

$$\int \mid \phi(g) \mid^2 dg = \sum_{r\in\tilde{G}} \sum_{a,b=1}^{\dim r} \frac{\dim r}{\text{vol}(G)} \mid \tilde{\phi}^r_{ab} \mid^2$$

The original proof of Peter and Weyl used the theory of compact integral operators, which had already been developed in Schmidt's thesis with Hilbert. This is still a good way of understanding the proof. We will make a digression

to review the theory of such operators [28], which is also useful in many other places in mathematical physics (e.g., the Green's function of the Laplacian in a compact manifold is such an operator). We won't give complete proofs, only a summary of the main ideas. A complete proof is given in the rigorous (yet remarkably clear) course meant for mathematicians by P. Etingof at MIT, available online [29].

13.6.1. *Compact integral operators*

Let us start by considering a self-adjoint linear operator (i.e., hermitian matrix) $K : V \to V$ in a finite dimensional vector space V with an inner product. A basic result is the spectral theorem:

$$V = \ker K \oplus \bigoplus_{\lambda \neq 0} E_\lambda$$

That is, the vector space V can be decomposed orthogonally into

- $\ker K$, the subspace of vectors ψ such that $K\psi = 0$
- the eigenspaces E_λ of vectors ψ satisfying $K\psi = \lambda\psi$. These eigenvalues are real and non-zero; the eigenspaces corresponding to unequal eigenvalues are orthogonal to each other

Since V is finite dimensional, $\ker K$ and the eigenspaces E_λ are also finite dimensional.

Hilbert and Schmidt found a generalization of this spectral theorem to integral operators of the form

$$K\psi(x) = \int_X K(x, y)\psi(y)dy$$

where X is a compact topological space and dy is a measure on it with finite volume

$$\mathrm{vol}(X) = \int_X dx.$$

In the application of interest to us, X will be a compact group and the measure will be in the invariant volume element dg. Self-adjointness means in this case

$$K(x, y) = K^*(y, x)$$

In other words, we are treating x, y as continuous matrix indices, with sums replaced by integrals. The kernel of K is again the subspace of vectors with

eigenvalue zero:

$$\ker K = \{\psi \mid K\psi = 0\}$$

The eigenspaces E_λ are again defined similarly:

$$E_\lambda = \{\psi \mid K\psi = \lambda\psi\}, \quad \lambda \neq 0$$

The eigenvalues λ of hermitian integral operators such as K are again real. Each eigenspace E_λ with $\lambda \neq 0$ is finite dimensional; in other words, non-zero eigenvalues have finite multiplicity. (There can be an infinite number of such non-zero eigenvalues.) But the degeneracy of a zero eigenvalue can be infinite (for example, a compact operator can have finite rank. See below).

Remark 81. A word of explanation is perhaps needed to understand why non-zero eigenvalues have finite multiplicity, but a zero eigenvalue can have infinite multiplicity. Within each eigenspace, the operator K reduces to a multiple of the identity, $\lambda 1$ on E_λ. A multiple of the identity is compact only if either

- the space E_λ is finite dimensional
- the multiple λ is zero; so $\ker K$can be infinite dimensional

Theorem 82. ***(Hilbert–Schmidt) A compact self-adjoint operator*** $K : L^2(X) \to L^2(X)$ ***yields an orthogonal decomposition***

$$L^2(X) = \ker K \oplus \overline{\bigoplus}_{\lambda \neq 0} E_\lambda \tag{13.6.1}$$

Since there can be an infinite number of non-zero eigenvalues, we must take the a completion of their direct sums; this is the meaning of the $\overline{\bigoplus}$ symbol.

An equivalent statement is that there is an orthonormal basis for $L^2(X)$ consisting of eigenvectors of K.

13.6.1.1. *Example: Finite rank operators*

Especially simple examples are finite rank operators. These have only a finite number of non-zero eigenvalues. So for them we do not need do any completion in (13.6.1).

They are *finite* linear combinations of the form

$$K(x, y) = \sum_a \lambda_a \psi_a(x) \psi_a^*(y)$$

Note that $\ker K$ is infinite dimensional. The image of K (i.e., the set of vectors that can arise as $K\psi$ for some ψ) is finite dimensional.This dimension is called the rank of K.

13.6.1.2. *Example: Heat kernel*

An example of a compact linear operator which is not of finite rank is the heat kernel on the circle:

$$h_t(x,y) = \sum_{n\in\mathbb{Z}} e^{-n^2 t} e^{inx} e^{-iny}, \quad t > 0$$

13.6.1.3. *The identity operator on an infinite dimensional Hilbert space is not a compact operator!*

Its integral kernel is the Dirac delta, which is *not* a continuous function on $X \times X$. Its only eigenvalue is 1, which has infinite multiplicity.

But a sequence of compact operators can provide an approximation to the identity operator. For example, the heat kernel above satisfies

$$\int h_t(x,y)\psi(y)dy \to \psi(x)$$

as $t \to 0$, for all continuous functions on the circle. For each $t > 0$, the operator h_t is compact, but not in the limit $t \to 0$.

13.6.2. *Proof of Peter-Weyl*

Armed with this machinery, we can outline a proof of the Peter-Weyl theorem. The idea is to approximate the identity operator on $L^2(X)$ (the Dirac delta) by a sequence of continuous functions. Each of these will define a compact self-adjoint operator to which we can apply the above spectral decomposition. More precisely, let $k_N : G \to \mathbb{R}$ be a sequence of continuous functions satisfying

$$\int k_N(g)dg = 1, \quad k_N(g) = k_N(g^{-1})$$

Also, we want the support of k_N (the set of g for which $k_N(g) \neq 0$) to shrink to just the identity element of G as $N \to \infty$.

It is not hard to construct such functions. From such a sequence we can construct the integral operators

$$K_N\psi(g) = \int K_N(g,h)\psi(h)dh$$

with

$$K_N(g,h) = k_N(gh^{-1})$$

Since $k_N(g^{-1}) = k_N(g)$ we get that $K_N(g,h) = K_N(h,g)$. That is, K_K is symmetric. Since it $k_N(g)$ is real we also get self-adjointness of K_N. The group

being compact we get that K_N is a compact integral operator. So, we have a spectral decomposition

$$L^2(G) = \ker K_N \oplus \bar{\bigoplus}_{\lambda \neq 0} E_{\lambda < N}$$

where $E_{\lambda,N}$ are the eigenspaces with non-zero eigenvalues. They are finite dimensional:

$$\psi \in E_{\lambda,N}, \quad \psi(g) = \sum_{\mu} c_\mu u_\mu(g),$$

the sum being finite.

Moreover, we can see that each $E_{\lambda,N}$ carries a representation of G. For this we note the symmetry

$$K_N(gh', hh') = K_N(g, h)$$

So K_N commutes with the right-translation $R_{h'}$:

$$R_{h'}\psi(g) = \psi(gh'),$$

$$K_N R_{h'} = R_{h'} K_N$$

Therefore

$$\psi \in E_{\lambda,N} \implies R_{h'}\psi \in E_\lambda$$

Thus there are matrices $\rho_{\mu\nu}(g)$ such that

$$u_\mu(gh') = \sum_{\nu} \rho_{\mu\nu}(h') u_\nu(g)$$

These matrices provide a finite dimensional representation of G

$$\rho_{\mu\nu}(h'h'') = \sum_{\sigma} \rho_{\mu\sigma}(h) \rho_{\sigma\nu}(h').$$

Thus the elements of E_λ are algebraic functions: They are finite linear combinations of irreducible representation matrices. Thus we see that

$$L^2(G) = \overline{L^2_{\text{alg}}(G)}$$

This is one statement of the Peter–Weyl theorem.

This page intentionally left blank

Chapter 14

QUANTUM GROUPS

Modern theoretical physics needs notions of symmetries that go beyond groups. The story begins with the solution by Bethe of the "isotropic spin chain" with hamiltonian

$$H = J \sum_{i=1}^{n} \sigma_i^a \sigma_{i+1}^a, \quad \sigma_{n+1}^a \equiv \sigma_1^a$$

where $\sigma_i^a = 1 \otimes \cdots \otimes \sigma^a \otimes 1 \cdots \otimes 1$ are Pauli matrices associated to the ith site on a chain (one dimensional lattice). It is invariant under rotations of all the spins simultaneously (i.e., global $SU(2)$ symmetry). This symmetry is crucial to solve the problem, by a fiendishly clever guess (the Bethe ansatz).

Later, Yang and Baxter were able to solve this problem even when the rotation invariance is broken:

$$H = \sum_{i=1}^{n} \sum_{a=1}^{3} J_a \sigma_i^a \sigma_{i+1}^a,$$

This was possible because of a mysterious identity satisfied by the scattering matrix of the spin waves (the "Yang-Baxter Relation"). Faddeev and collaborators discovered that this is related to the quantum analogue of the integrability of the corresponding classical spin chains. Although the systems do not appear to be rotation invariant at first, they are invariant under a "quantum deformation" of the rotation group.

It was Drinfeld who realized that the underlying mathematical structure is a "Hopf algebra", a generalization of the idea of a group. Hopf had postulated them, motivated by applications to Algebraic Topology. Even when a group is non-abelian, the algebra of functions on the group (by point-wise multiplication) is commutative. Hopf algebra is a generalization in which this operation is also

non-commutative. Drinfeld coined the phrase "Quantum Group" to describe this new kind of non-commutativity.

Hopf and followers (Sweedler, Taft) had discovered the "quantum" analogue of finite groups while Faddeev *et al.* had found examples of "Quantum Lie Groups".

All this happened over nearly a century (starting with Bethe). The forty years since Drinfeld's work were especially productive mathematically: The classification of Hopf algebras and their modules (i.e., representations). There is no doubt that this is a major breakthrough in mathematics.

Groups were discovered first in a corner of mathematics far from physics("exact" solution of polynomial equations by radicals) . Much later they were found to be central to quantum mechanics and to modern particle physics.

"Quantum Groups" were discovered in a somewhat specialized area of physics.It is possible that Hopf Algebras will be more broadly important to physics and central to the next generation of fundamental physical theories. Glimpses of the future we have today (e.g., quantum gravity might be described by non-commutative geometry) point in this direction. There are already several textbooks [31, 33] devoted entirely to Quantum Groups. We will discuss only one approach to this rapidly evolving subject, along with the simplest examples.

14.1. Algebras and Co-Algebras

It is a theme of modern mathematics that the algebra of functions on an object contains all the information about it. For example, the algebra of (continuous, differentiable, analytic) functions of a (topological, differentiable, complex) manifold determines it.

The idea is to first translate all the properties of a group to those of the space of functions on it. Then we will generalize the idea of a group by generalizing the properties of this space. The end result will be an algebraic structure (the Hopf algebra) which may not any more be the space of functions on anything! This is similar to the way observables of a quantum system are no longer functions on any phase space: The process of quantization forces us to give up on the idea that observables are some kind of functions.

Let G be a finite group and let $C(G)$ be the space of complex-valued functions on it. This is a commutative algebra under point-wise multiplication:

$$f_1 f_2(g) = f_1(g) f_2(g).$$

The dimension of the vector space $C(G)$ is the number of elements of G; at each element of the group, we can specify one independent complex number as

the value of a function. By linearity we can extend this multiplication as a bilinear map

$$m : C(G) \otimes C(G) \to C(G).$$

But this multiplication has no information about the underlying group multiplication law: We can define such a commutative algebra on functions over any set. The group law allows us to take a function of one group element and turn it into a function of two elements:

$$\Delta(f)(g_1, g_2) = f(g_1 g_2).$$

In other words, Δ takes a function of one variable and turns it into a function of two variables.

We can think of a function of two elements as a "matrix"[1] labelled by g_1, g_2. More precisely, $\Delta(f) \in C(G) \otimes C(G)$. Thus, Δ is a map that goes in the opposite direction from a multiplication $\Delta : C(G) \to C(G) \otimes C(G)$: It is a co-multiplication. We can make this precise by defining the idea of a co-algebra: A kind of mirror image of an algebra.

Definition. An algebra is a vector space A along with a linear map $m : A \otimes A \to A$. A co-algebra is a linear map in the opposite direction $\Delta : B \to B \otimes B$ on some vector space B.

It is useful to recall how to think of an algebra in terms of its structure constants in some basis e_a. (For now, let us think of finite dimensional algebras for simplicity. Also, we will use the summation convention.) Then the product $e_a e_b$ can be expanded as a linear combination of the basis elements:

$$e_a e_b = m^c_{ab} e_c.$$

The quantities m^c_{ab} are the components, in this basis, of a third order tensor. They are called the structure constants of the algebra. Various properties of the algebra translate into identities satisfied by the structure constants

- Commutative: $e_a e_b = e_b e_a \iff m^c_{ab} = m^c_{ba}$
- Associative: $e_a(e_b e_c) = (e_a e_b)e_c \iff m^e_{ad} m^d_{bc} = m^d_{ab} m^e_{dc}$
- Unit element: $\eta = \eta^a e_a$, $\eta^b m^c_{ba} = \delta^c_a = m^c_{ab}\eta^b$
- Lie algebra: $m^c_{ab} = -m^c_{ba}$, (anti-symmetry) and $m^d_{ab}\ m^e_{dc} + m^d_{bc}\ m^e_{da} + m^d_{ca}\ m^e_{db} = 0$ (Jacobi identity)

The simplest way to construct a co-algebra is to take B to be the dual vector space of an algebra. The multiplication on A induces a co-multiplication on its dual.

[1] We saw such integral kernels already in the proof of the Peter-Weyl theorem.

Recall that an element of the dual of A is a linear function $f : A \to \mathbb{C}$. We can identify $B \otimes B$ with the space of bi-linear functions of two elements of A. Thus a co-algebra is a map

$$\Delta(f)(x, y) = f(xy).$$

So a co-algebra is a kind of mirror image to an algebra: Its dual. Thus there is a notion of a co-associative co-algebra, a notion of a co-unit etc. A co-unit for example, is a linear map

$$\epsilon : B \to \mathbb{C}$$

satisfying some conditions that are dual of the identity element in an algebra. Other ideas of algebra can also be translated. For example a co-algebra is co-commutative if $\Delta(b)$ is a symmetric tensor in $B \otimes B$.

Translating all this into components is again useful. Let e_a be a basis in B. Then the element $\Delta(e_a) \in B \otimes B$ can be expanded in terms of $e_b \otimes e_c$:

$$\Delta(e_a) = \Delta_a^{bc} e_b \otimes e_c$$

In this notation the various subtypes of co-algebras correspond to conditions on the structure constants of the co-multiplication:

- Co-Commutative: $\Delta(e_a) = \Delta(e_a)^T \iff \Delta_a^{bc} = \Delta_a^{cb}$
- Co-Associative: $\Delta_e^{ad}\Delta_d^{bc} = \Delta_d^{ab}\Delta_e^{dc}$
- Co-unit element: $\epsilon : B \to \mathbb{C}, \quad \epsilon(e_a) = \epsilon_a$, satisfying $\epsilon_b \Delta_c^{ba} = \delta_c^a = \Delta_c^{ab}\epsilon_b$
- Co-Lie algebra: $\Delta_c^{ab} = -\Delta_{bc}^c$(anti-symmetry) and $\Delta_d^{ab}\Delta_e^{dc} + \Delta_d^{bc}\Delta_e^{da} + \Delta_d^{ca}\Delta_e^{db} = 0$(co-Jacobi identity)

Thus a dual vector space of a co-Lie algebra is a Lie algebra and so on.

If the notion of a co-algebra is just the mirror image of an algebra, why do we need a separate theory for them? The point is that you may have the notion of a multiplication and a co-multiplication on the *same* vector space.

Definition. A bi-algebra is a vector space A on which there is a multiplication $m : A \otimes A \to A$ as well as a co-multiplication $\Delta : A \to A \otimes A$ which are compatible with each other. That is, the co-multiplication is an algebra homomorphism of A to $A \otimes A$.

Implicit here is that if A is an algebra, there is a natural multiplication on $A \otimes A$:

$$(x \otimes y)(s \otimes t) = xs \otimes yt.$$

Again, the conditions on the multiplication and co-multiplication of a bi-algebra can be translated into properties of the structure constants. Let us compute:

$$\Delta(e_a e_b) = \Delta(e_a)\Delta(e_b), \quad \Delta(e_a) = \Delta_a^{fg} e_f \otimes e_g, \quad \Delta(e_b) = \Delta_b^{hi} e_h \otimes e_i$$

$$e_f e_h = m_{fh}^d e_d, \quad e_g e_i = m_{gi}^e e_e,$$

$$\Delta(e_a e_b) = m_{ab}^c \Delta(e_c) = m_{ab}^c \Delta_c^{de} e_d \otimes e_c,$$

so that the bi-algebra condition is

$$m_{ab}^c \Delta_c^{de} = \Delta_a^{fg} m_{fh}^d m_{gi}^e \Delta_b^{hi}.$$

Note that this condition is symmetric between the structure constants of the algebra and the co-algebra. So the dual of a bi-algebra is again a bi-algebra with the structure constants of multiplication and co-multiplication interchanged.

14.1.1. *Examples*

- The space of complex-valued functions of a group is a commutative bi-algebra: The co-multiplication defined above using the group product is compatible with the point-wise multiplication. The associativity of group multiplication implies that the co-algebra $C(G)$ is co-associative. The identity element $e \in G$ in the group defines a co-unit: $\epsilon(f) = f(e)$.
- Since we haven't used the existence of an inverse, we can get a commutative bi-algebra on any semi-group.
- Given a Lie algebra with basis e_a, its universal envelope is the associative algebra generated by e_a satisfying the the relations $[e_a, e_b] = f_{ab}^c e_c$. With the co-product

  $$\Delta(e_a) = 1 \otimes e_a + e_a \otimes 1,$$

  this is a co-commutative (but usually non-commutative) Hopf algebra with the co-unit defined by $\epsilon(e_a) = 0$.

14.1.2. *Sweedler notation*

There is an elegant notation [31] due to Sweedler that allows us to do computations on bi-algebras more easily. Recall that $\Delta(a)$ can be written as a linear combination

of tensor products of elements of A:

$$\Delta(a) = \sum_k a_{(1)k} \otimes a_{(2)_k}$$

Sweedler suggests we drop the index on this and write it as

$$\Delta(a) = \sum a_{(1)} \otimes a_{(2)}$$

or even drop the summation symbol and write

$$\Delta(a) = a_{(1)} \otimes a_{(2)}$$

After a bit of practice this becomes natural and quite useful,just like Einstein's summation convention. But it has to be used cautiously until you are used to it. Have fun going back and forth between the index notation and the Sweedler notation.

As an example, let us consider the notion of a convolution on bi-algbras that generalizes the convolution of functions on a group. Let $\phi, \psi : A \to A$ be linear maps in a bi-algebra. Then we can produce another linear map $\phi * \psi : A \to A$

$$\phi * \psi(a) = \phi(a_{(1)})\psi(a_{(2)})$$

Let us translate this into basis notation:

$$\phi(e_a) = \phi_a^b e_b, \quad \psi(e_b) = \psi_a^b e_b$$

$$\Delta(e_a) = \Delta_a^{bc} e_b \otimes e_c$$

$$\phi * \psi\ (e_a) = \Delta_a^{bc} m(\phi(e_b) \otimes \psi(e_c)) = \Delta_a^{bc} \phi_b^d \psi_c^e m(e_d e_e) = \Delta_a^{bc} \phi_b^d \psi_c^e m_{de}^h e_h$$

so that

$$(\phi * \psi)_a^h = \Delta_a^{bc} \phi_b^d \psi_c^e m_{de}^h.$$

In particular if id is the identity map in A

$$\phi * \mathrm{id}(a) = \phi(a_{(1)})a_{(2)},$$

$$(\phi * \mathrm{id})_a^h = \Delta_a^{bc} \phi_b^d m_{dc}^h, \quad (\mathrm{id} * \psi)_a^h = \Delta_a^{bc} \psi_c^d m_{bd}^h.$$

This will be useful soon.

14.2. The AntiPode

We still haven't used the existence of an inverse in the group. It defines an "antipode" map on the group bi-algebra $C(G)$

$$S(f)(g) = f(g^{-1}), \quad f : G \to \mathbb{C}, \quad g \in G$$

satisfying certain compatibility relations. Also, $gg^{-1} = e$ becomes

$$(1 \otimes S)\Delta(f) = \Delta(\epsilon(f))$$

etc.Given a basis on $C(G)$ the anti-pode can be thought of in terms of its components: $S(e_a) = S_a^b e_b$.

Definition. A Hopf algebra is a bi-algebra H with a unit η and a co-unit ϵ which admits a linear map $S : H \to H$ (called the antipode) satisfying

$$S * \mathrm{id} = \eta \circ \epsilon = \mathrm{id} * S$$

In index notation this is $\Delta_a^{bc} S_b^d m_{dc}^h = \epsilon_a \eta^h = \Delta_a^{bc} S_c^d m_{bd}^h$.

There are also some (more obvious) conditions on the unit and co-unit which we omit [31]. Not all bi-algebras admit an antipode. If a bi-algebra does admit an antipode, it is unique [31].

The antipode reverses the multiplication and the co-multiplication:

$$S(ab) = S(b)S(a), \quad S_c^d m_{ab}^c = S_b^e S_a^f m_{ef}^d$$

$$\Delta^{\mathrm{op}}(S(d)) = (S \otimes S)\Delta(d), \quad S_d^c \Delta_c^{ab} = S_e^b S_f^a \Delta_d^{ef}$$

We have written out each statement in the index notation for clarity.

14.2.1. *Elementary examples of Hopf algebras*

- The bi-algebra of functions on a group become a Hopf algebra with the choice $S(f)(g) = f(g^{-1})$ mentioned above. This is commutative.
- The bi-algebra of a Lie algebra becomes a Hopf algebra with $S(e_a) = -e_a$. This is co-commutative.

14.2.2. *First example of a quantum group*

To go beyond groups and Lie algebras, we should look for Hopf algebras which are neither commutative nor co-commutative. Such Hopf algebras are called quantum groups: In a sense they are more non-commutative than either groups or Lie algebras.

The elements of such a general Hopf algebra may no longer be functions on any set! To generalize any property of a group we first translate it to a property of the space of functions on it; then find a way to generalize it to a Hopf algebra that may not be either commutative or non-commutative.

The first example was found by Sweedler and a more general one by Taft. These are the first "Quantum Groups" although that name was not coined until later.

Example 83. The Taft Algebra [32]. (The special case $n = 2$ is the Sweedler algebra.) Let ω be a primitive root of unity of order n; i.e., $\omega^n = 1$ and no smaller power of ω is equal to one. Define an algebra generated by a, u satisfying the relations

$$ua = \omega au, \quad u^n = 1, \quad a^n = 0.$$

The co-multiplication is

$$\Delta(u) = u \otimes u, \quad \Delta(a) = a \otimes 1 + u \otimes a$$

and the co-unit

$$\epsilon(a) = 0, \quad \epsilon(u) = 1$$

and anti-pode

$$S(a) = -u^{-1}a, \quad S(u) = u^{-1}$$

So this is neither commutative nor co-commutative. There is no underlying set on which elements such as a and u are functions. This is what we give up in order to have such a generalization.

Exercise 84. Verify that the above satisfies the axioms for a Hopf Algebra of dimension n^2.

14.2.2.1. *"Classical limit"*

In the limit $n \to \infty$ the Taft Algebra reduces to the co-commutative Hopf algebra corresponding to a familiar Lie algebra. Thus, in a sense it is a "quantization" of this Lie algebra. To see this, let us note that

$$\omega = 1 + \frac{2\pi i}{n} + \mathrm{O}\left(\frac{1}{n^2}\right)$$

Set

$$u = e^{\frac{2\pi i}{n}L_0}, \quad a = L_+$$

The condition $ua = \omega au$ becomes

$$\left\{1 + \frac{2\pi i}{n}L_0\right\} L_+ = \left\{1 + \frac{2\pi i}{n}\right\} L_+ \left\{1 + \frac{2\pi i}{n}L_0\right\} + \mathrm{O}\left(\frac{1}{n^2}\right) \iff$$

Equating terms of order $\frac{1}{n}$ we get

$$[L_0, L_+] = L_+$$

which are the commutation relations of the two dimensional non-abelian Lie algebra. Also,

$$\Delta(u) = \Delta\left(1 + \frac{2\pi i}{n} L_0\right) = \left\{1 + \frac{2\pi i}{n} L_0\right\} \otimes \left\{1 + \frac{2\pi i}{n} L_0\right\} + \mathrm{O}\left(\frac{1}{n^2}\right)$$
$$= 1 + \frac{2\pi i}{n} (L_0 \otimes 1 + 1 \otimes L_0) + \mathrm{O}\left(\frac{1}{n^2}\right)$$

so that to leading order

$$\Delta(L_0) = L_0 \otimes 1 + 1 \otimes L_0.$$

And, again to leading order, $\Delta(a) = a \otimes 1 + u \otimes a$ reduces to

$$\Delta(L_+) = L_+ \otimes 1 + 1 \otimes L_+.$$

14.3. Primitives, Group-Like Elements, Skew–Derivations

The following statements are easy to prove:

- Let H be a Hopf algebra. The subset of elements of Hsatisfying

  $$\Delta(g) = g \otimes g$$

  form a group. Such elements said to be "group-like".
  In the example above, u and its powers are group-like; they form the group Z_n.
- Elements of $v \in H$ satisfying

  $$\Delta(v) = v \otimes 1 + 1 \otimes v \tag{14.3.1}$$

  are said to be "primitives". If v and w are primitives, so is $vw - wv$. Thus the sub-space of primitives is a Lie algebra. The condition (14.3.1) is a version of the Leibnitz identity of differentiation:

  $$v(\phi\psi) = v(\phi)\psi + \phi v(\psi).$$

  Primitives are derivations in some sense. This will become clear when we look at Hopf modules.

- Given a group-like element $g \in H$ we can define a generalization of a primitive, a skew primitive w.r.t. g (also called a $g-$derivation) by the condition

$$\Delta(v) = v \otimes 1 + g \otimes v$$

They are derivations up to an action by a group-like element:

$$v(\phi\psi) = v(\phi)\psi + g(\phi)v(\psi)$$

In the Taft algebra, a is a skew-primitive w.r.t. u. Indeed we will see it defines a "quantum vector field" on a non-commutative torus.

14.4. Hopf-Modules

The idea of a group becomes concrete when we study its representations. The analogue for a Hopf-algebra is a Hopf-module.

Definition 85. An algebra A is a Hopf-module of the Hopf-algebra H if there is a linear map $H \otimes A \to A$ (denoted by $h \otimes \phi \to h(\phi)$) such that

$$h\tilde{h}(\phi) = h(\tilde{h}(\phi)), \quad h(1) = \epsilon(h)1, \quad h(\phi\psi) = h_{(1)}(\phi)h_{(2)}(\psi)$$

We are using Sweedler's notation here. The first condition is analogous the condition for a Lie algebra representation. The last condition is a generalization of the "Leibnitz rule" of differentiation.

14.4.1. *Example: A Hopf-module of the Taft algebra*

We seek an algebra A of matrices and a map $H \otimes A \to A$. The conditions they must satisfy are $h(\phi\psi) = h_{(1)}(\phi)h_{(2)}(\psi)$ when h is each generator u, a of H. If we expand this out, we get

$$u(\phi\psi) = u(\phi)u(\psi), \quad a(\phi\psi) = a(\phi)\psi + u(\phi)a(\psi)$$

If A is generated by z_1, z_2 and

$$u(\phi) = z_2\phi z_2^{-1}, \quad a(\phi) = z_1\phi - z_2\phi z_2^{-1}z_1$$

we can satisfy these conditions. The conditions on u^n and a^n can be satisfied if we choose

$$z_2z_1 = \omega z_1z_2, \quad z_1^n = 1, \quad z_2^n = 1, \quad \omega = e^{\frac{2\pi i}{n}}$$

This is a representation of the finite Heisenberg group, $Heis(Z_n)$ where the central generator takes the value of ω . We already know how to realize z_1 and z_2 as $n \times n$ matrices. It is not hard to see that any $n \times n$ matrix ϕ can be written as a "polynomial" in z_1 and z_2:

$$\phi = \sum_{k,l=0}^{n-1} \phi_{kl} z_1^k z_2^l, \quad \phi_{kl} \in \mathbb{C}.$$

Thus we have the Taft algebra realized as transformations on the algebra A of $n \times n$ matrices. The elements of A can be viewed as functions on a Non-Commutative Torus (NCT) with z_1 and z_2 as co-ordinates.Since the dimension of A is n^2, this NCT is a sort of non-commutative lattice with n^2 points.

14.5. $SL_q(2)$

We saw that the Taft algebra is a quantization of the two dimensional Lie algebra. There is also a Hopf-algebra that reduces in the classical limit to the Lie group of 2×2 matrices with determinant one. It is best to construct it as a symmetry of the quantum plane [31].

Let us start with two variables z_1, z_2 satisfying

$$z_2 z_1 = q z_1 z_2 \tag{14.5.1}$$

They are co-ordinates of the "quantum plane". Here, $q \in \mathbb{C}$ is a complex number[2]. The special case $q = e^{\frac{2\pi i}{n}}$ (where n is an integer greater than 2) will be of special interest. The special case $q = 1$ is the usual plane $\mathbb{C}^2$.

Suppose $g = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix}$ acts on z to get some new variables z', z'':

$$\begin{pmatrix} z_1' \\ z_2' \end{pmatrix} = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}, \quad \begin{pmatrix} z_1'' \\ z_2'' \end{pmatrix} = \begin{pmatrix} g_{11} & g_{21} \\ g_{12} & g_{22} \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$$

We are not to think of g_{11} etc, as complex numbers: they are some abstract quantities whose multiplication properties we are about to derive.

What conditions should be satisfied[3] by $g_{11}, g_{12}, g_{21}, g_{22}$ in order that z_1', z_2' as well as z_1'', z_2'' also satisfy the condition (14.5.1)? By direct calculation, they are

$$\begin{gathered} g_{12} g_{11} = q g_{11} g_{12}, \quad g_{22} g_{12} = q g_{12} g_{22} \\ g_{21} g_{11} = q g_{11} g_{21}, \quad g_{22} g_{21} = q g_{21} g_{22} \\ g_{12} g_{21} = g_{21} g_{22}, \quad g_{11} g_{22} - q^{-1} g_{12} g_{21} = g_{22} g_{11} - q g_{12} g_{21} \end{gathered} \tag{14.5.2}$$

[2]For reasons that will become clear later, we require that $q^2 \neq 1$.

[3]We are assuming that the elements of g commute with the elements of z.

These relations (14.5.2)imply that the "quantum determinant": $\det_q g = g_{11}g_{22} - q^{-1}g_{12}g_{21}$ is central; i.e., that it commutes with $g_{11}, g_{12}, g_{21}, g_{22}$. So we can set it to one:

$$g_{11}g_{22} - q^{-1}g_{12}g_{21} = 1. \tag{14.5.3}$$

The algebra defined by the relations (14.5.2,14.5.3) is called $SL_q(2)$. We can turn it into a bi-algebra with the co-product

$$\Delta(g_{11}) = g_{11} \otimes g_{11} + g_{12} \otimes g_{21}, \quad \Delta(g_{12}) = g_{11} \otimes g_{12} + g_{12} \otimes g_{22}$$
$$\Delta(g_{21}) = g_{21} \otimes g_{11} + g_{22} \otimes g_{21}, \quad \Delta(g_{22}) = g_{21} \otimes g_{12} + g_{22} \otimes g_{22}$$

etc. It has a co-unit given by

$$\epsilon(g_{11}) = 1 = \epsilon(g_{22}), \quad \epsilon(g_{12}) = 0 = \epsilon(g_{21}).$$

Also define

$$S(g_{11}) = g_{22}, \quad S(g_{12}) = -qg_{12}$$
$$S(g_{21}) = -q^{-1}g_{21}, \quad S(g_{22}) = g_{11}$$

This can be verified to satisfy the conditions of an antipode. The Hopf algebra $SL_q(2)$ is thus a symmetry of the quantum plane. That is, the quantum plane is a Hopf-module of $SL_q(2)$.

14.5.1. *Finite approximations to Lie groups*

We know that the Lie group $U(1)$ can be approximated by the finite sub-group Z_n: In the limit $n \to \infty$ the group algebra of $Z_n$tends to that of $U(1)$. This allows us to approximate the Fourier series on $U(1)$ by finite series. For simple Lie groups such as $SU(2)$ or $SL(2, C)$ there is no obvious analogue for this. The obvious idea of approximating by finite sub-groups fails: There are only a very small collection of such finite subgroups (e.g., the symmetries of the Platonic solids in the case of $SO(3)$).

But if we look at the larger category of quantum groups (allowing for co-products to be non-co-commutative) there is a way around this. We illustrate this with the example of $SL_q(2)$.

If q is a primitive nth root of unity we can impose additional relations that reduce the dimensions of the above algebras. For example, z_1^n and z_2^n are now central and we can impose

$$z_1^n = 1 = z_2^n.$$

This is the NCT we discussed earlier. Then $g_{11}^n, g_{22}^n, g_{12}^n, g_{21}^n$ are also central in $SL_q(2)$ and we can impose

$$g_{11}^n = 1 = g_{22}^n, \quad g_{12}^n = 0 = g_{21}^n$$

This reduces the dimension of $SL_q(2)$ to n^3. Recall that the dimension of a group algebra is the number of elements of the group (hence infinite for all Lie groups). This $SL_q(2)$ for q a primitive nth root of unity can be thought of as a "quantum group" with n^3 points.

By setting $q = e^{\frac{2\pi i}{n}}$, we get a sequence of finite dimensional Hopf algebras which tend to the Hopf algebra of functions on the group $SL(2, C)$ as $n \to \infty$.

To get finite approximations to compact Lie groups such as $SU(2)$ we need a generalization of the hermitian conjugate: An anti-linear map which preserves multiplication but reverses the co-multiplication.[4]

$$a^\dagger = d, \quad b^\dagger = -qc, \quad c^\dagger = -q^{-1*}b, \quad d^\dagger = a$$

Such finite approximations could be useful in numerical computations in lattice gauge theory [34].

[4]This differs slightly from [31] since we are allowing q to be a complex number.

This page intentionally left blank

Chapter 15

EULER-ARNOLD DYNAMICS

Let us begin with two important physical systems, both due to Euler.

15.1. The Rigid Body

A rigid body is one whose molecules remain at a constant distance between each other as the body as a whole moves. We will consider the case when there is no external force or torque on the body. So, the translational degree of freedom is uninteresting: The center of mass moves at a constant velocity. Its rotations can be quite intricate if the shape of the body is not symmetric, so that its three principal moments of inertia are mutually unequal. Euler discovered the equations that describe this rotation. It is a staple of classical mechanics courses [13]. But we will briefly review it to set the stage for the vastly more complicated examples to follow.

The absence of external torques mean that the angular momentum of the body is constant in time, as measured by an inertial observer. But, the angular velocity would not be constant as the moment of inertia (which is a symmetric matrix in general)varies with the orientation of the body. A simpler description can be obtained in the reference frame that is moving with the body. There is such a frame in which the moment of inertia is a diagonal matrix. Then the components of the angular momentum $\mathbf{L}$ and the angular velocity $\mathbf{\Omega}$ are proportional to each other :

$$\Omega_1 = h_1 L_1, \quad \Omega_2 = h_2 L_2, \quad \Omega_3 = h_3 L_3 \tag{15.1.1}$$

The constants h_1, h_2, h_3 are positive and are the inverses of the principal moments of inertia. L_1, L_2, L_3 depend on time, as they are measured in a reference frame that is rotating with the body, with an angular velocity $\mathbf{\Omega}$. So the total time

derivative, including the effect of this rotation, is zero:

$$\frac{d\mathbf{L}}{dt} + \mathbf{\Omega} \times \mathbf{L} = 0 \tag{15.1.2}$$

Combining (15.1.1,15.1.2) we get the Euler Equations for a rigid body

$$\frac{dL_1}{dt} + (h_2 - h_3)\, L_2 L_3 = 0, \quad \frac{dL_2}{dt} + (h_3 - h_1)\, L_3 L_1 = 0,$$
$$\frac{dL_3}{dt} + (h_1 - h_2)\, L_1 L_2 = 0 \tag{15.1.3}$$

When $h_1 = h_2 = h_3$ these equations simply say that $\mathbf{L}$ is a constant: The angular momentum is constant even in the rotating reference frame. When just one pair of the parameters are equal (say $h_1 = h_2 \neq h_3$) a component of (here L_3) of angular momentum is constant. The remaining components of $\mathbf{L}$ and $\mathbf{\Omega}$ are then expressible as trigonometric functions of time.

This is a good approximation for the rotation of the Earth. Because of the bulge at the Equator, the moments of inertia are not all equal. But the shape is, to a good approximation, circularly symmetric: A pair of the moments of inertia are equal. The axis of rotation of the earth itself precesses around a fixed direction, with a period of about 26,000 years. Ancient astronomers (e.g., Aryabhatta of 5th century AD) knew of this precession. Newton himself gave the first mechanical explanation.

Jacobi solved the general case of the Euler equations when h_1, h_2, h_3 are mutually unequal. This depends on the conservation of the magnitude of angular momentum

$$\mathbf{L}^2 = L_1^2 + L_2^2 + L_3^2$$

even in this general case. In addition, the energy

$$H = \frac{1}{2}\mathbf{\Omega} \cdot \mathbf{L} = \frac{1}{2}(h_1 L_1^2 + h_2 L_2^2 + h_3 L_3^2)$$

is also conserved.

Exercise 86. Directly verify that $\frac{d\mathbf{L}^2}{dt} = 0 = \frac{dH}{dt}$ using the Euler equations of a rigid body.

The solution $\mathbf{L}(t)$ describes a curve in $\mathbb{R}^3$. This curve is the intersection of the sphere given by constant $\mathbf{L}^2$ and the ellipsoid on which H is a constant. Jacobi

found the solution explicitly in terms of his elliptic functions

$$L_1(t) = A_1\mathrm{cn}(\nu t), \quad L_2(t) = A_2\mathrm{sn}(\nu t), \quad L_3(t) = A_3\mathrm{dn}(\nu t)$$

These elliptic functions satisfy differential equations that generalize those of the trigonometric functions

$$\frac{d}{dz}\mathrm{sn}(z) = \mathrm{cn}(z)\mathrm{dn}(z), \quad \frac{d}{dz}\mathrm{cn}(z) = -\mathrm{sn}(z)\mathrm{dn}(z), \quad \frac{d}{dz}\mathrm{dn}(z) = -m\mathrm{cn}(z)\mathrm{sn}(z)$$

They also depend on a parameter m called the modulus.

Exercise 87. Find the relation of the constants A_1, A_2, A_3, m and ν to h_1, h_2, h_3 and the two conserved quantities.

More details are in [13].

15.2. Euler Equations of a Fluid

There is another equation in physics also named for Euler. It describes the flow of an ideal incompressible fluid. Unlike the rigid body, this is very far from being exactly solvable. It exhibits a virulent, as yet mysterious, kind of chaos called turbulence. Understanding turbulence is on everyone's list of the most important problems of physics.

The equations of motion of a fluid can be derived [35] from the mechanics of a fluid element[1]:

$$\mathrm{div}\ \mathbf{v} = 0, \quad \frac{\partial \mathbf{v}}{\partial t} + \mathbf{v}\cdot\nabla\mathbf{v} = -\nabla p \tag{15.2.1}$$

Here

- $\mathbf{v}$ is the velocity of the fluid, which depends on position and time.
- The first condition is the conservation of mass under the assumption that the density is constant. (This is the meaning of incompressibility.)
- The second condition is Newton's law for a fluid element. The left hand side is the acceleration of a fluid element. In addition to the explicit time derivative, it has a term describing the acceleration of a small fluid element being carried along by the fluid flow itself.

[1] A fluid element is a region whose size is small compared to the vessel containing the fluid but large compared to the distance between molecules. Fluid mechanics is the effective field theory of particle mechanics in this approximation.

- p is the pressure divided by density. Even in the absence of external forces (e.g., gravity) the fluid has to exert some forces on itself to keep the density constant. An equivalent point of view is that p is the Lagrange multiplier that enforces the condition of incompressibility.

We can eliminate p from the equations by taking a curl. It is useful to think of the fluid flow in terms of vorticity:

$$\omega = \text{curl } \mathbf{v}$$

Given appropriate boundary conditions this equation can be inverted

$$\mathbf{v} = \text{curl}^{-1}\omega.$$

curl^{-1} is an integral operator called Biot-Savart operator[2].

Remark 88. These boundary conditions on $\mathbf{v}$ usually say that the normal component of $\mathbf{v}$ vanishes at the boundary. In other words, that the fluid does not cross the boundary. We will find it simpler to assume the fluid fills the whole of $\mathbb{R}^3$ with the "boundary condition" that velocity and vorticity vanish at infinity faster than any power.

The Euler equation in vorticity form is

$$\frac{\partial \omega}{\partial t} + [\mathbf{v}, \omega] = 0. \tag{15.2.2}$$

Here

$$[\mathbf{u}, \mathbf{w}] = \mathbf{u} \cdot \nabla \mathbf{w} - \mathbf{w} \cdot \nabla \mathbf{u}$$

is the commutator of vector fields.

Exercise 89. Derive (15.2.2) from (15.2.1). The following identity will be useful:

$$\text{curl } (\mathbf{w} \times \mathbf{u}) = \mathbf{w}\text{div } \mathbf{u} - \mathbf{u}\text{div } \mathbf{w} + [\mathbf{u}, \mathbf{w}]$$

Show also that the commutator of two vector fields with zero divergence again has zero divergence.

[2]It also arises in magnetism, relating a magnetic field to its current source.

15.2.1. *Energy of a fluid*

The two assumptions we made

- there are no external forces (such as gravity)
- the fluid is incompressible

imply that all the energy of the fluid is kinetic. With density set to one

$$H = \frac{1}{2}\int \mathbf{v}^2 dx.$$

Exercise 90. Verify by direct calculation that $\frac{dH}{dt} = 0$.

15.2.2. *Helicity*

There is another conserved quantity, called Helicity:

$$\mathcal{H} = \int \mathbf{v}\cdot\boldsymbol{\omega}\, dx, \quad \text{Helicity}$$

The proof of its conservation is straightforward enough:

$$\frac{d}{dt}\mathcal{H} = -2\int \mathbf{v}.[\mathbf{v},\boldsymbol{\omega}]dx = 2\int \mathbf{v}.\text{curl}(\mathbf{v}\times\boldsymbol{\omega})dx = 2\int \boldsymbol{\omega}.(\mathbf{v}\times\boldsymbol{\omega})dx = 0.$$

In the last step we do an integration by parts using the identity

$$\text{div}\,(\mathbf{a}\times\mathbf{b}) = (\text{curl}\,\mathbf{a})\cdot\mathbf{b} - (\text{curl}\,\mathbf{b})\cdot\mathbf{a} \tag{15.2.3}$$

Traditionally, it is given a topological interpretation [36,37] . But in our context, it is more appropriate to understand it as an invariant of the Lie algebra V.

For more on Fluid Mechanics see [35, 37].

15.3. Euler–Arnold Equations on a Metric Lie Algebra

Now we will describe a theory of Arnold [37] which unifies the above two examples into a single framework. It addition it gives us a host of examples of intermediate complexity. Since turbulence is such a hard problem, it is useful to have these simple examples. We will study one of them in more detail later.

Let V be a real Lie algebra and G a positive inner product[3] on it. That is,

$$G : V\times V \to \mathbb{R}$$

[3]We will not assume that G is invariant. In fact the interesting cases are precisely the non-invariant ones.

is a symmetric bilinear on S. Positive means that

$$G(u,u) \geq 0$$

and that

$$G(u,u) = 0 \implies u = 0.$$

This means that there is an invertible linear map (also denoted by G to save on notation)

$$G : V^* \to V$$

where V^* is the dual vector space[4] of V.

Recall that every Lie algebra has a representation on itself, the adjoint representation:

$$\mathrm{ad}_u(w) = [u,w], \quad u,w \in V$$

Therefore it also has a representation on its dual, the co-adjoint representation

$$\mathrm{ad}^*_u\alpha(w) = \alpha\left([u,w]\right), \quad u,w \in V, \quad \alpha \in V^*$$

Definition 91. A dynamical system on V^* evolving in time according to the differential equation

$$\frac{d\omega}{dt} + \mathrm{ad}^*_{G\omega}\omega = 0$$

is called an Euler-Arnold system.

Thus, if V is abelian or if G is invariant, this dynamics would be trivial: ω would be time-independent.

15.3.1. *Index notation*

It is useful to express these ideas more explicitly, in a basis $e_a \in V$ and a dual basis $e^b \in V^*$; i.e.,

$$e^b(e_a) = \delta^b_a$$

We will assume for the moment that V is finite dimensional, although the general theory makes sense even for infinite dimensional cases. The Lie algebra

[4] That is, V^* is the space of real-valued linear functions on V.

structure constants f^c_{ab} are defined by

$$[e_a, e_b] = f^c_{ab} e_c$$

Then the positive inner product defines a symmetric matrix

$$G_{ab} = G(e_a, e_b)$$

with an inverse G^{ab} defined by

$$G^{ab} G_{bc} = \delta^a_c.$$

The adjoint representation is, in component form

$$[\mathrm{ad}_u w]^a = f^a_{bc} u^b w^c, \quad u, w \in V$$

Given $\alpha \in V^*$ the co-adjoint representation is determined by the condition

$$\left[\mathrm{ad}^*_u \alpha\right]_a w^a + \alpha_a \left[\mathrm{ad}_u w\right]^a = 0$$

That is

$$\left[\mathrm{ad}^*_u \alpha\right]_a w^a + \alpha_c f^c_{ba} u^b w^a = 0$$

so that

$$\left[\mathrm{ad}^*_u \alpha\right]_a = -\alpha_c f^c_{ba} u^b$$

Thus

$$\left[\mathrm{ad}^*_{G\omega} \omega\right]_a = -\omega_c f^c_{ba} G^{bd} \omega_d = G^{bd} f^c_{ab} \omega_c \omega_d.$$

The Euler-Arnold equation is then

$$\frac{d\omega_a}{dt} + G^{bd} f^c_{ab} \omega_c \omega_d = 0$$

This can be thought of as a mechanical system with hamiltonian

$$H = \frac{1}{2} G^{bd} \omega_b \omega_d$$

and Poisson brackets

$$\{\omega_a, \omega_b\} = f^c_{ab} \omega_c$$

for the dynamical variables. In particular, the hamiltonian is a conserved. So these are ideal systems, where there is no dissipation of energy: Frictional effects such as viscosity are ignored. There may be symmetric invariant tensors (Casimir elements) in the Lie algebra which can lead to additional conserved quantities.

If there is an invariant inner product

$$\langle e_a, e_b \rangle = \eta_{ab}$$

we would have

$$\langle e_a, [e_b, e_c] \rangle + \langle [e_b, e_a], e_c \rangle = 0$$

which means

$$f^d_{bc}\eta_{ad} + f^d_{ba}\eta_{cd} = 0.$$

The inverse matrix η^{ae} defines an inner product on V^*. By multiplying the above equation by $\eta^{ce}\eta^{af}$ we get

$$\eta^{ce}\eta^{af}\left\{f^d_{bc}\eta_{ad} + f^d_{ba}\eta_{cd}\right\} = 0 \iff$$

$$\eta^{ce} f^f_{bc} + \eta^{af} f^e_{ba} = 0.$$

Switching indices $f \to c, c \to a$, in the first term and just $f \to c$ in the second

$$\eta^{ae} f^c_{ba} + \eta^{ac} f^e_{ba} = 0. \tag{15.3.1}$$

The quantity

$$C = \eta^{ab}\omega_a\omega_b$$

is then a conserved quantity. For,

$$\frac{dC}{dt} = -2\eta^{ae}\{G^{bd} f^c_{ab}\omega_c\omega_d\}\omega_e = 2G^{bd}\eta^{ae} f^c_{ba}\omega_c\omega_d\omega_e$$

We now symmetrize in e, c:

$$= G^{bd}\{\eta^{ae} f^c_{ba} + \eta^{ac} f^e_{ba}\}\omega_c\omega_d\omega_e$$

which is zero by the above identity (15.3.1).

15.3.2. *Special case: The rigid body*

In this case the Lie algebra is $so(3)$. The Lie bracket is the cross product of vectors in $\mathbb{R}^3$. The angular momentum is the dynamical variable $L_a \equiv \omega_a$. The Moment of inertia matrix is the matrix G_{ab} of the inner product. Its inverse relates angular velocity to angular momentum:

$$\Omega^a = G^{ab} L_b$$

Thus the energy of the rigid body is

$$H = \frac{1}{2} G^{ab} L_a L_b.$$

The quantities h_1, h_2, h_3 are the eigenvalues of G^{ab} (i.e., inverses of the eigenvalues of G_{ab}, the moment of inertia).

The extra conserved quantity L^2 arises from the fact the dot product

$$\mathbf{u} \cdot \mathbf{v} = \delta_{ab} u^a v^b$$

is an invariant inner product in the Lie algebra[5]. Equivalently, $\mathbf{L}^2 = L_a L_b \delta^{ab}$ has zero Poisson brackets with all the components of L_a.

$$\left\{\mathbf{L}^2, L_a\right\} = 0.$$

Because of this invariant inner product, we can identify $so(3)$ with its dual space. In the index notation this is obvious as the covariant and contra-variant components are the same because of the Kronecker delta. So in this case we can think of Ω and L as belonging to $so(3)$.

The structure constants are completely anti-symmetric with $f_{123} = 1$. Thus the Euler-Arnold equations become

$$\frac{dL_a}{dt} + \sum_b h_b f_{abc} L_c L_b = 0$$

For $a = 1$ this reduces to

$$\frac{dL_1}{dt} + (h_2 - h_3)\omega_2 \omega_3 = 0$$

as needed. The remaining equations are cyclic permutations of this.

15.3.3. *Special case: The incompressible fluid*

Vector fields satisfying the condition of incompressibility

$$\mathrm{div} u = 0$$

[5]Thus there are two inner products of interest here: G is not invariant and gives the energy. The dot product is invariant and gives another conserved quantity.

form an infinite dimensional[6] Lie algebra with the Lie bracket $[\mathbf{u}, \mathbf{w}] = \mathbf{u} \cdot \nabla \mathbf{w} - \mathbf{w} \cdot \nabla \mathbf{u}$. (Recall the exercise showing that the commutator of two incompressible vector fields is again incompressible). We will call it S. (There are analogous vector fields on any manifold with a volume form.)

At constant density, (which we choose to be one by a choice of units) the total energy of the fluid is

$$H = \frac{1}{2} \int \mathbf{v}^2 dx$$

15.3.3.1. *Helicity as an invariant inner product*

Let us define a bilinear

$$\langle \mathbf{u}, \mathbf{w} \rangle = \int \mathbf{u} \cdot \mathrm{curl}^{-1} \mathbf{w} dx, \quad \mathbf{u}, \mathbf{w} \in S$$

where curl^{-1} is the Biot–Savart operator we mentioned earlier. The first thing to note is that this is symmetric.

If we put[7]

$$\mathbf{u} = \mathrm{curl}\, \mathbf{a}, \quad \mathbf{w} = \mathrm{curl}\, \mathbf{b}, \quad \mathrm{div}\, \mathbf{a} = 0 = \mathrm{div}\, \mathbf{b}$$

These $\mathbf{a}, \mathbf{b}$ are unique due to the invertibility of curl. Then,

$$\langle \mathbf{u}, \mathbf{w} \rangle = \int (\mathrm{curl}\, \mathbf{a}) \cdot \mathbf{b} dx$$

so that

$$\langle \mathbf{u}, \mathbf{w} \rangle - \langle \mathbf{w}, \mathbf{u} \rangle = \int \{(\mathrm{curl}\, \mathbf{a}) \cdot \mathbf{b} - (\mathrm{curl}\, \mathbf{b}) \cdot \mathbf{a}\} dx$$

This is a total divergence because of the identity (15.2.3).

So, the anti-symmetric part is a surface integral which vanishes for our boundary conditions.

Due to the invertibility of curl in S, we can see that this bilinear is non-degenerate:

$$\langle \mathbf{u}, \mathbf{w} \rangle = 0, \quad \forall \mathbf{w} \in S \implies \mathbf{u} = 0.$$

Thus $\langle \mathbf{u}, \mathbf{w} \rangle$ is an inner product on S.

[6] Any such vector field is determined by two independent functions: Three components satisfying one condition. Since there are an infinite number of linearly independent functions, the space of incompressible vector fields is infinite dimensional.

[7] Recall that the curl of. a vector field is unchanged if we add a gradient to it. We impose the condition of zero divergence on $\mathbf{a}$ and $\mathbf{b}$ to remove this ambiguity.

It is however, not positive. For example, parity reverses its sign. (Recall that curl $\mathbf{u}$ contains the cross-product which changes sign under parity).

Next we will see that $\langle \mathbf{u}, \mathbf{w} \rangle$ is an invariant inner product:

$$\langle [\mathbf{s}, \mathbf{u}], \mathbf{w} \rangle + \langle \mathbf{u}, [\mathbf{s}, \mathbf{w}] \rangle = 0, \quad \forall \mathbf{s}, \mathbf{u}, \mathbf{w} \in S.$$

For,

$$[\mathbf{s}, \mathbf{w}] = \operatorname{curl}(\mathbf{w} \times \mathbf{s})$$

so that

$$\langle \mathbf{u}, [\mathbf{s}, \mathbf{w}] \rangle = \int \mathbf{u} \cdot (\mathbf{w} \times \mathbf{s}) dx.$$

By the anti-symmetry of the triple product and the symmetry of $\langle , \rangle$,

$$\langle \mathbf{u}, [\mathbf{s}, \mathbf{w}] \rangle = - \int \mathbf{w} \cdot (\mathbf{s} \times \mathbf{u}) dx = -\langle \mathbf{w}, [\mathbf{s}, \mathbf{u}] \rangle = -\langle [\mathbf{s}, \mathbf{u}], \mathbf{w} \rangle$$

proving the result.

Thus $\langle \mathbf{u}, \mathbf{w} \rangle$ is analogous to the dot product in $so(3)$. Helicity is the special case

$$\mathcal{H} = \langle \boldsymbol{\omega}, \boldsymbol{\omega} \rangle$$

which is analogous to $\mathbf{L} \cdot \mathbf{L}$. Its conservation is thus, a direct consequence of the invariance of the inner product.The main difference with $\mathbf{L}^2$ is that helicity is not positive.

15.3.3.2. *Energy as a positive inner product*

Expressed in terms of vorticity,

$$H = \frac{1}{2} \int \operatorname{curl}^{-1} \boldsymbol{\omega} \cdot \operatorname{curl}^{-1} \boldsymbol{\omega} dx$$

After an integration by parts this can be written in terms of the Green's function of the Laplace operator on vector fields on $\mathbb{R}^3$.

$$H = \frac{1}{2} \int \omega_i(x) G_{ij}(x, y) \omega_j(y) dx dy$$

This is a positive inner product on S^* which is however, not invariant.

We can also write this in terms of the invariant inner product above as

$$H = \frac{1}{2} \langle \operatorname{curl}^{-1} \omega, \omega \rangle$$

which is analogous to the formula $H = \frac{1}{2} \boldsymbol{\Omega} \cdot \mathbf{L}$ for a rigid body.

15.4. Euler–Arnold Dynamics on $SO(3,1)$

When physicists are confronted with a problem that is too hard to solve, we seek a simpler "toy model" which captures some of the essential aspects of the original problem. The rigid body is not a useful toy model for fluids, because it is exactly solvable. It does not exhibit the essential chaotic phenomena. We seek an example which is close to being exactly solvable (so that we have hopes of understanding it) but has some chaos as well. We would also like it to have an invariant inner product (analogous to helicity) which is not positive. This suggests a non-compact Lie algebra. It would be better for it to be simple, as the Lie algebra of incompressible vector fields does not seem to have any ideals.

The best choice appears to be $SO(3,1)$, the Lorentz group. It is familiar to us from studying relativistic wave equations: A completely unrelated use of the same group. Thought of as a real Lie algebra it is six dimensional.

$SO(3,1)$ is a rank two Lie algebra, only one step more complicated than $SO(3)$ of the rigid body.Other familiar rank two Lie algebras such as $so(4) \sim so(3) \oplus so(3)$ and $so(2,2) \sim so(1,2) \oplus so(1,2)$ are not simple.

15.4.1. *The Lie algebra*

Let us begin by choosing a basis in $so(3,1)$. The Minkowski metric on $\mathbb{R}^{3,1}$ is[8]

$$\eta = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

$so(3,1)$ is the vector space of anti-symmetric 4×4 matrices with the Lie bracket

$$[X, Y] = X\eta Y - Y\eta X.$$

The Minkowski metric on $\mathbb{R}^{1,3}$ induces not one, but two, invariant inner product on the Lie algebra $so(3,1)$:

$$\langle X, Y\rangle_1 = \frac{1}{4}\epsilon^{\mu\nu\rho\sigma} X_{\mu\nu} Y_{\rho\sigma}, \quad \langle X, Y\rangle_2 = \frac{1}{4}\left(\eta^{\mu\rho}\eta^{\nu\sigma} - \eta^{\mu\sigma}\eta^{\nu\rho}\right) X_{\mu\nu} Y_{\rho\sigma}$$

[8] We chose a different sign convention when we discussed wave equations earlier, because it is more common in particle physics.

Here $\epsilon^{\mu\nu\rho\sigma}$ is the completely anti-symmetric tensor with $\epsilon^{0123} = 1$. $\langle X, Y\rangle_1$ changes sign under parity, but it is still invariant under the connected component of the Lorentz group.

15.4.2. *A basis*

A basis in $so(3,1)$ is the set of six matrices

$$\alpha_1 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}, \quad \dots, \quad \alpha_4 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \cdots$$

The $\cdots$ denote the cyclic permutations $1 \to 2 \to 3 \to 1$, $4 \to 5 \to 6 \to 4$. That the components of $so(3,1)$ can be split into two groups of three, with a cyclic symmetry Z_3 in each subset is a common theme in the calculations below.The inner products are brought to standard forms (for $a, b = 1, \cdots 6$) in this basis[9]:

$$\langle \alpha_a, \alpha_b\rangle_1 = \begin{pmatrix} 0_3 & 1_3 \\ 1_3 & 0_3 \end{pmatrix}, \quad \eta_{ab} \equiv \langle \alpha_a, \alpha_b\rangle_2 = \begin{pmatrix} 1_3 & 0_3 \\ 0_3 & -1_3 \end{pmatrix}$$

The structure constants are defined by

$$[\alpha_a, \alpha_b] = f^c_{ab}\alpha_c \iff f^c_{ab} = \eta^{cd}\langle [\alpha_a, \alpha_b], \alpha_d\rangle_2$$

It is not hard to compute them explicitly in Mathematica; we do not see the need to list them as a table.

15.4.3. *Aside: Embedding so(3,1) in S*

We note in passing that $so(3,1)$ is a Lie-sub-algebra of S, the Lie algebra of incompressible vector fields. Thus, in some sense, we are studying a sub-system of fluid mechanics. However, the metric of S does not split as a direct sum of a metric in $so(3,1)$ and S: Solutions of the Euler-Arnold equations on $so(3,1)$ are not special solutions of the Euler equations for the fluid.

To see that $so(3,1) \subset S$ we need to find a set of six incompressible vector fields that satisfy the commutation relations of $so(3,1)$. Fact familiar to particle physicists will help us do this. The Lorentz group acts on the time-like hyperboloid;

[9]We abuse notation slightly here. What is meant is that the inner product $\langle \alpha_a, \alpha_b\rangle_1$ is the a, b component of the matrix on the RHS.

this is the "mass-shell" of a particle of unit mass and positive energy. The Lorentnz invariant volume measure is

$$\frac{d^3\mathbf{p}}{\sqrt{1+\mathbf{p}^2}} = 2\delta(p_0^2 - \mathbf{p}^2 = 1)\theta(p_0 > 0)dp_0 d^3\mathbf{p}$$

Knowing how $so(3,1)$ acts on vectors in $\mathbb{R}^{3,1}$ the basis above translates to

$$e_1 = -p_3\frac{\partial}{\partial p_2} + p_2\frac{\partial}{\partial p_3}, \ldots e_4 = \sqrt{1+\mathbf{p}^2}\frac{\partial}{\partial p_1}, \ldots$$

where again $\cdots$ denotes cyclic permutations over $1, 2, 3$.

Now, the hyperboloid is diffeomorphic to $\mathbb{R}^3$ as a manifold. The change of variables

$$\mathbf{x} = \frac{s(|\mathbf{p}|)}{|\mathbf{p}|}\mathbf{p}$$

with

$$\frac{s^3(|\mathbf{p}|)}{3} = \frac{1}{2}\left(|\mathbf{p}|\sqrt{1+|\mathbf{p}|^2} + \log\left[-|\mathbf{p}| + \sqrt{1+|\mathbf{p}|^2}\right]\right).$$

maps the volume $\frac{d^3\mathbf{p}}{\sqrt{1+\mathbf{p}^2}}$ to $d^3\mathbf{x}$. The expressions for the vector fields e_a in these coo-ordinates is quite messy. But we don't need them, as commutation relations are unchanged under such co-ordinate transformations.

15.4.4. *The Poisson brackets*

The Poisson brackets among the components ξ_a of $\xi \in so(3,1)$ (in the basis above) are

$$\{\xi_a, \xi_b\} = f^c_{ab}\xi_c.$$

These brackets are degenerate: There are functions (arising from the two invariant inner products)

$$C_1(\xi) = \xi_1\xi_4 + \xi_2\xi_5 + \xi_3\xi_6, \quad C_2(\xi) = \frac{1}{2}\left(\xi_1^2 + \xi_2^2 + \xi_3^2 - \xi_4^2 - \xi_5^2 - \xi_6^2\right)$$

that commute with every function of ξ_a. They are constants of motion and are therefore determined by the initial conditions.

15.4.5. *The phase space*

In the case of the rigid body, the constraint $\mathbf{L}^2 =$ constant picks out a two dimensional sphere. The dynamics of the rigid body takes places on this space: It is the phase space.[10] In our case, we can take an element $\xi \in so(3,1)$ and hold fixed two quantities C_1, C_2 coming from the two invariant metrics:

The resulting four-dimensional manifold is the phase space of our system. Thus we study a hamiltonian system with two degrees of freedom: The minimum needed to have chaos. If we have one more conserved quantity (in addition to H, C_1, C_2) our system would be integrable.

The geometric shape of the phase space depends on the values of the constants C_1 and C_2. The simplest choice $C_1 = 0 = C_2$ corresponds to a kind of cone.

15.4.6. *The hamiltonian and the equations of motion*

The hamiltonian should be a quadratic function of the components of ξ. For simplicity, we will choose this to be diagonal quadratic form (in the basis above)

$$H(h) = \frac{1}{2}\sum_{a=1}^{6} h_a \xi_a^2.$$

The six parameters h_a, along with the Poisson bracket relations

$$\{\xi_a, \xi_b\} = f_{ab}^c \xi_c$$

determine the equations of motion.

$$\frac{d\xi_1}{dt} = \xi_2\xi_3\,(h_2 - h_3) + \xi_5\xi_6\,(h_5 - h_6)\,, \quad \cdots$$

$$\frac{d\xi_4}{dt} = \xi_3\xi_5\,(-h_3 - h_5) + \xi_2\xi_6\,(h_2 + h_6)\cdots$$

The $\cdots$ denote the permutations $1 \to 2 \to 3 \to 1,\ 4 \to 5 \to 6 \to 4$ as usual.

[10] More geometrically, $\mathbb{S}^2$ is a co-adjoint orbit of the rotation group with a standard symplectic structure given by a construction of Kirillov. But we do not need this more abstract perspective.

15.4.7. *Commuting hamiltonians*

Now we use a remarkable discovery[11] of [38–40]. If a certain function of the parameters h_a is zero, the system is integrable. So the magnitude of this quantity is a measure of the chaos in the system.

Theorem 92. *If* $K(h) \equiv (h_1h_2 - h_3h_6)(h_4 - h_5) + \cdots = 0$ *the system is integrable*

The original proof of [39, 40] uses the idea of a Lax pair, a technique popular in the field of integrable systems. We will describe a more direct proof, which may be along the lines of [38]. This price for the directness is more calculations, which can be done in Mathematica.

Proof. We can calculate the Poisson Brackets $\{H(h), H(h')\}$ for two different choices h, h' of the parameters. The result is a cubic polynomial in the six variables ξ_a. By inspection, there are four conditions for it to vanish:

$$h_1h_2' - h_2h_1' + \cdots = 0$$
$$h_1(h_6' - h_5') - (h_6 - h_5)h_1' + h_5h_6' - h_6h_5' = 0, \cdots$$

The first equation is a single condition. The second condition its cyclic permutations give the remaining ones.

A little algebra (again in Mathematica) shows that if there are six parameters Y_a such that

$$h = \left(\frac{Y_5 - Y_6}{Y_2 - Y_3}, \frac{Y_6 - Y_4}{Y_3 - Y_1}, \frac{Y_4 - Y_5}{Y_1 - Y_2}, -\frac{Y_4}{Y_1}, -\frac{Y_5}{Y_2}, -\frac{Y_6}{Y_3}\right)$$
$$h' = \left(\frac{Y_5' - Y_6'}{Y_2 - Y_3}, \frac{Y_6' - Y_4'}{Y_3 - Y_1}, \frac{Y_4' - Y_5'}{Y_1 - Y_2}, -\frac{Y_4'}{Y_1}, -\frac{Y_5'}{Y_2}, -\frac{Y_6'}{Y_3}\right)$$

these conditions are satisfied, for **any** choice of Y'. Then there would be another conserved quantity $H(h')$, making the system integrable.

Thus the question reduces to: " when is there such Y_a"?

[11] They were studying a closely related system on $so(4)$; a few tweaks of signs is all that is needed to use there work for our purpose.

We can rewrite the first set of these equations as a system of linear equations for Y

$$\begin{pmatrix} 0 & h_1 & -h_1 & 0 & -1 & 1 \\ -h_2 & 0 & h_2 & 1 & 0 & -1 \\ h_3 & -h_3 & 0 & -1 & 1 & 0 \\ h_4 & 0 & 0 & 1 & 0 & 0 \\ 0 & h_5 & 0 & 0 & 1 & 0 \\ 0 & 0 & h_6 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{pmatrix} = 0$$

So, the condition is that the matrix above must have zero determinant. This determinant is

$$K(h) \equiv (h_1 h_2 - h_3 h_6)(h_4 - h_5) + \cdots$$

Which proves what we claimed. □

15.4.8. *Numerical examples*

We still do not know if the system is chaotic when $K \neq 0$. It is hard to prove that a system is chaotic; indeed there isn't even a precise definition of what chaos is. There is a quantity due to Kolmogorov and Sinai measuring the rate of entropy production which comes close to quantifying chaos. Here we have a system where such ideas can be tried out. For now, the test for chaos is,[12]

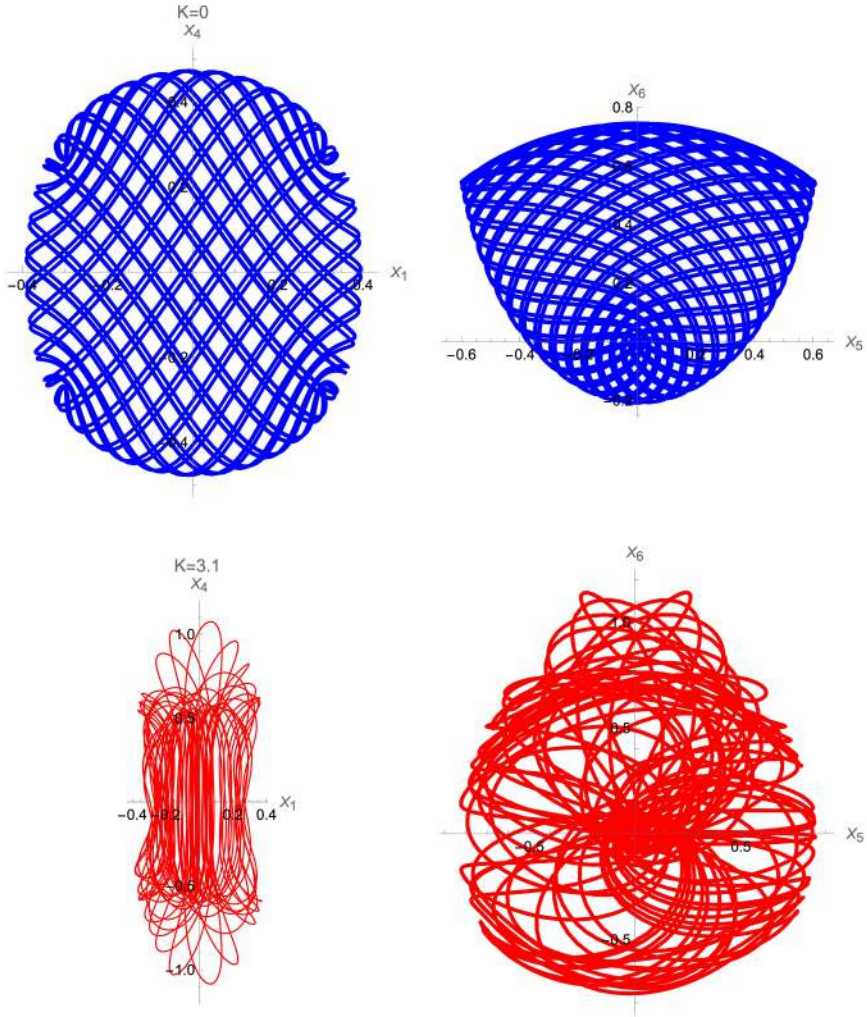
> "I know it when I see it."

It is not difficult to solve the six ODEs for different choices ξ_a and h_a. We can choose K, C_1, C_2 and choose the remaining parameters randomly to generate many examples.

By plotting these solutions,[13] it is obvious to the eye that the cases $K = 0$ (in Blue in the diagram) and $K = 3.1$ (in Red) look very different, even with exactly the same initial conditions.

[12] This is a quote of Justice Potter of the US Supreme Court. He was talking about something else.

[13] To visualize a curve in six dimensions is not easy. We have chosen to plot the projection to two different planes.



Exercise 93. Write a Mathematica program to (1) pick ξ_a and h_a at random for agiven choice of K, C_1, C_2; (2) Solve the equations of motion numerically for these choices (3) Plot the solutions in different projections. By generating many examples this way, understand the difference between $K = 0$ and $K \neq 0$.

More work is needed to bring ideas from statistical physics to study this system in detail: Markov Chain Monte Carlo looks like a promising idea to apply here.

BIBLIOGRAPHY

[1] D. Mumford, *Indra's Pearls: The Vision of Felix Klein,* Cambridge University Press (2015)

[2] O. Nash, *On Klein's Icosahedral Solution of the Quintic*, arXiv:1308.0955v1 [math.AG]

[3] M. Hausner and J. T. Schwartz *Lie Groups and Lie Algebras*, Gordon & Breach Science Publishers, Inc. (1968)

[4] C. Reuttenauer, *Free Lie Algebras,* Clarendon Press (1993)

[5] J. J. Rotman, *Advanced Modern Algebra,* Prentice-Hall (2002)

[6] J. Munkres, *Topology,* Pearson (2024)

[7] R. Shankar, *Principles of Quantum Mechanics*, Plenum (1980)

[8] P. A. M. Dirac, *Principles of Quantum Mechanics*, Cambridge (1959)

[9] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics,* Pergamon Press (1977)

[10] L. O'Raifeartaigh, *Group Structure of Gauge Theories*, (Section 9.4) Cambridge (1986)

[11] R. P. Feynman *Lectures in Physics* Vol. 3, Chapter 12. https://www.feynmanlectures.caltech.edu/III_12.html

[12] D. J. Griffiths, *Am. J. Phys.* **50**, 698 (1982)

[13] S. G. Rajeev, *Advanced Mechanics*, Oxford (2013)

[14] D. Shale Trans. *Amer. Math. Soc.*, **103**, 149–167 (1962)

[15] C. Itzykson and J-B. Zuber, *Quantum Field Theory,* Dover (2006)

[16] G. B. Folland, *Harmonic Analysis in Phase Space*, Princeton (1989)

[17] V. Bargmann, *Ann. Math.* **48**, 568–540 (1947)

[18] J. R. Klauder and E.C.G. Sudarshan *Fundamentals of Quantum Optics,* Dover (2006)

[19] H.Georgi, *Lie Algebras in Particle Physics,* Benjamin-Cummings (1982)

[20] K. Huang, *Statistical Mechanics*, 2nd Ed., Wiley (1987)

[21] M. L. Mehta, *Random Matrices,* Academic, 3rd Ed., Press (2004)

[22] V. Kargin, *Lecture Notes on Free Probability*, arxiv:1305.2611v1

[23] P. Deift and D. Gioev, Random Matrix Theory: Invariant Ensembles and Universality, *Amer. Math. Soc.* (2009)

[24] T. Ceccherini-Silberstein, F. Scarabotti and F. Tolli, *Harmonic Analysis on Finite Groups,* Cambridge (2008)

[25] A. Terras, *Fourier Analysis on Finite Groups and Applications*, Cambridge University Press, Cambridge 1999
[26] D. Bump, P. Diaconis, A. Hicks, L. Miclo and H. Widom, Ann. de la faculté des sciences de Toulouse, XXVI, pp. 263–288 (2017) [arxiv:1502.04160]
[27] E. M. Stein and R. Shakarchi, *Fourier Analysis: An Introduction*, Princeton (2003)
[28] R. Courant and D. Hilbert *Methods of Mathematical Physics, Vol. 1* Wiley-VCH (1989)
[29] P. Etingof, *Lie Groups and Lie Algebras,* MIT Open Courseware, https://ocw.mit.edu/courses/18-755-lie-groups-and-lie-algebras-ii-spring-2024/
[30] L. Faddeev and Takhtajan
[31] C. Kassel, *Quantum Groups*, Springer-Verlag (1995)
[32] Taft, Earl J. The order of the antipode of finite-dimensional Hopf algebra, *Proc. Natl. Acad. Sci.* U.S.A., **68**, 2631–2633, (1971)
[33] V. Chari and A. Pressley, *A Guide to Quantum Groups*, Cambridge (1994)
[34] T. V. Zache, D. Gonzalez-Cuadra, P. Zoller, *Phys. Rev. Lett.* 131(17), 71902 (2023),[arxiv:2304.02527].
[35] S. G. Rajeev, *Fluid Mechanics*, Oxford (2018)
[36] H. K. Moffat, J. *Fluid Mech*. **35,** 117–129 (1969)
[37] V. I. Arnold and B. Khesin, *Topological Methods in Hydrodynamics*, Springer (2021).
[38] L. A. Dikii Funkts. Anal. Prilozhen. **6**(4), 83–84 (1972)
[39] S. V. Manakov *Funct. Anal. Appl.* **10**, 328–329 (1977)
[40] M. Adler and P. van Moerbeke, *Invent. Math.* **67**, 297–331 (1982)

INDEX