Structure and Symmetry in Particle Physics

An Introduction For Students Interested in Research

Christopher B. Verhaaren

Brigham Young University
Inspire-HEP

verhaaren@physics.byu.edu

Contents

I Introduction 1

1 How to Use This Book 1

2 Acknowledgements 4

3 An Orientation 4

II Particles and Fields 7

4 Units of High Energy Physics 8

5 Special Relativity 11
5.1 Relativistic Kinematics 19

6 Collider Basics 26
6.1 Detectors and Cross Section 34

7 Particles and the Lorentz Group 40

8 The Klein-Gordon Equation 43
8.1 The Klein-Gordon Lagrangian 45

9 The Dirac Equation 53
9.1 The Dirac Lagrangian 59

10 Spin-1 Lagrangian 68

11 Massless Spin-1 74
11.1 Connections to E&M 76

III Quantum Electrodynamics 80

12 Lagrangian from Local Symmetry 81

13 Perturbation Theory & Feynman Diagrams 85
13.1 Interaction Vertices 91

14 Running Couplings 95
14.1 Renormalization 98

15 Discrete Symmetries 100
15.1 Parity 101
15.2 Charge Conjugation 104
15.3 Time Reversal 109

IV Quantum Chromodynamics 111

16 Nonabelian Local Symmetry 112

17 A Simple SU(2) Example 118

18 QCD from SU(3) 120
18.1 Confinement 122
18.2 Partons, Fragmentation, and Showers 127

V Electroweak Symmetry Breaking 129

19 Nonabelian SU(2)$_L$ 130

20 The Higgs Mechanism 132
20.1 Spontaneous Symmetry Breaking 132
20.2 Breaking Continuous Symmetries 135
20.3 Breaking Gauged Symmetries 140

21 Electroweak Symmetry Breaking 141
21.1 Charges and Couplings 147

22 Beta Decay 150

VI The Standard Model 153

23 Fermion Masses 153

24 Flavor and the CKM Matrix 156
24.1 Weak CP-Violation 163

25 Bound States & Flavor 167
25.1 SU(3) Flavor 170

26 Symmetries of the Standard Model 175

VII Beyond The Standard Model 180

27 Gravity 180

28 Neutrino Masses 186
28.1 Majorana Fermions 192
28.2 Seesaw Mechanism 195

29 Strong CP Problem 196
29.1 Quark Mass Connection 199
29.2 The QCD Axion 201

30 Grand Unified Theories 203

31 Electroweak Hierarchy Problem 216

VIII Physics Prerequisites 225

32 Variational Methods 225
32.1 Hamiltonian Dynamics 232
32.2 Field Theories 236
32.3 Symmetries and Noether's Theorem 239

33 Quantum Essentials 243
33.1 Position and Momentum Operators 247
33.2 Quantum Symmetries 251

33.3 Time Evolution 253
33.4 Simple Harmonic Oscillator 254

IX Mathematical Resources 260

34 Vector Spaces 260

35 Index Notation 269

36 Groups 281

37 Lie Groups 287

38 Lie Algebras 290

39 Representations 299
39.1 Representations of SU(3) 319
39.2 Representations of Subgroups 339

40 Contour Integration 341

41 Green Functions 348

42 Statistics and Uncertainty 359

PART

I

Introduction

Sec. 1. How to Use
Sec. 2. Acknowledgements
Sec. 3. Orientation

Section 1

How to Use This Book

The intent of this text is to prepare interested undergraduate and beginning graduate students for research in high energy particle physics. This focus is somewhat different than much of the existing literature, which is primarily focused on experienced graduate students or undergraduate surveys. Rather than acting as a survey of many different particle physics facts and concepts this text tries to build up a theoretical framework that can connect and explain those facts and concepts. No attempt is made to outline the history of various experiments or discoveries that led to this framework. While such histories are interesting and often instructive [1, 2], I do not feel that they aid the beginning student in learning the concepts and techniques needed to understand particle physics.

The book contains a significant amount of introductory mathematical material following the main text. The purpose for putting this last is *not* to imply that it is less important. In many cases it is a necessary foundation for the physics. However, students come with different preparation and exposure to these topics. Therefore, I have left these introductions as a resource. The beginning student may find it useful to work through much of these resources, or at to least read through them so that they know where to go when they come upon unfamiliar concepts in the main text.

These introductions are neither mathematically rigorous nor comprehensive. They are more of a minimal set of information that I think can be useful to beginning students. It is very likely that others would include more or less. Certainly most mathematicians would be unsatisfied with the treatment, though I hope they would at least find the examples and exercises useful.

In addition to the mathematics resources I have included overviews of some parts of a typical undergraduate physics curriculum after the central material concludes. As with the mathematics, these are nothing like comprehensive. They are meant to provide some minimal introduction to ideas that are built upon in the main text. Specifically, the focus is on those topics and concepts that are essential to understanding the Lagrangian formulation of the Standard Model of particle physics and something of the methods that use that Lagrangian to make concrete calculations.

Throughout the text I have tried to include many exercises as they somewhat naturally apply to the course content. Serious students should work through these exercises as they provide an essential part of the real learning. There is a certain amount of truth in the idea that "If you can't do it, you don't understand it." By struggling through exercises students are likely to find out what they do not yet understand and begin the process coming to know it. They also gain essential practice in doing particle physics calculations which may make subsequent sections of this text easier to follow. In short, the term exercise fulfills the same purpose here as it does in regard to physical health and skill. By exercising one builds strength, skill, and ability to do increasingly difficult things. Learning particle physics can be difficult, but with diligent practice (exercise) any student can improve.

This is not a book on quantum field theory even though quantum field theory is the typical and essential language of high energy physics. As important as quantum field theory is, much of particle physics can be understood before grappling with the full complexity, nuance, and precision it provides. Indeed, my hope is that students can use this text to learn first many concepts and ideas about the physics and mathematics that are foundational to quantum field theory. Then, when they take their first quantum field theory course they can focus on the field theory, learning it deeply and thoroughly. Ideally, they will be better equipped, after working through this material, to understand how to use what they are being taught. In short, I hope that this text can serve as semester zero of quantum field theory.

That being said, there are times when it makes sense (at least to me) to point to certain quantum field theory features or consequences that relate to the concepts being discussed. At these points I do make these connections as best I can, but without taking the time required to justify them. Again, I hope that students take these as invitations to further study and some concepts to watch out for in those studies. There are few quantum field theory texts that seem more accessible to undergraduate students [3, 4]. A beginning text that I found useful is "A First Book of Quantum Field Theory" by Lahiri and Pal [5]. Unfortunately, it seems to be out of print. A similar presentation of some of the material is made by one of the authors in the particle physics text [6], which I found very useful.

As mentioned above, this text is also not a historically based or discovery based treatment of high energy or particle physics. There are many texts that follow these routes and interested students are encouraged to read them. Among these I make special mention of Larkoski's text [7]. As I was searching for a text to use in teaching advanced undergraduates and beginning graduate students particle physics, I found Larkoski's book to the best suited for the course I planned to teach. I used it over the next few years, but the course drifted further and further from this text. In the

end I begin writing these notes for that class, but I am sure that a lot of Larkoski's ideas can be found in what I've written. I still like the book, it just ended up being the right book for a different course and I am sure that some will prefer to what I have written.

A primary aim of mine is that this text not be a survey of quick facts whose connections and relations are unclear. Such surveys can be useful for orientation within the larger landscape of physics ideas. However, I have often found them unsatisfying. Rather than a great many facts that must all be taken to be true based on the authority of the text, I have sought to make a smaller number of assumptions and work out their consequences. Therefore, this text aims to be a fairly systematic construction of the Standard Model Lagrangian. Many concepts and experimental consequences follow from this structure and I do try to emphasize them when they arise from the concepts being discussed. However, the organization is based on the structures, not on the phenomena that result from that structure. Of course, that we can claim these structures have anything to do with nature relies on the experimental determination of these resulting phenomena. So, I hope that my focus is not seen as implying that the phenomena and methods of detection are unimportant, that is most emphatically not my intent. I am simply separating the issues of what the structures are and why we associate them with nature. This text focuses strongly on the former, in part because there are many very good texts that address the latter.

The shape of this text has been influenced by the one semester course I have taught on particle physics each year since 2022. My teaching follows the outline of the text, but also takes detours into the mathematical resources and physics prerequisites early in the course. Those detours are marked in the text by comments that advise the reader to review those concepts if they are not yet familiar with them. My teaching of this detour material has not covered all the content, for instance I do not go through all of the detail regarding weights and roots of Lie group representations. That material is simply there as reference for the interested student.

When I have taught I have not covered much of the material in the Beyond the Standard Model section of the text. I usually get to gravity on the last day of class. The remaining sections can serve interested students or might be included in courses that have more days in their semester or that cover less review material.

My course also included projects (not currently included in the text) that employed a collider event generator (I used MadGraph [8] though I am sure others would work just as well) to tie the theory to experimental outcomes. The intent is to connect the mathematical structures built in the text with concrete measurements in the real world. Preparation for these projects included working through the section on statistics in the mathematical resources.

Section 2

Acknowledgements

This text has benefited greatly from readers and students who found holes or other areas for improvement by working through previous versions. A partial list includes Michael Vaka, Logan Page, Josh Forsyth, Thomas Cochran, Martin Clemens, Carson Tenney, and Nate Garey. I also gratefully acknowledge support from the National Science Foundation under Grant No. PHY-2210067.

In addition to students, I am indebted to a great many teachers, mentors, and collaborators who have helped me to better understand these topics over the years. I have also learned from a great many books on these topics. Rather disappointingly for me, several times while preparing this text I have stumbled upon discussions or derivations very like the ones I had put in this book, but had no memory of reading before. This leads me to suspect that very little, if anything, in what follows is novel, even if I have not yet tracked down the correct source. If you find that my referencing is incorrect in some way or incomplete, please let me know. I have tried give recognition for various ideas correctly, though likely imperfectly.

In particular several books on quantum field theory have greatly shaped my understanding [5, 9–15]. It is likely that any particularly good presentation of material is due to these sources. I have made some efforts to point them out to readers when appropriate. Similarly, I have benefited from many books that discuss particle physics and the standard model [7, 16–19]. Of course, any all misunderstandings of these references belongs to me alone.

Section 3

An Orientation

It can be difficult to communicate exactly what elementary particle physics is. First introductions often include a list or table of particles including various characteristics about them. Often descriptions say a lot about symmetry and how important it is to understanding particle physics. But the reader is not given much guidance as to why this might be, they are forced to take the author's word for it. In this text we build up the component pieces of what we need to see how symmetry is used in particle physics. This takes a significant investment of effort before we start connecting things to particles and it is easy to feel a little lost along the way.

In order to provide an end goal to navigate by, in this section we provide some context. We also list particles along with characteristics so

that students might recognize why we are learning certain concepts. Let us begin with what is perhaps the most familiar of particles: The electron.

Mass	Spin	Electric Charge	Size	Lifetime
9.109×10^{-31} kg	1/2	-1.601$\times10^{-19}$ C	$< 2 \times 10^{-20}$ m	$> 2 \times 10^{36}$ s

Table 1. Table of some qualities of the electron. Additional information can references regarding experimental results can be found in the Review of Particle Physics [20].

Table 1 displays some aspects of electrons and conveys a few important ideas. First, the units that are useful to daily life, like meters, seconds, and kilograms are not well suited to elementary particles. Other units are introduced in what follows that are very cumbersome for daily life, but useful for describing very small. Second, there is often a disconnect between our present models and our experimental probes. In the standard model of particle physics the electron is treated as having zero size and infinite lifetime. Of course, we cannot measure an infinite lifetime, all we can do is place a lower bound. Similarly, we cannot claim that the electron does not have a finite size, all we can say is that so far we have seen no hint of nonzero size. Finally, characteristics like mass and electric charge may be familiar, while spin may not be. Many of these less familiar characteristics become easier to understand as the mathematical description is clarified.

As discussed in what follows, the intrinsic angular momentum of particles (what we call their spin) can be understood as a consequence of the special theory of relativity. That is, special relativity assumes that space and time follow a certain structure, which might also be described as a set of symmetries. These symmetries are described by a mathematical object called a group. This group structure, or symmetry structure, is very general and can be represented in several distinct ways. In the case of special relativity, particles that are associated with different representations of these symmetries can have different values of spin. In order to give the idea in this paragraph concrete meaning, one must take the time to learn about groups and representations.

Representing symmetry groups may seem like esoteric mathematics, but it pervades this course. Symmetries are deeply related to conserved quantities from energy and momentum to particle number and electric charge. The fundamental forces, the strong and weak nuclear forces along with electromagnetism, are also tied to symmetry groups and each particle that experiences these forces is related to one of their representations. So, while it takes time and effort to understand something of group and representation the payoff is substantial and used over and over.

One issue with listing particles, like in Table 1, is that our most accurate models of nature are not exactly models of particles, they are particles of fields. These fields are taken to exist throughout all of space and time.

What are often called particles are certain types of waves in these fields. The fields that make up the standard model of particles physics could be listed and grouped by how they represent the various symmetries.

The distinction between fields and particles might seem pedantic, but it can be important and aid understanding. It is the fields that are in specific representations of various groups and provide more information than the particles themselves. For instance, the particles may or may not be simply related to the fields. The electron can be thought of as an excitation of an electron field ψ_e, but this field is actually the combination for two more primal fields, that we might call ψ_{eL} and ψ_{eR}. These fields are in different representations of the symmetries of the standard model but are linked such that they produces the electron field, at least at low energies.

The intent of this text is to introduce you to the different types of fields that models of nature (assuming the symmetries of special relativity) can be built out of. Then, additional types of symmetries are considered including those that govern how different fields might interact with each other. This eventually leads to the symmetry structure of the standard model and the various particles associated with it.

With background, we can introduce the many of the fields that make up the Standard Model of Particle Physics. There are three fields that are like the electron. These are the electron itself along with the muon and tau. The muon and tau are distinguished from the electron by being about 200 and 3500 times more massive, respectively. These differences in mass are due to their different sized interactions with the Higgs field.

One important difference between the electron-like fields and the Higgs field is they have different intrinsic angular momentum, or spin. The electron-like fields are spin-1/2, which means they have an intrinsic angular momentum with magnitude

$$\hbar\sqrt{\frac{1}{2}\left(\frac{1}{2}+1\right)} = \hbar\frac{\sqrt{3}}{2}\ , \tag{3.1}$$

where

$$\hbar = \frac{h}{2\pi} \approx 1.055\times 10^{-34}\ \mathrm{J}\cdot\mathrm{s}\ , \tag{3.2}$$

is sometimes called the reduced Planck constant. Note that it has dimensions of angular momentum. The Higgs field is spin-0, it has no intrinsic angular momentum.

Each of the electron-like fields has electric charge equal to that of the electron). This means that they interact with light, or the photon field. The photon is a massless particle particle with spin-1, so it has intrinsic angular momentum of $\hbar\sqrt{1(1+1)} = \hbar\sqrt{2}$. Each of the electron like-fields is also related to a spin-1/2 field without electric charge and very small mass. These are called neutrinos. Collectively, the neutrinos and electron-like

fields are called leptons.

Protons and neutrons are the primary building blocks of atomic nuclei. However, they are not elementary, they are bound states of other particles that we call quarks and gluons. The quark are spin-1/2 fields and come in six types, or flavors. Three of these, the down quark, the strange quark, and the bottom quark all have an electric charge that is one-third that of the electron. Of these the down quark has the smallest mass, with the strange and bottom quarks about 20 and 870 times more massive, respectively. As with the electrically charged leptons, the differences in the quark masses are due to their different interactions with the Higgs field. The other three quark flavors all have electric charge that is $-2/3$ times that of the electron. They are called the up quark, the charm quark, and the top quark. The up quark is just a little less massive than the down quark, while the charm and top quarks are about 500 and 75000 times the up quark mass, respectively.

The gluons are eight massless spin-1 fields that interact strongly with the quarks and with each other. It is these interactions that ultimately give rise to the strong nuclear force. The weak nuclear force is due to two other spin-1 fields. One of these, the $W^{\pm}$ has a mass about 80 times that of the proton and has the same magnitude of electric charge as the electron. The Z particle is about 90 times more massive than the proton and has no electric charge. The masses of both the $W^{\pm}$ and Z are due to the Higgs field.

A few other quark bound states are also discussed, but the above particles make up the primary focus of this text. It may be that the descriptions above primarily leave you with more questions. I hope that many of these questions are given satisfying answers throughout the remainder of this text.

PART

II

Particles and Fields

Sec. 4. Units
Sec. 5. Relativity
Sec. 6. Collider Basics
Sec. 7. Lorentz Group
Sec. 8. Klein-Gordon Eq
Sec. 9. Dirac Equation
Sec. 10. Spin-1 Equations
Sec. 11. Massless Spin-1

Particle physics is typically thought of as the study of nature on very small scales. Resolving the minute structure of nature as expressed through the interactions of various particles (a term that is defined a little more carefully below). This study of small scales is also referred to as high energy physics.

It is reasonable to ask how we can know about scales this removed from everyday human experience. There is a sense in which we learn to put energy into the basic structure of nature and observe how that energy propagates. To do this most effectively, we assume that this structure manifests the symmetries of special relativity, which turns out to be powerfully constraining. First, however, we discuss a basic set of units preferred in high energy physics.

Section 4

Units of High Energy Physics

Before we begin discussing physics we introduce a convenient and ubiquitous system of units. These units make the equality of high energy physics with small distance physics. This relationship can already be motivated by thinking about light.

When we use light to discover the characteristics of an object (like when we use our eyes) then we can only resolve features that are larger than the wavelength λ of the light we are using. The energy of a photon (a particle of light) is given by

$$E = \frac{h\,c}{\lambda}\ , \tag{4.1}$$

where h is Planck's constant and c is the speed of light in vacuum. The two number in the numerator are constants whose exact value depend on the system of units we are using to measure quantities. The significant physics is the inverse relationship between the energy of a photon and its wavelength. Longer wavelength photons have lower energy while shorter wavelengths correspond to higher energies.

The human eye is sensitive to wavelengths between about 450 and 600 nanometer (nm). In the SI units of meters (m) seconds (s) and kilograms (kg) etc, Plank's constant and the speed of light are

$$h = 6.626 \times 10^{-34} \frac{\mathrm{m}^2\mathrm{kg}}{\mathrm{s}}\ , \qquad c = 2.99 \times 10^{8} \frac{\mathrm{m}}{\mathrm{s}}\ . \tag{4.2}$$

One sees immediately that in SI units, which are based on the scales of human day-to-day experience, that Planck's constant is very small and the speed of light is very big. This is why quantum mechanical and relativistic effects are not obvious in most aspects of life. Newtonian physics is essentially the $h \to 0\,, c \to \infty$ limit of physics as we currently understand it. Using these values we find that the energy of the photons our eyes are sensitive to vary from 4.41×10^{-19} Joules to 3.31×10^{-19} Joules. These are quite small, which is good for our eyes, but not very helpful for our understanding.

Other energy units are often used for small scale processes. The electron Volt (eV) is the amount of energy an electron gains under the influence of a 1 Volt electric potential:

$$1 \text{ eV} = 1.60218 \times 10^{-19} \text{ Joules}\ . \tag{4.3}$$

This scale is useful for atomic physics and chemical reactions. For instance, the energy required to remove the bound electron from a Hydrogen atom is

13.6 eV. But this binding energy does not correspond to the size of atomic systems.

Exercise 4.1 What photon wavelength can be associated to the binding energy of the Hydrogen atom, 13.6 eV? What physical distance scale does this correspond to?

Exercise 4.2 The Bohr radius for a Hydrogen atom is

$$a_0 = 5.29 \times 10^{-11} \text{ meters} . \tag{4.4}$$

If we take this to be the wavelength of a photon, what energy in eV does it correspond to? How does this compare with the Hydrogen binding energy. Note that thousands of eV are denoted with kilo-eV or keV.

Exercise 4.3 The characteristic size of an atomic nucleus is

$$1 \text{ fm} = 10^{-15} \text{ meters} . \tag{4.5}$$

If we take this to be the wavelength of a photon, what energy in eV does it correspond to? Note that millions of eV are denoted by mega-eV or MeV.

The point of these exercises is to motivate the inverse relationship between energy and length. In high energy physics a system of units (sometimes called natural units) makes this, and other relationships, obvious. As a first step we choose units in which $c = 1$. For example, if we measure time in seconds and length in light-seconds (ls) (the distance light in vacuum travels in one second) then

$$c = 2.99 \times 10^8 \frac{\text{m}}{\text{s}} \times \frac{1 \text{ ls}}{2.99 \times 10^8 \text{ m}} = 1 \frac{\text{ls}}{\text{s}} . \tag{4.6}$$

In practice, the speed of light is a physical parameter that relates distance to time. We express this by

$$[\text{Length}] = c\,[\text{Time}] \quad \Rightarrow \quad [\text{Length}] = [\text{Time}] \ , \tag{4.7}$$

where the terms in brackets indicate generic dimensional quantities. By choosing units in which $c = 1$ we can treat space and time on the same footing. From Einstein's famous results we also have

$$E = mc^2 \quad \Rightarrow \quad E = m \ , \tag{4.8}$$

or

$$[\text{Energy}] = [\text{Mass}] \ . \tag{4.9}$$

That is, we see that energy and mass are deeply related. They are not quite the same, but mass is a way of storing energy. The danger here is that in reality space and time carry different dimensions as do energy and mass. The factors of c are not written explicitly so we as the physicists must keep track ourselves of what types of quantities we are considering.

So far, we have developed a connection between space and time and a connection between energy and mass. We now connect these two. The dimensions of Planck's constant are those of 'action' or angular momentum, which is energy times time. It is most convenient to choose units in which $\hbar = h/(2\pi) = 1$. This means that

$$\hbar = 1 = [\text{Energy}] \times [\text{Time}] \quad \Rightarrow \quad [\text{Time}] = [\text{Energy}]^{-1} \ . \tag{4.10}$$

This allows us to describe energy, mass, space, and time by one dimensionful unit. The others are all determined from this scale by factors of c and $\hbar$.

Example

A useful energy scale is the giga-eV or GeV, which is a billion electron-Volts. We can then find what length corresponds to 1 GeV.

$$[\text{Length}] = c\,[\text{Time}] = c\frac{\hbar}{[\text{Energy}]} \ . \tag{4.11}$$

Putting in SI numbers for, we find

$$1 \text{ GeV}^{-1} = 1.97 \times 10^{-16} \text{ meters} \ . \tag{4.12}$$

Exercise 4.4

Show that in natural units

$$1 \text{ GeV} = 1.602 \times 10^{-10} \text{ Joules} \tag{4.13}$$
$$1 \text{ GeV}^{-1} = 6.58 \times 10^{-25} \text{ seconds} \tag{4.14}$$

In what follows we refer to all dimensions in terms of GeV or other energy scales. We often loosely refer to this as a mass scale. Most quantities are given as some power of GeV, with the understanding that the above formulas can be used to convert them into SI units if needed.

Exercise 4.5

Hideki Yukawa hypothesized that a new particle, which we now called the pion, can be used to model the strong nuclear force between nucleons, protons and neutrons. Assuming that atomic nuclei have a characteristic radius of about 1 fm, use dimensional analysis to estimate the pion mass. Check the PDG (the publication from the Particle Data Group) and find the measured mass of the pion. While there are two types of pions, the neutral π^0 and the electrically charged $\pi^\pm$, their masses are quite similar.

These natural units can be extended by including other constants of

nature. For instance, Boltzmann's constant is

$$k_{\rm B} = 1.38 \times 10^{-23} \frac{\text{Joules}}{\text{Kelvin}} \ . \tag{4.15}$$

By choosing unity in which $k_{\rm B} = 1$ we can write temperatures in terms of energy.

Exercise 4.6 Show that

$$1 \text{ eV} = 1.61 \times 10^4 \text{ Kelvin} \ . \tag{4.16}$$

The core of the sun has a temperature of about 15 million Kelvin. Express this in keV.

Sometimes Newton's gravitational constant G (or sometimes $8\pi G$) is chosen to be one. This uniquely specifies a scale for lengths, times, masses, and temperatures. These are typically referred to as the Planck length, Planck time, Planck mass, and Planck temperature. However, these are not particularly useful scales for contemporary particle physics, so we do not use them in this text.

Now that the standard units of particle physics have been outlined, we can start in on the physics. In doing so we assume a working knowledge of vector spaces and linear algebra. Many of these topics are reviewed in Sec. 34. Quite a lot of physics has linear algebra at its heart. In this text we make use of several distinct vector spaces. In order to keep track of these several notations are used, with abstract index notation (reviewed in Sec. 35) perhaps the most powerful.

Section 5

Special Relativity

The stage for much of particle physics is vector space that describes special relativity. Therefore, we include a brief introduction to special relativity and highlight aspects that are more relevant to particle physics. Students may find the relevant chapters of [21] and [22] to be useful complements to and extensions of these ideas.

Einstein's theory of special relativity highlights a deep and significant connection between space and time. The structure of this spacetime determines much of the vocabulary used to describe higher energy physics. The essential connection has to do with describing physical systems in coordinate systems that are related to each other by rotations or changes of velocity. We already expect that rotations should have no effect on the system being described.[1]

[1] This refers to two coordinate systems that are related by a rotation about some point, not a coordinate system that continues to rotate with respect to another.

Consider Newton's 2nd Law written in two spatial dimensions as a simple example

$$\vec{F} = m\,\vec{a}\ . \tag{5.1}$$

The familiar vector notation communicates that this equation is true no matter how a given system is oriented in space. It is a geometric statement, no matter how the net force is oriented the acceleration is oriented in the same direction.

Using index notation (see Sec. 35 if unfamiliar) we write this same equation as

$$F^i = m\,a^i\ , \tag{5.2}$$

and consider rotating the vectors counterclockwise by an angle θ using the matrix

$$R^i{}_j = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}\ . \tag{5.3}$$

That is we can define rotated vectors

$$F'^i = R^i{}_j F^j\ ,\quad a'^i = R^i{}_j a^j\ , \tag{5.4}$$

and the correct realization of Newton's 2nd Law is

$$F'^i = m\,a'^i\ . \tag{5.5}$$

The way these equations are written, in vector notation, can then be understood as conveying an assumption we make about physical laws: that they do not depend on orientation in space. Or, in other words, that the expression of these laws takes the same form no matter what orientation is chosen.

The vector space considered above is $\mathbb{E}^2$ or two dimensional Euclidean space. The metric for this space δ_{ij} can be expressed in Cartesian coordinates as

$$\delta_{ij} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\ . \tag{5.6}$$

This defines the inner-product in this space

$$a^i a^j \delta_{ij} = \vec{a}\cdot\vec{a} = (a^1)^2 + (a^2)^2\ , \tag{5.7}$$

and

$$F^i a^j \delta_{ij} = \vec{F}\cdot\vec{a} = F^1 a^1 + F^2 a^2\ . \tag{5.8}$$

By our assumption that physics does not depend on orientation, these inner products must be unchanged by rotations it must be that

$$F^i a^j \delta_{ij} = F'^i a'^j \delta_{ij}\ . \tag{5.9}$$

Exercise 5.1 Show that Eqs. (5.4) and (5.9) lead to the condition

$$\delta_{ij} = R^m{}_i R^n{}_j \delta_{mn} \ . \tag{5.10}$$

As explained in Sec. 37, this is the condition the defines orthogonal matrices.

Exercise 5.2 Show that the matrix in Eq. (5.3) satisfies the condition in Eq. (5.10).

We have shown this condition in two dimensions, but the generalization to three dimensions is nearly identical. The three-by-three rotation matrices still satisfy Eq. (5.10) where the metric in Cartesian coordinates is the three-by-three identity matrix.

The jump to four dimensional spacetime is similar, but there are a few important distinctions. In addition to the rotation invariance we have discussed, there is now an invariance among inertial frames. That is two coordinate systems that differ by one moving with a velocity v relative to the other. Consider a four-vector of Cartesian coordinates

$$x^\mu = \begin{pmatrix} t \\ x \\ y \\ z \end{pmatrix} \ . \tag{5.11}$$

Here, and throughout this text, Greek indices take values from $\{0, 1, 2, 3\}$ with $x^0 = t$ and $x^{1,2,3}$ give the usual spatial components of a three vector. When we want to indicate that the index only takes spatial values (that is, from $\{1, 2, 3\}$) we use Latin indices like j or k. The coordinates of an equivalent frame, either by rotation or by boosting to a relative velocity v, is

$$x'^\mu = \begin{pmatrix} t' \\ x' \\ y' \\ z' \end{pmatrix} \ . \tag{5.12}$$

In special relativity it is not the length of a vector that needs to be preserved by all transformations. You may recall that observers from different frames may disagree about the lengths L and times T associated with some physical process. However, they always agree on the value of

$$T^2 - L^2 \ . \tag{5.13}$$

This implies that for special relativity we should require that

$$t'^2 - x'^2 - y'^2 - z'^2 = t^2 - x^2 - y^2 - z^2 \ . \tag{5.14}$$

Notice that in the special case that $t = t'$ this reduces to the usual requirement that vector lengths are unchanged. In other words, the metric of special relativity, often referred to as Minkowski space, is

$$\eta_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} . \tag{5.15}$$

The lower left three-by-three block has the same form (up to the minus sign) as the Euclidean metric. This simply encodes the fact that space alone does have a Euclidean character (we are not accounting for the curved spacetime of General Relativity). The relative sign between space and time has the new properties we are interested in exploring. This also means that there is a sign difference between the spatial components of four-vector and its dual:

$$x_\mu = x^\nu \eta_{\nu\mu} = (t, -x, -y, -z) \ . \tag{5.16}$$

Let us consider a simple case that captures these effect. Imagine a single space direction x. Suppose we have two systems of coordinates for describing space and time. The first is $(t,\, x)$ and the second is $(t',\, x')$. This primed coordinates correspond to a frame that is moving in the positive x direction with velocity v relative to the unprimed coordinates. Before special relativity we would simply say that

$$t' = t \ , \qquad x' = x + vt \ , \tag{5.17}$$

where we have chosen $t = 0$ to be the time when $x' = x$. This transformation between two inertial frames is sometimes referred to a Galilean relativity.

Definition An **inertial frame** is a system of coordinates that are not accelerating.

The reader may well ask what this statement means. How does one define acceleration? It turns out that one can make measurements within a frame that are sensitive to whether or not you are accelerating. Once you have found one inertial frame every other that is not accelerating with respect to that frame is also inertial.

Exercise 5.3 Show that the transformation law given in Eq. (5.17) does not satisfy

$$t^2 - x^2 = t'^2 - x'^2 \ . \tag{5.18}$$

Galilean relativity does not satisfy the requirements of Minkowski spacetime, but it should agree when velocities are small compared to the speed

of light. We can parameterize the possibilities by a linear transformation

$$t' = c_1 t + c_2 v x \ , \quad x' = c_3 x + c_4 v t \ . \tag{5.19}$$

This transformation is linear in the sense that no power of higher powers of x or t appear in the transformation law. This is similar to the types of transformations that describe rotations. Note that they are restricted by the $v \to 0$ limit. For instance $c_1 = c_3 = 1$ when $v = 0$ as in this case the two frames should be identical.

Exercise 5.4 Insert the transformation law given in Eq. (5.19) into the 1+1 (one space and one time dimension) Minkowski space requirement (5.18). Conclude that

$$c_1 c_2 = c_3 c_4 \ , \quad c_1^2 - v^2 c_4^2 = 1 \ , \quad c_3^2 - v^2 c_2^2 = 1 \ . \tag{5.20}$$

These three constraints on four parameters do not uniquely determine them. It is useful to define

$$\alpha = \frac{c_4}{c_1} \ . \tag{5.21}$$

Show that all the c_is can be written in terms of α as

$$c_1 = c_3 = \frac{1}{\sqrt{1 - v^2\alpha^2}} \ , \quad c_2 = c_4 = \frac{\alpha}{\sqrt{1 - v^2\alpha^2}} \ , \tag{5.22}$$

where we have assumed that c_1s do not change sign as $v \to 0$.

Finally, our transformation should match the Galilean form for small, but nonzero, v. Keeping on terms linear in v/c we have

$$ct' = ct + \frac{v}{c} x\alpha \ , \quad x' = x + \frac{v}{c}\alpha\, ct \ . \tag{5.23}$$

The factors of c have been made explicit in this equation to make clear that velocity correction to the t coordinate is actually much smaller than the correction to the x coordinate. Neglecting all v/c terms we have

$$t' = t \ , \quad x' = x + v\alpha t \ , \tag{5.24}$$

which agrees with the Galilean expectation in Eq. (5.17) only when $\alpha = 1$. This motivate taking all the $c_i = \gamma$ where

$$\gamma = \frac{1}{\sqrt{1 - v^2}} \ . \tag{5.25}$$

The transformation law is then

$$t' = \gamma(t + vx) \ , \quad x' = \gamma(x + vt) \ . \tag{5.26}$$

Exercise 5.5 Show that $\gamma \geq 1$ and determine the situation in which $\gamma = 1$.

Returning to $3+1$ Minkowski space the transformation matrices $\Lambda^{\mu}{}_{\nu}$ that preserve the metric according to

$$\eta_{\mu\nu} = \Lambda^{\alpha}{}_{\mu}\Lambda^{\beta}{}_{\nu}\eta_{\alpha\beta} \ , \tag{5.27}$$

are called Lorentz transformations. The simple transformation that boosts to a frame moving with velocity v in the x-direction is

$$\begin{pmatrix} t' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \gamma & v\gamma & 0 & 0 \\ v\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} t \\ x \\ y \\ z \end{pmatrix} \ . \tag{5.28}$$

Exercise 5.6 Show that the transformation in Eq. (5.28) satisfies the defining relationship for Lorentz transformations, Eq. (5.27). If you plan to use matrix multiplication, is may be useful to remember that

$$\Lambda^{\alpha}{}_{\mu} = \Lambda^{T\alpha}_{\mu} \ . \tag{5.29}$$

So, we can write

$$\eta_{\mu\nu} = \Lambda^{T\alpha}_{\mu}\eta_{\alpha\beta}\Lambda^{\beta}{}_{\nu} \ . \tag{5.30}$$

Exercise 5.7 Write down the Lorentz transformations for boosts in each of the y and z directions. Next, prove that the boosting along an arbitrary velocity vector

$$\vec{v} = \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix} \ , \tag{5.31}$$

leads to the transformation law

$$t' = \gamma\left(t + \vec{v}\cdot\vec{x}\right) \ , \qquad \vec{x}' = \gamma\left(\vec{x} + \vec{v}t\right) \ , \tag{5.32}$$

where

$$v = \sqrt{v_x^2 + v_y^2 + v_z^2} \ . \tag{5.33}$$

The full group of Lorentz transformations are composed of combinations of boosts along the three spatial directions and rotations about each spatial axis. Note that there is a qualitative difference between boosting and rotating. Rotating about an axis by 2π radians is the same as no rotation at all. This is related to the compact nature of the Lie group associated to rotations. In contrast, there is no finite boost that is equivalent to no boost and there is no maximum value of γ that can be employed. Correspondingly, the Lie group related to boosting is not compact.

Because the boosting symmetry is not compact the operators that generate it are not Hermitian! Consider a boost along the x direction for a very small v and expand to linear order in the velocity. From Eq. (5.28) we find

$$\begin{pmatrix} t' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} 1 & v & 0 & 0 \\ v & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} t \\ x \\ y \\ z \end{pmatrix} + \mathcal{O}(v^2)$$
$$= (\mathbb{I} - ivK_x) \begin{pmatrix} t \\ x \\ y \\ z \end{pmatrix} , \tag{5.34}$$

where

$$K_x = \begin{pmatrix} 0 & -i & 0 & 0 \\ -i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} , \tag{5.35}$$

is an *anti*-Hermitian operator, one with $K_x^\dagger = -K_x$. This anti-Hermian form ends up preventing us from finding finite dimensional, unitary representations of the full Lorentz group.

While the general possibilities for rotations and boosts looks quite complicated for many of our purposes we choose coordinates that make transformations simple. For instance, in particle collisions we typically take the motion of a particle to be along the z-axis. Then going from the particle at rest to the lab frame where it has a velocity v are accomplished by a boost in the z direction.

Definition

The **rest frame** of an object is the inertial frame in which it has zero velocity.

This structure leads to some of the famously nonintuitive predictions of special relativity. First, suppose you measure the length of an object at rest. We choose this length to be along the z direction and denote it $\Delta z = z_2 - z_1$. These positions can be measured at any time, so we say the ends of the object are located at (t_2, z_2) and (t_1, z_1) in its rest frame.

Suppose the same object is moving past you with velocity v in the positive z direction. If you measure its length now, it is being measured in a frame that is moving with velocity v in the *negative* z direction relative to its rest frame. (Stop and draw a picture of this in the margin to make sure you understand the negative sign.) This implies that measurements in the moving frame would be

$$z_2' = \gamma(z_2 - vt_2) , \qquad z_1' = \gamma(z_1 - vt_1) . \tag{5.36}$$

This means that

$$\Delta z' = \gamma\left(z_2 - vt_2 - z_1 + vt_1\right) = \gamma\left(\Delta z - v\Delta t\right) \ , \tag{5.37}$$
$$\Delta t' = \gamma\left(t_2 - vz_2 - t_1 + vz_1\right) = \gamma\left(\Delta t - v\Delta z\right) \ . \tag{5.38}$$

How do we define the length of the object in the moving frame? It is the distance between the spatial coordinates at one fixed time. In other words, we require $\Delta t' = 0$.

Exercise 5.8

Find the relationship between Δt and Δz when $\Delta t' = 0$. Use this condition to show that

$$\Delta z' = \Delta z/\gamma \ . \tag{5.39}$$

Recall that $\gamma \geq 1$, so we see that the length of the moving object ($\Delta z'$) is *less* than the length of the object at rest (Δz). This is usually referred to as **Lorentz contraction**.

We have shown that the structure of special relativity implies that moving objects are contracted relative to their rest length. What about the time of moving objects? The transformations in Eqs. (5.37) and (5.38) still apply. But in this case we are interested in some process of fixed time Δt in the rest frame. This process takes place at one fixed point, because the object is at rest, so $\Delta z = 0$.

Exercise 5.9

Use the condition $\Delta z = 0$ to show that

$$\Delta t' = \gamma \Delta t \ . \tag{5.40}$$

Because $\gamma \geq 1$ we see that the time of the moving object ($\Delta t'$) is *greater* than the time of the object at rest (Δt). This is usually referred to as **time dilation**.

Example

The classic example of the reality of time dilation comes from the muon lifetime. Muons are particles that are very much like electrons except that their mass is about 200 times larger than the electron mass. Muons also have a finite lifetime of

$$\tau_\mu = 2.197 \times 10^{-6} \text{ seconds} \ . \tag{5.41}$$

This is related to the half-life for their decay into an electron and two neutrinos. Muons are often created by cosmic rays (typically protons moving close to the speed of light) that hit the gas molecules of the upper atmosphere. The average height of the atmosphere above the surface of the earth is about 12 km. The muons produced in these collisions have velocities close to the speed of light.

If we neglect time dilation then the typical distance these muons travel

from the upper atmosphere is about

$$d = c\tau_\mu \approx 600 \text{ meters.} \tag{5.42}$$

This implies that essentially none of them make it to the surface of the earth. (Note that we cannot make a stronger statement because the lifetime is really related to a half-life. The process of decay is quantum mechanical and hence probabilistic.) However, ground based observations do measure a flux of muons produced from these cosmic ray collisions. This has been shown to agree with the time dilation extending the muon lifetime in the frame of the experiment. (In the muon frame it is the distance between the atmosphere and the experiment that has contracted, leading to the same physical results.)

Subsection 5.1

Relativistic Kinematics

We can describe the location of the particle in space and time by a position 4-vector

$$x^\mu = \begin{pmatrix} t \\ x \\ y \\ z \end{pmatrix} . \tag{5.43}$$

However, often we are more interested in the energy E and momentum $\vec{p}$ of an object. These also transform as a four-vector

$$p^\mu = \begin{pmatrix} E \\ p_x \\ p_y \\ p_z \end{pmatrix} . \tag{5.44}$$

If we take this object to truly be a four-vector it means that its inner-product with itself is the same in all inertial frames

$$p^\mu p^\nu \eta_{\mu\nu} = m^2 . \tag{5.45}$$

The value of this inner product is the squared mass of the object in question.

Exercise 5.10 Show that

$$p^\mu p^\nu \eta_{\mu\nu} = E^2 - \vec{p}\cdot\vec{p} . \tag{5.46}$$

This agrees with the standard relativistic relationship between energy and momentum

$$E^2 - \vec{p}\cdot\vec{p} = m^2 . \tag{5.47}$$

Definition 1 The **proper time** of a particle parameterizes its position in four dimensional spacetime. That is $x^\mu(\tau)$ with $t(\tau)$ and $x^i(\tau)$. In general, the proper time of an individual object is different from the proper time of another object and distinct from the coordinate t.

This seems like the right place to introduce a little particle physics jargon. When a system or particle satisfies the above relation, that is

$$E = \sqrt{m^2 + \vec{p}\cdot\vec{p}}\,, \tag{5.48}$$

it is said to be "on-shell." In Fig. 1 we see that each specific mass picks out a curve, a shell, in the energy-momentum plane. You may be wondering why this term exists. Surely all relativistic particles are on-shell, right? As we shall see a little further down things are not quite so simple.

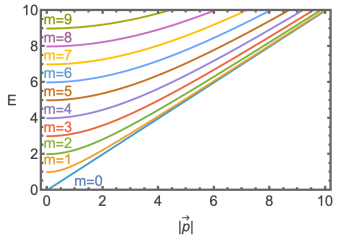


Figure 1. Plot of several "mass shells."

Exercise 5.11 Show that in the rest frame of an object that

$$E = m\ . \tag{5.49}$$

Note that this is Einstein's famous formula with $c = 1$. Then boost to a frame in which the object is moving with velocity v. Show that the energy and momentum of the object satisfy

$$E = \gamma\, m\ , \qquad p^i = m\, v^i\, \gamma\ , \tag{5.50}$$

where p^i and v^i are the components of three vectors. This motivates the form of the four-velocity v^μ as

$$v^\mu = \gamma \begin{pmatrix} 1 \\ v^1 \\ v^2 \\ v^3 \end{pmatrix}\ , \tag{5.51}$$

where, for instance,

$$v^1 = \frac{dx^1}{dt}\ . \tag{5.52}$$

This results agrees with

$$\frac{d}{d\tau}x^\mu = v^\mu\ , \tag{5.53}$$

where τ is the proper time of the particle. Show this means that

$$\frac{dt}{d\tau} = \gamma\ , \tag{5.54}$$

and that

$$v^\mu v_\mu = 1\ . \tag{5.55}$$

The relation $E = \gamma m$ is what leads some people to say that the mass of a an object with velocity is larger than the mass of an object at rest. It is usually more useful to think of an object as having a fixed rest mass given by Eq. (5.45). We then simply say that the energy of a an object is larger than its energy at rest, which is no different that our Newtonian intuition from kinetic energy.

Exercise 5.12

We can define the four acceleration as

$$a^\mu = \frac{d}{d\tau} v^\mu \ . \tag{5.56}$$

First, show that

$$\frac{d}{dt}\gamma = \gamma^3 \vec{v}\cdot\vec{a} = -\bar{a}^\nu v_\nu \ , \tag{5.57}$$

where

$$\bar{a}^\mu = \gamma^2 \begin{pmatrix} 0 \\ a^1 \\ a^2 \\ a^3 \end{pmatrix} \ . \tag{5.58}$$

Use this result to demonstrate that the four acceleration is

$$a^\mu = \bar{a}^\mu - \bar{a}^\nu v_\nu v^\mu \ , \tag{5.59}$$

and that

$$a^\mu v_\mu = 0 \ . \tag{5.60}$$

Exercise 5.13

Suppose you bring a particle of mass m to an energy E. Using the relativistic energy equation in (5.50) to determine the velocity of the particle. Argue that this velocity cannot be equal to or larger than one.

Newton's second law can also be generalized to a relativistic four-vector equation. It is

$$\frac{dp^\mu}{dt} = ma^\mu = F^\mu \ , \tag{5.61}$$

where F^μ is the four-vector that includes the three dimensional force F^i. The time component of this equation

$$\gamma^4 m\vec{a}\cdot\vec{v} = F^0 \ . \tag{5.62}$$

What is the meaning of the time-component of force? Remember that the three-vector part is given by the time derivative of the three-momentum. The time component of the four-momentum is the energy, so the time-component of the force is the time derivative of the energy, which is the power associated with a process.

We are ready to synthesize a few concepts. We argued in the previous

section that on dimensional grounds if we are interested in probing a length scale L then we need to use energies of $E \sim 1/L$. Suppose then we are interested in probing the physics of the proton. This means we need to resolve proton size scales.

Exercise 5.14

The proton has a characteristic size of about 1 fm $= 10^{-15}$ meters. What energy scale should we expect to need to probe its structure? Suppose we want to use electrons, whose mass is about

$$m_e \approx 5 \times 10^{-4} \text{ GeV} , \tag{5.63}$$

to probe this energy scale. What velocity must the electrons have? Compare this velocity to the speed of light in vacuum.

As the previous exercise makes clear, when we seek to investigate the small scale structure of Nature we are nearly always pushed into situations where relativity is essential, its effects cannot be neglected. These small scales are also dominated by quantum mechanical effects. The framework that unifies special relativity and quantum mechanics is Quantum Field Theory (QFT), the details of which are beyond the scope of this text.

Even without the full power of QFT we can learn a lot about the interactions between particles from relativity. The familiar requirements of energy and momentum conservation (in other words, of four-momentum conservation) along with the new requirement that all inertial frames are equivalent greatly constrain interactions and dynamics.

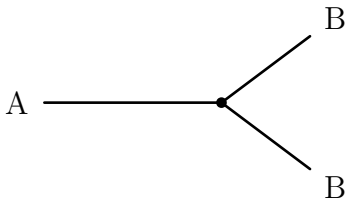


Figure 2. Decay of an A particle into two B particles. Time proceeding from left to right.

Example

Let us explore the decay of a particle. That is, consider a process in which we begin with one particle of type A and end with two particles of type B, $A \to B + B$. At this point we assume nothing about how this process occurs. In general, each particle has an associated momentum four-vector of the form

$$p^\mu = \begin{pmatrix} E \\ p_x \\ p_y \\ p_z \end{pmatrix} , \tag{5.64}$$

where

$$E = \sqrt{m_A^2 + p_x^2 + p_y^2 + p_z^2} . \tag{5.65}$$

Because rotations between frames are equivalent, we choose a frame in which the three-momentum of the particle A is along the positive z-axis

$$p_A^\mu = \begin{pmatrix} E \\ 0 \\ 0 \\ p \end{pmatrix} , \tag{5.66}$$

where here the A subscript is a label, not an index. Now we invoke special relativity. Because two frames related by a boost are equivalent, we choose a frame in which particle A has zero three-momentum. (You might complain that we could have done this from the outset, but for clarity I have taken extra steps.) This leads us to

$$p_A^\mu = \begin{pmatrix} m_A \\ 0 \\ 0 \\ 0 \end{pmatrix} , \tag{5.67}$$

What about the B particles? To distinguish them mathematically we call them particle 1 and particle 2. In the frame we have chosen they have four-momenta

$$p_1^\mu = \begin{pmatrix} E_1 \\ p_1^i \end{pmatrix} , \quad p_2^\mu = \begin{pmatrix} E_2 \\ p_2^i \end{pmatrix} . \tag{5.68}$$

While we have written these four-momenta in a general way, they are constrained by four-momentum conservation

$$p_A^\mu = p_1^\mu + p_2^\mu . \tag{5.69}$$

The $\mu = 0$ component of this equation is simply conservation of energy

$$m_A = E_1 + E_2 , \tag{5.70}$$

while the spatial components enforce three-momentum conservation

$$0 = p_1^i + p_2^i . \tag{5.71}$$

The three momentum conservation implies that the B particles have momentum with equal magnitude, but opposite direction. In other words,

$$p_1^i \equiv p^i , \quad p_2^i = -p^i . \tag{5.72}$$

This, in turn, implies that their energies

$$E_1 = \sqrt{m_B^2 + p^2} = E_2 \equiv E_B , \tag{5.73}$$

are equal. (We have used the notation $p^2 = p^i p_i$.) We put this result into the conservation of energy equation (5.70) to find

$$m_A = 2\sqrt{m_B^2 + p^2} . \tag{5.74}$$

This can be solved for $p = \sqrt{p^i p_i}$ as

$$p = \frac{m_A}{2}\sqrt{1 - \frac{4m_B^2}{m_A^2}} \ . \tag{5.75}$$

Let's consider what we have achieved so far. Without specifying anything about the mechanism that leads to particle A to decay into two B particles we have determined the energy and the magnitude of the momentum of the decay products. The only thing we do not know is the direction of the decay products, though we do know they are back-to-back in the rest frame of particle A. Notice also that Eq. (5.75) implies that the momentum is only real when $m_A \geq 2m_B$. Making the (reasonable) requirement that momenta must be real valued, we find a condition. The decaying particle must have at least as much mass as the sum of the particles it decays into. This makes good intuitive sense considering conservation of energy, but it is nice to see it enforced by the mathematics.

Exercise 5.15

Generalize the previous example by allowing the two decay products of particle A to have different masses, say m_B and m_C, so $A \to B + C$. You may find it useful to use the relationship $p^2 = m^2$ each particle. This relation and the conservation of four-momentum can be combined to show that

$$m_A^2 = m_B^2 + m_C^2 + 2\left(E_B E_C - \vec{p}_B \cdot \vec{p}_C\right) \ . \tag{5.76}$$

Use this result in the A rest frame to solve for the magnitude of the B and C three-momenta. From these results show that the daughter particle energies are

$$E_B = \frac{m_A}{2}\left(1 + \frac{m_B^2 - m_C^2}{m_A^2}\right) \ , \qquad E_C = \frac{m_A}{2}\left(1 + \frac{m_C^2 - m_B^2}{m_A^2}\right) \ . \tag{5.77}$$

Finally, what relation amongst the three particle masses *must* be satisfied for this decay to occur, that is what relation keeps the B and C three momenta real valued?

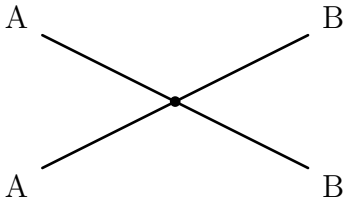


Figure 3. Production of B particles from A particle collision. Time proceeds from left to right.

Example

The next simplest process is two particles colliding and producing two other particles. Consider the simple case of two type A particles combining a producing two B type particles, $A + A \to B + B$. Conservation of four-momentum leads to

$$p^\mu_{A1} + p^\mu_{A2} = p^\mu_{B1} + p^\mu_{B2} \ . \tag{5.78}$$

In the previous example considering the rest frame of the decaying particle simplified much of the analysis. A similarly useful choice for the 2 to 2 process is the center-of-momentum frame with the A particle traveling along the z-axis. (Of course, choosing any particle axis is just as good, but typically the z-axis is chosen to exploit coordinates with cylindrical symmetry about the collision axis.)

In the center-of-momentum frame, the three momenta of the incoming and outgoing particles must be equal in magnitude and opposite in direction

$$p^i_{A1} + p^i_{A2} = 0 \ , \quad p^i_{B1} + p^i_{B2} = 0 \ . \tag{5.79}$$

So, we define the quantities

$$p^i_{A1} \equiv p^i_A \ , \quad p^i_{A2} = -p^i_A \ , \tag{5.80}$$

and similar definitions are made for p^i_B. This makes clear, that

$$E_{B1} = \sqrt{m_B^2 + p_B^2} = E_{B2} \equiv E_B \ , \tag{5.81}$$

and similarly that $E_{A1} = E_{A2} \equiv E_A$. Therefore, the $\mu = 0$ component of Eq. (5.78) implies

$$2E_A = 2E_B \equiv E_T \ . \tag{5.82}$$

He we have defined E_T as the total energy of the interaction. This leads to

$$\frac{1}{4}E_T^2 = m_B^2 + p_B^2 \ , \tag{5.83}$$

which can be solved for p_B:

$$p_B = \frac{E_T}{2}\sqrt{1 - \frac{4m_B^2}{E_T^2}} \ . \tag{5.84}$$

The B particle momentum looks similar to the momentum of particles produced from a decay given in Eq. (5.75). We have found again that much of the 2-to-2 scattering process is dictated by the structure of special relativity and the conservation of four-momentum. However, there is a significant difference in at least one aspect. Recall that the decay of an A particle into two B particles can only occur when $m_A \geq 2m_B$. In contrast, if we collide two A particles, we find from Eq. (5.84) that two B particles can be produced as long as $E_T \geq 2m_B$, that is even if the A particles have smaller masses than B particles we can use them to produce the Bs if we give the As enough energy. This is an essential part of why particle colliders are useful.

Exercise 5.16 Generalize the previous example to $A+A \to B+C$, that is, two different particles are created by the collision of two A particles. Begin in the center of momentum frame. As before, define the total energy of the collision is $E_T = 2E_A$. Relate the total energy to the B and C energies and solve for the magnitude of the B and C three-momenta. Show that the produced particles have energies

$$E_B = \frac{E_T}{2}\left(1 + \frac{m_B^2 - m_C^2}{E_T^2}\right) , \qquad E_C = \frac{E_T}{2}\left(1 + \frac{m_C^2 - m_B^2}{E_T^2}\right) . \tag{5.85}$$

What relation *must* E_T satisfy in order for this process to proceed, that is what relation keeps the B and C three momenta real valued?

Section 6

Collider Basics

We have seen in the last section that simply making an assumption about the nature of space and time (the spacetime of special relativity) leads to precise predictions about how particles interact. One way to test these and other predictions is to collider beams of particles and track the consequences. After decades of increasingly sophisticated analyses of particle collisions a useful language for describing the interactions has been developed.

The schematic set-up is to take two colliding particles moving on the z axis of a cylindrical coordinate system. One particle is moving in the positive z direction, the other is moving along negative z. The particles collide at $z = 0$, which is called the interaction point (IP). The angle θ sweeps from $\theta = 0$ on the positive z axis to $\theta = \pi$ on the negative z axis. The angle ϕ goes from $\phi = 0$ all the way around the z axis to $\phi = 2\pi$.

In a particle collision with the colliding particles along the z axis, the products of the collision can be tracked in the cylinder of the detector according to the z, θ, and ϕ coordinates of various detector components. However, these are not the variables typically used.

The **rapidity** of a particle with energy E and momentum in the z direction p_z is defined to be

$$y = \frac{1}{2}\ln\frac{E + p_z}{E - p_z} . \tag{6.1}$$

This dimensionless quantity is useful because of its behavior under boosts in the z direction.

Exercise 6.1 Boost a particle's four-momentum

$$p^\mu = \begin{pmatrix} E \\ p_x \\ p_y \\ p_z \end{pmatrix} , \tag{6.2}$$

by some velocity v along the z direction. Recall that a boost in the x direction was given in Eq. (5.28). Find the boosted energy and momentum in terms of the original energy and momentum. Then, write the boosted rapidity in terms of the original rapidity. Use this result to show that the difference between two rapidities is invariant under boosts in the z-direction.

In any real experiment there is variation in the incoming particle energies and the true location of the collision. This means that actual collisions depart from the idealized collision by some amount, which can often be related to the idealized center-of-momentum frame of collisions. However, if we describe the outgoing particles by their rapidities then the difference in rapidity of any two particles will be unchanged amongst frames related to the ideal conditions by a boost in z.

Another quantity that, like differences in rapidity, is invariant under boosts along the z-axis is the component of the momentum that are transverse to the z-axis. This motivates defining the **transverse momentum**

$$p_T = \sqrt{p_x^2 + p_y^2} \, . \tag{6.3}$$

This immediately allows us to use

$$\vec{p}^{\,2} = p_T^2 + p_z^2 = p_T^2 + \vec{p}^{\,2} \cos^2\theta \, , \tag{6.4}$$

to find that

$$p_T = |\vec{p}| \sin\theta \, . \tag{6.5}$$

Exercise 6.2 Show that p_T is unchanged by boosts along the z direction.

This property of invariance under boosts along the beam axis makes rapidity and transverse momentum very useful. Indeed, we can map cylindrical coordinates to these quantities. We can make some connection through the relation

$$p_z = |\vec{p}| \cos\theta \, , \tag{6.6}$$

which is similar to the usual relationship between Cartesian and spherical polar coordinates $z = r\cos\theta$.

Exercise 6.3

Show that

$$\frac{p_z}{E} = \tanh y \ , \tag{6.7}$$

and

$$E^2 = m^2 + p_T^2 + E^2 \tanh^2 y \quad \Rightarrow \quad E = \cosh y \sqrt{m^2 + p_T^2} \ . \tag{6.8}$$

Use these results to show that

$$p^\mu = \begin{pmatrix} E \\ p_x \\ p_y \\ p_z \end{pmatrix} = \begin{pmatrix} E \\ |\vec{p}| \cos\phi \sin\theta \\ |\vec{p}| \sin\phi \sin\theta \\ |\vec{p}| \cos\theta \end{pmatrix} = \begin{pmatrix} \sqrt{m^2 + p_T^2} \cosh y \\ p_T \cos\phi \\ p_T \sin\phi \\ \sqrt{m^2 + p_T^2} \sinh y \end{pmatrix} \tag{6.9}$$

We see that we can exchange the magnitude of the particle momentum $|\vec{p}|$ and the angle θ for p_T and y.

It is somewhat easier to map to a quantity called **pseudorapidity**

$$\eta \equiv -\ln \tan \frac{\theta}{2} \ , \tag{6.10}$$

which is directly related to the angle θ. Though not immediately obvious this is the rapidity of a massless particle. For a massless particle the relativistic energy

$$E = \sqrt{m^2 + \vec{p}^2} \ , \tag{6.11}$$

simply becomes $E = |\vec{p}|$. This means we can write rapidity for a massless particle as

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z} = \frac{1}{2} \ln \frac{|\vec{p}| \, (1 + \cos\theta)}{|\vec{p}| \, (1 - \cos\theta)} = -\ln \sqrt{\frac{1 - \cos\theta}{1 + \cos\theta}} \ . \tag{6.12}$$

It is a standard trigonometric identity that

$$\sqrt{\frac{1 - \cos\theta}{1 + \cos\theta}} = \tan \frac{\theta}{2} \ , \tag{6.13}$$

for $0 \leq \theta \leq \pi$. This means that for a massless particle

$$y = -\ln \tan \frac{\theta}{2} = \eta \ . \tag{6.14}$$

Exercise 6.4 Show that a *massless* momentum four-vector can be expressed as

$$p^\mu = p_T \begin{pmatrix} \cosh\eta \\ \cos\phi \\ \sin\phi \\ \sinh\eta \end{pmatrix} \ . \tag{6.15}$$

You may find it useful to show that

$$\sin\theta = \frac{1}{\cosh\eta} \ . \tag{6.16}$$

The form of the four-vector shows that for a massless particle we can exchange $|\vec{p}|$ and θ for p_T and η.

When the momentum of a particle is large compared to its mass we can write the energy as

$$E = |\vec{p}|\sqrt{1+\frac{m^2}{\vec{p}^2}} = |\vec{p}| + \frac{m}{2}\frac{m}{|\vec{p}|} + \dots \ . \tag{6.17}$$

If we only take the first term we have the relation between energy and momentum for a massless particle. This leads to the understanding that particles with momenta much larger than their masses can be approximated as massless. For example, consider a massless particle with momentum of 1 GeV and a particle with 1 MeV mass with a momentum of 1 GeV. The energy of the first particle is 1 GeV and the energy of the second is about 1.0000005 GeV. The finite mass has an effect, but it is a very small one compared to the energy of the particle.

Exercise 6.5 This effect can also be seen using η and p_T to characterize a particle. Begin by showing

$$E^2 = m^2 + \frac{p_T^2}{\sin^2\theta} = m^2 + p_T^2\cosh^2\eta \ . \tag{6.18}$$

Use this result to show that

$$p^\mu = p_T \begin{pmatrix} \cosh\eta\sqrt{1+\frac{m^2}{p_T^2\cosh^2\eta}} \\ \cos\phi \\ \sin\phi \\ \sinh\eta \end{pmatrix} \ . \tag{6.19}$$

Note that from above we have $|\vec{p}| = p_T\cosh\eta$. Therefore, when the momentum of a particle $\vec{p}|$ is larger than its mass m we can expand the

square root

$$\sqrt{1+\frac{m^2}{p_T^2\cosh^2\eta}} \approx 1+\frac{m^2}{2p_T^2\cosh^2\eta}+\dots\ . \tag{6.20}$$

This means that for a massive particle with large momentum (compared to its mass) we have

$$p^\mu = p_T\begin{pmatrix}\cosh\eta\\ \cos\phi\\ \sin\phi\\ \sinh\eta\end{pmatrix}+\frac{m}{2p_T\cosh\eta}\begin{pmatrix}1\\0\\0\\0\end{pmatrix}+\dots\ . \tag{6.21}$$

As expected the first term is the parameterization of a massless particle.

In general, the tracking components of a particle detector, at a collider, is described by ϕ and η. The particles are described in terms of their p_T. These quantities are most seem most useful for massless particles, but this same intuition can be used for particles with energy much larger than their mass.

Given two particles in a collision event a useful measure of how close the particles are is

$$\sqrt{(\Delta y)^2+(\Delta\phi)^2}\ . \tag{6.22}$$

Because ϕ and Δy are unchanged by boosts along the z-axis, so is this quantity. For massless particles we can replace the rapidity by the pseudo-rapidity. We therefore define

$$\Delta R = \sqrt{(\Delta\eta)^2+(\Delta\phi)^2} \tag{6.23}$$

to describe the angular distance between particles. This quantity is independent of boosts along the beamline (z-axis) only when the particles are massless. However, as we have seen, massive particles with mass much smaller than their momentum are can be approximated well as being massless.

We now discuss some basic principles of various types of particle colliders. The first type of experiment we hear about is something like Rutherford's gold foil experiment where a beam of particles is directed toward a stationary target. Let us suppose that experimentally we can create a beam of particles with mass m_B with energy E_B. The four momentum of

one beam particle can be taken to be

$$p_B^\mu = \begin{pmatrix} E_B \\ 0 \\ 0 \\ p_B \end{pmatrix} , \tag{6.24}$$

where p_B is the magnitude of the momentum, which is along the beam axis, taken to be the z-axis.

The target particles are taken to be at rest and have a mass of m_T. Their four-vector is simply

$$p_T^\mu = \begin{pmatrix} m_T \\ 0 \\ 0 \\ 0 \end{pmatrix} . \tag{6.25}$$

Now what are possible target particles? This depends on our beam particles. If we are use a high energy electron beam then we cannot consider the target to be bowling balls because in reality the electrons will only interact with the constituents of a single atom within the bowling ball. Indeed, if the electrons are at very high energy a proton is too large an object, and the electron primarily scatters off a single quark within the proton.

Given a beam particle and target particle, what is the total energy that is available to make new particles or otherwise probe nature's structure? The final energy from the two particles is simply

$$E_F = E_B + m_T , \tag{6.26}$$

but how much of this is available to the experiment? One way to determine this is to find the invariant mass of the combination of the two four-vectors

$$m_F^2 = (p_B^\mu + p_T^\mu)\,(p_{B\mu} + p_{T\mu}) . \tag{6.27}$$

Exercise 6.6 Show that

$$m_F = \sqrt{m_T^2 + m_B^2 + 2E_B m_T} . \tag{6.28}$$

The invariant mass is unchanged by under transformations, including a boost into the center of momentum frame. In that frame the final state has no net momentum and so all the incoming particle momentum feeds into energy available to the final state.

Example We would like to boost to a frame in which the beam and target particles have equal momentum in opposite directions. We denote this boosted

frame with primes and find

$$E'_B = \gamma(E_B - vp_B) \qquad E'_T = \gamma m_T \tag{6.29}$$
$$p'_B = \gamma(p_B - vE_B) \qquad p'_T = -\gamma v m_T \ , \tag{6.30}$$

where v is the velocity of the primed frame relative to the lab frame.

Being in the center of momentum frame means that

$$p'_B = -p'_T \ . \tag{6.31}$$

This implies that

$$v = \frac{p_B}{m_T + E_B} \ . \tag{6.32}$$

The total energy in this frame is simply

$$E'_F = E'_B + E'_T \ . \tag{6.33}$$

Exercise 6.7 Show that

$$E'_B + E'_T = \sqrt{m_T^2 + m_B^2 + 2E_B m_T} = m_F \ . \tag{6.34}$$

This confirms that the invariant mass of the initial state particles describes the final state energy in the center of momentum frame.

For high energy collisions, we have that $E_B \gg m_T, m_B$. This means that

$$m_F = \sqrt{2E_B m_T}\sqrt{1 + \frac{m_T^2 + m_B^2}{2E_B M_T}} \approx \sqrt{2E_B m_T} + \dots \ . \tag{6.35}$$

This is our characterization of how much energy can be extracted from these, so-called fixed target, collisions. Notice that the total energy is related to the geometric mean of the target mass and beam energy.

A fixed target experiment is somewhat simple. The beam is aimed at a collection of target particles larger than the beam width, which does not require delicate adjustments. It is much more difficult to direct two different particle beams at each other and obtain collisions. However, there are additional gains that can make the trouble worth it.

If we consider two beams directed along the z-axis in opposite directions then the characteristic four-momenta for the two beams are

$$p_1^\mu = \begin{pmatrix} E_B \\ 0 \\ 0 \\ p_B \end{pmatrix} \ , \quad p_2^\mu = \begin{pmatrix} E_B \\ 0 \\ 0 \\ -p_B \end{pmatrix} \ . \tag{6.36}$$

Here, for simplicity, we have assumed the beams are of the same energy

and particle type. As in the fixed target case, the quantity of interest is the invariant mass obtained from the sum of two four-vectors

$$\begin{aligned} m_F^2 =& \left(p_1^\mu + p_2^\mu\right)\left(p_{1\mu} + p_{2\mu}\right) \\ =& 4E_B^2 \ . \end{aligned} \tag{6.37}$$

Therefore, the ratio of the energy available from a colliding beams experiment to that of a fixed target experiment is approximately

$$\sqrt{2\frac{E_B}{m_T}} \ . \tag{6.38}$$

This means that for beam energies larger than the mass of a target particle we get a significant increase in the energy that can go into making new particles, or probing higher energies.

This, essentially, is why particle experiments often go to the trouble of trying to collider particle beams. As one might imagine, this is a nontrivial feat. If the collider design is simply a straight tube with a beam directed through at either end (this is really not exactly how linear colliders are set-up, but I think you get the idea) then the beams need to be fully accelerated to collision energy as well as carefully shaped and aimed in the short time it takes to get form the beam sources to the interaction point.

If instead of a straight line path toward the collision the beams travel around a circular ring then many revolutions can be used to get the particle up to their final energies and the beams prepared for collision. But of course, everything comes at a price. Typically beams are accelerated and steered using electromagnetic interactions, meaning that the beam particles must have electric charge. But accelerating electric charges radiate energy.

Both a linear and a circular collider need to accelerate the beams up to a given energy, but the circular collider also must accelerate the particles to keep them on a circular path. This additional radiation from circular colliders is called synchrotron radiation. The power radiated is given by the relativistic version of the standard Larmor formula

$$\mathcal{P}_{\text{rad}} = \frac{2\alpha q^2}{3}\gamma^4 a^2 \ , \tag{6.39}$$

where

$$\alpha = \frac{e^2}{4\pi} \approx \frac{1}{137} \ , \tag{6.40}$$

is the fine structure constant, q is the charge of accelerating particle ($q = -1$ for an electron), and a is the acceleration. We then use the definition of relativistic energy to express

$$\gamma = \frac{E_B}{m_B} \ . \tag{6.41}$$

The usual relationship for the acceleration required for circular motion is

$$a = \frac{v^2}{R} \ , \tag{6.42}$$

where v is the velocity of the particle and R is the radius of the circular path. This allows us to express the radiated power as

$$\mathcal{P}_{\text{rad}} = \frac{2\alpha}{3}\left(\frac{qE_B^2 v^2}{m_B^2 R}\right)^2 \ . \tag{6.43}$$

This result describes the amount of energy lost (per unit time and per particle) due to the circular path of the particle beams. This leads to two issues. First, the accelerator must supply this much power (per particle) to keep the beam at a given energy. Second, all of this energy is radiated by the particles into the structure of the collider and must be efficiently removed from the system to prevent damage to collider components.

Notice that as the beam energy increases the radiated power goes up by the fourth power. This means increasing the beam energy by a factor of 2 increases the radiated power by a factor of 16! This formula also shows how the radiated power can be reduced. If the radius of the collider ring is increased or the mass of the colliding particles are increased the radiated power is reduced. This is why circular colliders have become increasingly large over the years and why proton colliders have generally been higher energy machines than electron colliders of equal radius.

Exercise 6.8 The Large Electron Positron collider (LEP) reached a maximum center of mass energy of number GeV. If the electron beams were be replaced with protons what energy collider would lead to the same power in synchrotron radiation? Compare this to the center of mass energy of the Large Hadron Collider (LHC).

Exercise 6.9 The muon is a particle with identical quantum numbers to the electron, except that it mass m_μ is about 200 times larger than the electron mass m_e. If the electron beams at LEP could be replaced with muons what energy collider would lead to the same power in synchrotron radiation?

Subsection 6.1

Detectors and Cross Section

We have now developed some ideas about the collider beams. Before going on we should say a little about the detectors. In many collider experiments two beams are brought together so that they collide. The point of collision is called the Interaction Point (or sometimes the IP). From this point a spray of particles of various type and momenta are produced. It is these

produced particles that we are typically interested in and the collider detectors are designed to provide as much information about them as possible. We also introduce the idea of the cross section associated with particle interactions. The discussion assumes some basic understanding of quantum mechanics. Some readers may find it useful to read the first portion of Sec. 33 (up to 33.1) to prepare.

At the Large Hadron Collider (LHC) the detectors are very large, about the size of a four or five story building. This size is taken up with several different detector layers, which can be roughly divided into four types. These all surround the beam like a solid cylinder. The innermost layer is the tracker, which is filled with layers of material that register the flight of electrically charged particles, allowing them to be tracked. The trajectories of these particles also provide information about them. The next layer is the electronic calorimeter or ECAL. This is where many light electrically charged particles (like electrons) and photons are stopped, depositing their energy. Around this is the hadronic calorimeter (HCAL) which stops particles that carry the strong nuclear force. Finally, the muon system tracks muons (and things that look like muons). These particles are too massive to be stopped by the ECAL and don't interact much with the HCAL. The muon system does not stop them, but gives more information about their trajectories.

All of this detection produces a huge amount of information for each collision (about one Megabyte [23]) and the LHC generates about 1 billion collisions per second [24]. This generates much more data (about 1 Petabyte per second) than can be reasonably stored or read out by the experimental collaborations. In addition, it is likely that many of these collisions do not produce anything new or particularly interesting. This is why the experiments employ triggers. Only when certain requirements are satisfied are the collision data are recorded, that is only for events that satisfy the triggers. At the LHC only a tiny fraction of the collision data are currently recorded. These data are then used for various specific experimental analyses.

In a collider experiment the output is simply the number of events of some particular type. Once we decide what type of event we are interested in we simply count (actually it is not simple at all to both run the machine and design and operate the detectors that do the counting, but the big picture idea is simple) the number of events of that type that occur while the collider is running. But how do we connect this number of events that we count to structure of nature?

The basic idea is that the interactions related to the particles that compose the beams is what we are studying. To focus on these interactions we need to characterize the number of events we measure in terms of these interactions. To do this, let us consider two beams of particles that are directed into a collision course. These beams are typically composed of

discrete bunches that have a characteristic length ℓ and cross sectional area $\mathcal{A}$. Within each bunch there is some number of particles. We consider the collision of bunch A with bunch B. Bunch A has some number N_A of particles or, more usefully, a number density

$$n_A = \frac{N_A}{\mathcal{A} \cdot \ell} \ . \tag{6.44}$$

This bunch is traveling at velocity v_A. Similar definitions apply to bunch B.

What happens when these two bunches collide? First, we can determine how long it takes a particle from one bunch to pass through the other. This simple the magnitude of their relative velocity $|v_A - v_B|$ divided by the length of the bunch ℓ

$$t = \frac{|v_A - v_B|}{\ell} \ . \tag{6.45}$$

Next, how many interaction events occur within this time? We expect

$$\frac{\text{Events}}{\text{Time}} \propto \frac{N_A N_B}{t} \ , \tag{6.46}$$

because the larger the number of particles in each bunch the more likely they are to interact. To define an equality, we multiply the right-hand side by $\sigma/\mathcal{A}$. We call σ the **cross section** and note that by definition it has dimensions of area. In essence, it describes how likely one of the particles in one bunch is to interact with one of the particles in the other bunch. If these particles behaved like classical hard spheres then σ would simply be the cross sectional area of one sphere. The ratio of that area to the total area of the beam (assuming they are randomly distributed) provides the probability of two particles colliding. In the quantum mechanical world that governs these collider events the cross section is not simply geometrical, but plays the same role.

We can then write the number of events per time as

$$\begin{aligned} \frac{\text{Events}}{\text{time}} &= \frac{N_A N_B}{t} \frac{\sigma}{\mathcal{A}} = \frac{N_A N_B}{\mathcal{A} \cdot \ell} |\vec{v}_A - \vec{v}_B| \sigma = n_A n_B \mathcal{A} \cdot \ell |\vec{v}_A - \vec{v}_B| \sigma \\ &\equiv \mathcal{L}\, \sigma \ . \end{aligned} \tag{6.47}$$

In this last line we have defined the **luminosity** $\mathcal{L}$. This is a function of quantities that are determined by the experimenter. Things like the relative velocity, the number of particles in a bunch and the size of the bunches. Only the cross section is related to the fundamental physics that governs the particle interactions. This makes Eq. (6.47) exactly what we want. The number of events we record over some time is related to the set-up of the experiment and the cross section, which is set by nature.

An individual particle physics model makes specific predictions for the cross sections of various processes. By using colliders and other experiments we can compare various types of cross sections, those that correspond to various specific types of events, to actual experimental results. If these disagree then we need to consider a new model. If they agree, then we gain confidence that we are on the right track.

As mentioned, cross sections are measured in units of area. During the early days of nuclear physics a standard unit of cross section was defined: the barn

$$1 \text{ barn } = 10^{-24}\text{cm}^2 \ . \tag{6.48}$$

Of course, in natural units we associate lengths with inverse energy, so an area is inverse energy squared. However, typically cross sections are still reported in terms of barns, or more usually fractions of barns. For instance, picobarns (pb) which are 10^{-12} barns or femtobarns (fb) which are 10^{-15} barns.

Exercise 6.10 Show that

$$\text{GeV}^{-2} = 3.894 \times 10^{-4} \text{ barns} \ . \tag{6.49}$$

The integrated luminosity is the total luminosity over a given time. Because the integrated luminosity multiplied by the cross section produces a pure number, the number of events, the dimensions of integrated luminosity are inverse area. For ease of use with cross sections, these are often reported in terms of inverse barns. For instance, a certain LHC search might use 36 fb^{-1} of recorded luminosity, this is read as 36 inverse femtobarns.

We now consider how the cross section is calculated. Because it is related to the probability of particles interacting to produce a specific final state we expect

$$\sigma \propto |\langle \text{out}|\text{in}\rangle|^2 \ , \tag{6.50}$$

where $|\text{in}\rangle$ is the state that describes the incoming particles and $|\text{out}\rangle$ is the state that describes the outgoing particles. Then complex number $\langle \text{out}|\text{in}\rangle$ is the inner product, or overlap, of these two states. The cross section can only be proportional to this squared quantum amplitude because the amplitude is dimensionless while the cross section goes like inverse energy squared. If either particle is going faster, having larger energy, then it "spends less time" near other particles and hence its cross section goes down. This motivates

$$\sigma \propto \frac{1}{E_A E_B} |\langle \text{out}|\text{in}\rangle|^2 \ . \tag{6.51}$$

However, this does not quite have the properties we expect. Suppose we take the colliding particles to travel along the z-direction. The area they present should be invariant under boosts along the z-direction, just like a geometrical area would be.

Exercise 6.11 Recall that after a boost in the z-direction

$$E \to \gamma\,(E + v\,p_z)\ , \quad p_z \to \gamma\,(p_z + v\,E)\ , \tag{6.52}$$

and the velocity (in the z direction) of the ith particle is

$$v_i = \frac{p_{iz}}{E_i}\ . \tag{6.53}$$

With these results in mind, show that

$$\frac{1}{E_A E_B} \to \frac{1}{\gamma^2 E_A E_B\,(1 + v\,v_A)\,(1 + v\,v_B)}\ , \tag{6.54}$$

and

$$|v_A - v_B| \to \frac{|v_A - v_B|}{\gamma^2\,(1 + v\,v_A)\,(1 + v\,v_B)}\ . \tag{6.55}$$

From these results conclude that under boosts along the z-direction that

$$\frac{1}{E_A E_B |v_A - v_B|} \to \frac{1}{E_A E_B |v_A - v_B|}\ . \tag{6.56}$$

The above exercise shows that

$$\sigma \propto \frac{1}{E_A E_B |v_A - v_B|}\,|\langle \text{out}|\text{in}\rangle|^2\ . \tag{6.57}$$

Which has the correct dimensions and transformation properties. A careful quantum field theory calculation produces

$$\sigma = \frac{1}{2E_A\,2E_B\,|v_A - v_B|} \times \int \left[\prod_{i=1}^{n} \frac{d^4 p_i}{(2\pi)^4} 2\pi\delta(p_i^2 - m_i^2)\right] \times |\mathcal{M}|^2\,(2\pi)^4 \delta^{(4)}\left(p_A + p_B - \sum_{i=1}^{n} p_i\right)\ , \tag{6.58}$$

where $\mathcal{M}$ is referred to as the **Lorentz invariant amplitude**. The coefficient in front we understand from dimensions and behavior under boosts. The product under the integral sign spans all n particles that are produced by the collision of particles A and B. The first terms are integrals over their four-momenta, subject to the constraint $p_i^2 = m_i^2$. We also enforce total momentum conservation.

Exercise 6.12 Suppose we know that some particle X of mass M sometimes decays to two Y particles. For simplicity we assume the process is spherically symmetric and that Y is effectively massless. The formula, in the rest

frame of the X particle, is similar to that of the cross section

$$\Gamma = \frac{1}{2M}\int \frac{d^3\vec{p}_1}{2E_1(2\pi)^3}\frac{d^3\vec{p}_2}{2E_2(2\pi)^3}(2\pi)^4\delta(M-E_1-E_2)\delta^{(3)}(\vec{p}_1+\vec{p}_2)\,|\mathcal{M}|^2 \ , \tag{6.59}$$

where $\vec{p}_1$ and $\vec{p}_2$ are the three-momenta of the decay products and $E_i = |\vec{p}_i|$ because the decay products are massless. Assuming that $|\mathcal{M}|^2$ is spherically symmetric (does not depend on θ and ϕ), show that

$$\Gamma = \frac{1}{16\pi M}\,|\mathcal{M}|^2\Big|_{|\vec{p}_1|=|\vec{p}_2|=M/2} \ . \tag{6.60}$$

Within these general considerations for the cross section of a particular process, only the value of $\mathcal{M}$ depends on the particular model in question. Thus, the main goal of quantum field theory calculations is to determine $\mathcal{M}$ for a given process so that those predictions can be compared with experiment.

While the details of any process depends on what model is being used, we can make a few general statements. For instance, the dimensions of $|\mathcal{M}|^2$ changes with the number of particles in the final state. Each particle's phase space factor is dimension two, but the decay width is alway dimension one and the cross section always has dimensions of minus two. This implies that the dimension of the amplitude $\mathcal{M}$ changes as the number of final state particles changes.

It is also true that the decay width or cross section decreases with the number of final state particles. This can be understood by considering the phase space integral over a constant argument

$$\int \frac{d^3\vec{p}}{2E(2\pi)^3} = \frac{4\pi}{2(2\pi)^3}\int \frac{dp\,p^2}{\sqrt{m^2+p^2}} \ , \tag{6.61}$$

where m is the mass of the final state particle. If the total amount of energy E_f that flows into this particle is much larger than its mass then this integral is approximately

$$\frac{E_f^2}{8\pi^2} \ , \tag{6.62}$$

and if the $E_f \ll m$ then we find

$$\frac{E_f}{m}\frac{E_f^2}{3\pi^2} \ . \tag{6.63}$$

Both results illustrate the dimension of the phase space integral, which must be compensated by the amplitude. Aside from this dimensionful quantity, both integrals are small. Because the available energy must be shared among all the final state particles and because of these reductions that come from additional phase space integrals, processes with more particles

in the final state are typically less likely to occur than those with fewer particles.

Section 7

Particles and the Lorentz Group

We have already seen how powerful the structure of special relativity is on the simple kinematics of particle interactions. In this section we extend its reach further. Before particle physics can have real meaning, we need to know what is meant by a particle. In a heuristic way elementary particles are the smallest bits of stuff. But in seeking to efficiently and effectively model nature we want to used descriptions that are well suited to nature, at least as we currently understand it. All the evidence we currently understand indicates that nature respects the structure of special relativity. This is encapsulated mathematically in the Lorentz group O(1,3).

What does it mean for our model to be well suited to this symmetry? In essence, actions of the symmetry should not change the essential characteristics the objects we are considering. We do not want to have an important building block of our model significantly change its character when we rotate through an angle or boost to a new frame. So, what we call a particle is something in that some way stays similar, or of a certain type, under Lorentz transformations. Mathematically, this means that different types of particles should all be described by a single irreducible representation of the Lorentz Group. Further discussions of particle representations of the Lorentz group can be found in [11, 14, 15, 25].

Recall from Sec. 5 that much of relativity can be understood by using four-vectors and the Minkowski metric

$$\eta_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} . \tag{7.1}$$

We also discussed a collection of transformations Λ^{μ}_{ν} (Lorentz transformations) that satisfy

$$\eta_{\mu\nu} = \Lambda^{\alpha}{}_{\mu}\Lambda^{\beta}{}_{\nu}\eta_{\alpha\beta} . \tag{7.2}$$

These matrices form a group that is denoted by O(1,3) (or sometimes O(3,1)). If the concept of a mathematical group is unfamiliar, then you should review Sec. 36 before continuing. We mostly focus on the subgroup of transformations that can be continuously connected to the identity. For instance, taking a rotation matrix to an angle of $\theta = 0$ just gives the identity. This subgroup, labeled SO(1,3), encodes all the information about rotations in three dimensional space and also about boosting from one

inertial frame to another. Both rotations and boosts can be described by continuous parameters, indicating (correctly) that this group is a Lie (pronounced Lee) group. A short introduction to Lie groups can be found in Sec. 37.

Exercise 7.1 We can write a Lorentz transformation that is just a "little" change from the identity as

$$\Lambda^{\mu}{}_{\nu} = \delta^{\mu}_{\nu} + \omega^{\mu}{}_{\nu} \ , \tag{7.3}$$

where $\omega^{\mu}{}_{\nu}$ is understood to be small, that is all of its components are much smaller than one. Use the defining characteristic of Lorentz transformations given in Eq. (7.2) to first order (linear powers of) $\omega_{\mu\nu}$ to show that it is antisymmetric. From this conclude that Lorentz transformations are specified by six generators.

Like all Lie groups, this means that any particular group element Λ can be written as

$$\Lambda = e^{i\alpha^a T^a} \ , \tag{7.4}$$

where the α^as are real parameters and the T^as are elements of the associated Lie algebra $\mathfrak{so}(1,3)$. Lie algebras are reviewed in Sec. 38. The Lie algebra is defined by how basis elements T^a behave under the Lie bracket (which for our purposes is the commutator of two matrices)

$$\left[T^a,\, T^b\right] = T^a T^b - T^b T^a = i f^{abc} T^c \ , \tag{7.5}$$

where the f^{abc} are the structure constants of the algebra. There are six generators of the Lie algebra: three Hermitian generators related to rotations about the three spatial directions J_i and three anti-Hermitian generators related to boosts along the three spatial dimensions K_i. The structure of the algebra is determined by the following relations among them

$$[J_i,\, J_j] = i\varepsilon_{ijk} J_k \ , \quad [J_i,\, K_j] = i\varepsilon_{ijk} K_k \ , \quad [K_i,\, K_j] = -i\varepsilon_{ijk} J_k \ . \tag{7.6}$$

Exercise 7.2 Using the commutators given above, show that

$$J_i^{\pm} \equiv \frac{1}{2}\left(J_i \pm i K_i\right) \ , \tag{7.7}$$

are Hermitian operators and have the commutation relations

$$\left[J_i^+,\, J_j^+\right] = i\varepsilon_{ijk} J_k^+ \ , \quad \left[J_i^-,\, J_j^-\right] = i\varepsilon_{ijk} J_k^- \ , \quad \left[J_i^+,\, J_j^-\right] = 0 \ . \tag{7.8}$$

This is the algebra structure $(\mathfrak{su}(2)\oplus\mathfrak{su}(2))$ for the product group SU(2)$\times$ SU(2).

The exercise shows that we can rewrite the Lie algebra in terms of

two generators that behave like independent $\mathfrak{su}(2)$ algebras. This is very useful because we know a lot about $\mathfrak{su}(2)$. In particular, we know all the finite-dimensional, unitary representations of $\mathfrak{su}(2)$. These are labeled by a half-integer j.

Exercise 7.3 The group elements of SU(2)×SU(2) are of the form

$$O = e^{-i\theta_i J_i^+ - i\vartheta_i J_i^-} \ , \tag{7.9}$$

where θ and ϑ are real numbers. Show that these can be represented by unitary operators. The element of the Lorentz group SO(3,1) have the form

$$\Lambda = e^{-i\theta_i J_i - i\beta_i K_i} \ , \tag{7.10}$$

where θ and β are real numbers. Write Λ in terms of the $J_i^\pm$ generators and compare it to the operators O given above.

The previous exercise shows that while SO(3,1) has mathematical connection to SU(2)× SU(2) they are not quite the same. The latter is a compact group with finite dimensional unitary representations. The former, the group we are actually interested in, shares much of the local structure but has noncompact aspects and so its unitary representations are infinite dimensional. The groups are different, but the Lie algebras are the same

$$\mathfrak{so}(3,1) \simeq \mathfrak{su}(2) \oplus \mathfrak{su}(2) \tag{7.11}$$

So, we use the SU(2)× SU(2) language to develop finite dimensional representations of the Lorentz algebra. We then connect these to infinite dimensional unitary representations of the group.

Because we can label all the finite-dimensional representations of $\mathfrak{su}(2)$ by a half-integer j we can label all the finite-dimensional representations of $\mathfrak{so}(3,1)$ by an ordered pair of half-integers (j_1, j_2). For instance, the $(0, 0)$, $(\frac{1}{2}, 0)$, and $(0, \frac{1}{2})$ representations. This labeling is quite clear, but often not how different representations are referred to. Notice that $J_z = J_z^+ + J_z^-$, this means that a state in the (j_1, j_2) representation has angular momentum $j_1 + j_2$. In other words, representations with nonzero j_1 or j_2 carry an intrinsic angular momentum that has nothing to do with orbiting something. It is this intrinsic angular momentum that is given the name **spin**.

The spin-0 representation, $(0, 0) \to \mathbf{0} \otimes \mathbf{0} = \mathbf{0}$, is often called the scalar representation and is one dimensional. The Higgs boson is an example of a scalar particle. There are two spin-1/2 representations: $(\frac{1}{2}, 0) \to \mathbf{\frac{1}{2}} \otimes \mathbf{0} = \mathbf{\frac{1}{2}}$ and $(0, \frac{1}{2}) \to \mathbf{0} \otimes \mathbf{\frac{1}{2}} = \mathbf{\frac{1}{2}}$. It is important to note that while these two representations both have the same intrinsic angular momentum, they are distinct, two-dimensional representations of the Lorentz group. These are

referred to as **spinor** representations. Familiar objects like the electron and the proton have spin-1/2, and so are described by spinors.

Particles like the photon have spin-1. One might consider the three dimensional $(1,0)$ or $(0,1)$ representations to describe the photon. However, we shall find below that we actually want the $(\frac{1}{2},\frac{1}{2}) \to \mathbf{\frac{1}{2}} \otimes \mathbf{\frac{1}{2}} = \mathbf{1} \oplus \mathbf{0}$ representation. This is a direct sum of the three dimensional spin-1 and one dimensional spin-0 representations. This is exactly right for a Lorentz four-vector, which is why $(\frac{1}{2},\frac{1}{2})$ is often referred to as the vector representation. In the next few chapters we examine the fields that are related to these representations.

Section 8

The Klein-Gordon Equation

In this section we take our first steps in describing a the quantum field that gives rise to a particle. This section assumes some familiarity with quantum mechanics. A short introduction that covers the most relevant topics is found in Sec. 33.

The first field equation we consider can be motivated by the Schrödinger equation for a free particle:

$$i\partial_t \psi(t,x) = \frac{i^2}{2m}\partial_i\partial^i\psi(t,x) \ . \tag{8.1}$$

We have written this in a way that shows the use of the momentum operator in the position basis $\widehat{P}^i = i\eta^{ij}\partial_j$ where the index is only over the three spatial dimensions. Recall that using the full four-dimensional Minkowski spacetime metric

$$\eta_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \ , \tag{8.2}$$

we have $\widehat{P}_i = \eta_{ij}\widehat{P}^j = i\partial_i$, In other words, the Schrödinger equation is

$$i\partial_t \psi(t,x) = \frac{\widehat{P}_i\widehat{P}^i}{2m}\psi(t,x) \ . \tag{8.3}$$

The term on the right-hand side looks like the kinetic energy of a particle, and for a free particle it would be the total energy. This suggests we might consider an energy operator $\widehat{E} = i\partial_t$. The equation is then just a way of saying $E = \vec{p}^{\,2}/(2m)$.

In relativity the energy of a free particle is actually given by $E^2 - \vec{p}^{\,2} =$

m^2. If we then write this in terms of operators we have

$$-\partial_t^2\phi(t,x) - \partial_i\partial^i\phi(t,x) = m^2\phi(t,x) \ . \tag{8.4}$$

This is the Klein-Gordon equation, simply a relativistic generalization of the Schrödinger equation [26–28].

Exercise 8.1 Show that the Klein-Gordon equation can be written as

$$\left(\partial_\mu\partial^\mu + m^2\right)\phi(t,x) = 0 \ . \tag{8.5}$$

What are some aspects of this equation? First, we note that the equation is Lorentz invariant. This can been seen by noting that the inner product of any two four-vectors V^μ and U^ν is invariant under Lorentz transformations $V^\mu \to \Lambda^\mu{}_\alpha V^\alpha$ and $U^\nu \to \Lambda^\nu{}_\beta U^\beta$:

$$\begin{aligned} V^\mu U_\mu =& V^\mu U^\nu \eta_{\mu\nu} \to V^\alpha U^\beta \Lambda^\mu{}_\alpha \Lambda^\nu{}_\beta \eta_{\mu\nu} \\ =& V^\alpha U^\beta \eta_{\alpha\beta} = V^\alpha U_\alpha \ . \end{aligned} \tag{8.6}$$

In the second line we have used the defining characteristic of the Lorentz transformations, that they preserve the form of the Minkowski metric: $\eta^{\mu\nu} = \Lambda^\mu{}_\alpha \Lambda^\nu{}_\beta \eta^{\alpha\beta}$. The field ϕ is a scalar, so it is also Lorentz invariant. Because the equations are Lorentz in variant, so too must be the solutions.

Exercise 8.2 Show that for a normalization constant a_p and parameters E and $\vec{p}$

$$\phi(x) = a_k e^{-i(\omega t - \vec{k}\cdot\vec{x})} = a_k e^{-ix^\mu k_\mu} \ , \tag{8.7}$$

is a solution of the Klein-Gordon equation if the dispersion relation

$$\omega^2 = \vec{k}^{\,2} + m^2, \tag{8.8}$$

is satisfied. What does this dispersion relation remind you of?

The exercise above shows that plane-waves with a specific dispersion relation are solutions of the Klein-Gordon equation. Specifically,

$$\omega_k = \pm\sqrt{\vec{k}^{\,2} + m^2} \ . \tag{8.9}$$

This looks exactly like the equation for relativistic energy $E = \sqrt{m^2 + \vec{p}^2}$. However, the minus sign is a signal that we cannot interpret the Klein-Gordon equation as a single particle wavefunction. Were we to try we would find both positive energy and negative energy states. In fact, it is not just that the energy is negative (which we are used to when dealing with bound states) but that it has no lower bound, making $|\vec{k}|$ larger makes

the energy lower. Instead of a single particle picture, we consider energy and momentum flowing through a field that fills all of space. This field is an operator $\widehat{\phi}(t,x)$ that can act on the states $|\Psi\rangle$ that describe various physical scenarios. We see below that this system can be thought of as having only positive energy configurations. In what follows we also drop the hat on fields and treat them as classical fields.

A general solution to the Klein-Gordon equation is a linear combination of all the solutions we have found (the so-called plane wave solutions) where we identify $\vec{k}$ and the three-momentum $\vec{p}$. In terms of $\vec{p}$ dependent coefficients $a(\vec{p})$ and $b(\vec{p})$ we have

$$\phi = \int d^3p \left(a(\vec{p}) \exp\left[-i(\omega_p t - \vec{p}\cdot\vec{x})\right] + b(\vec{p}) \exp\left[-i(-\omega_p t - \vec{p}\cdot\vec{x})\right]\right) \ , \tag{8.10}$$

where the limits of integration go from $-\infty$ to ∞. By changing the integration variables $\vec{p} \to -\vec{p}$ in the second term only we find

$$\phi = \int d^3p \left(a(\vec{p}) e^{-ix^\mu p_\mu} + b(-\vec{p}) e^{ix^\mu p_\mu}\right) \ , \tag{8.11}$$

where $p_\mu = (\omega_p, \vec{p})$. Finally, we require that ϕ be a real field. This means that $\phi^\dagger = \phi$. This in turn implies that $b(-\vec{p}) = a^\dagger(\vec{p})$. In other words we have

$$\phi = \int d^3p \left(a(\vec{p}) e^{-ix^\mu p_\mu} + a^\dagger(\vec{p}) e^{ix^\mu p_\mu}\right) \ . \tag{8.12}$$

In quantum field theory $a^\dagger(\vec{p})$ and $a(\vec{p})$ are operators, specifically they are creation and annihilation operators like those of the quantum harmonic oscillator discussed in Sec. 33.4. This illustrates again that $\phi(x)$ is a quantum operator that acts on quantum states, not a quantum state itself. This also indicates that the negative frequency parts of the field do not correspond to a negative energy particle. Instead, they are essential to describing how positive energy flows from one part of the field to another.

This can be understood, schematically, as follows. As a quantum operator the field ϕ can both create particles with momentum $\vec{p}$ through $a^\dagger(\vec{p})$ and destroy particles of momentum $\vec{p}$ with $a(\vec{p})$. Again, the field cannot be interpreted like a wavefunction, it does not describe a quantum state. It is an operator that can be used to map quantum states with a certain number of particles to other quantum states with a different number of particles.

Subsection 8.1

The Klein-Gordon Lagrangian

Now that we have the Klein-Gordon the equation and its solutions we can try to find Lagrangian density that leads to that equation. If the term "Lagrangian" does not mean anything to you then you should review Sec. 32 up to the beginning of Sec. 32.1 before reading on. This, however,

only covers single particle Lagrangian dynamics. The Lagrangian densities of field theories are reviewed Sec. 32.2. In doing so, we remind the reader that the functional derivative

$$\frac{\partial(\partial_\mu\phi)}{\partial(\partial_\alpha\phi)} = \delta^\alpha_\mu \ . \tag{8.13}$$

In short, a lower index in the denominator becomes an upper index.

Exercise 8.3 Show that the Lagrangian density

$$\mathcal{L} = \frac{1}{2}\left(\partial_\mu\phi\right)\left(\partial^\mu\phi\right) - \frac{m^2}{2}\phi^2 \ , \tag{8.14}$$

produces the Klein-Gordon equation from the Euler-Lagrange equation

$$\partial_\mu \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)} = \frac{\partial\mathcal{L}}{\partial\phi} \ . \tag{8.15}$$

You may find it useful to first write the Lagrangian density with $\partial^\mu\phi$ written in terms of $\partial_\nu\phi$.

Not that we have a Lagrangian density in hand, we can determine the Hamiltonian density. When integrated over all space this defines the energy of the system and so allows us to check that it is bounded from below. Following the procedure outlined in Sec. 32.2 we find

$$\begin{aligned}
\mathcal{H} =& \dot\phi\frac{\partial\mathcal{L}}{\partial\dot\phi} - \mathcal{L} \\
=& \dot\phi^2 - \frac{1}{2}\left[\dot\phi^2 + (\partial_i\phi)(\partial^i\phi) - m^2\phi^2\right] \\
=& \frac{1}{2}\dot\phi^2 - \frac{1}{2}(\partial_i\phi)(\partial^i\phi) + \frac{1}{2}m^2\phi^2 \\
=& \frac{1}{2}\dot\phi^2 + \frac{1}{2}(\nabla\phi)\cdot(\nabla\phi) + \frac{1}{2}m^2\phi^2 \ .
\end{aligned} \tag{8.16}$$

The first term is sensitive to changes in the field with respect to time, similar to a velocity. This term is always positive or zero. When the field is static there is no contribution to this piece. The second term is nonzero when there are spatial changes, or gradients, in the field. This too is never negative. Then the field is uniform throughout space this term is zero. The final term is sensitive to the actual value of $\phi(x)$. It is always positive (for $m^2 > 0$) and is only zero when $\phi = 0$. In short, we see that the energy density (the Hamiltonian density) is nonnegative, meaning that the energy of all field configurations is bounded from below, as promised.

Before moving on, notice that the Klein-Gordon equation is easy to generalize to include nonlinear self-interactions of the field. If we define a

potential for ϕ

$$U(\phi) = \frac{1}{2}m^2\phi^2 + \frac{1}{3}c_3\phi^3 + \frac{1}{4}c_4\phi^4 + \dots \ , \tag{8.17}$$

then the Lagrangian density

$$\mathcal{L} = \frac{1}{2}\left(\partial_\mu\phi\right)\left(\partial^\mu\phi\right) - U(\phi) \ , \tag{8.18}$$

produces a similar equation.

Exercise 8.4 Find the equation of motion of the Lagrangian density

$$\mathcal{L} = \frac{1}{2}\left(\partial_\mu\phi\right)\left(\partial^\mu\phi\right) - \frac{1}{2}m^2\phi^2 - \frac{\lambda}{4}\phi^4 \ . \tag{8.19}$$

Notice that we can also generalize to several scalar fields:

$$\mathcal{L} = \frac{1}{2}\left(\partial_\mu\phi_1\right)\left(\partial^\mu\phi_1\right) + \frac{1}{2}\left(\partial_\mu\phi_2\right)\left(\partial^\mu\phi_2\right) - \frac{m_1^2}{2}\phi_1^2 - \frac{m_2^2}{2}\phi_2^2 \ , \tag{8.20}$$

which leads to the equations

$$\left(\partial_\mu\partial^\mu + m_1^2\right)\phi_1 = 0 \ , \qquad \left(\partial_\mu\partial^\mu + m_2^2\right)\phi_2 = 0 \ . \tag{8.21}$$

This, on its own, is not very interesting, the fields do not even interact with each other. However, if we assume $m_1 = m_2 \equiv m$ then we have

$$\mathcal{L} = \frac{1}{2}\left(\partial_\mu\phi_1\right)\left(\partial^\mu\phi_1\right) + \frac{1}{2}\left(\partial_\mu\phi_2\right)\left(\partial^\mu\phi_2\right) - \frac{m^2}{2}\left(\phi_1^2 + \phi_2^2\right) \ . \tag{8.22}$$

This may not look much more interesting, yet.

Exercise 8.5 Beginning with the Lagrangian density in Eq. (8.22), make the substitutions

$$\phi_1 = \phi_1'\cos\theta - \phi_2'\sin\theta \ , \qquad \phi_2 = \phi_1'\sin\theta + \phi_2'\cos\theta \ , \tag{8.23}$$

where θ is any real number. Show that in terms of these new fields the Lagrangian takes the form

$$\mathcal{L} = \frac{1}{2}\left(\partial_\mu\phi_1'\right)\left(\partial^\mu\phi_1'\right) + \frac{1}{2}\left(\partial_\mu\phi_2'\right)\left(\partial^\mu\phi_2'\right) - \frac{m^2}{2}\left(\phi_1'^2 + \phi_2'^2\right) \ . \tag{8.24}$$

The above exercise shows that there is a family of changes of variables that preserve the form of the Lagrangian. Any such transformation is called a symmetry of the Lagrangian and typically leads to useful and interesting physical consequences. The symmetry considered above is particularly

interesting because it is parameterized by a continuous variable, θ. Emmy Noether showed that any such continuous symmetry leads to a conserved quantity, as reviewed in Sec. 32.3.

Example

From Sec. 32 we have that

$$\begin{aligned} J^\mu =& \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi_1)} \left.\frac{d\phi_1}{d\theta}\right|_{\theta=0} + \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi_2)} \left.\frac{d\phi_2}{d\theta}\right|_{\theta=0} \\ =& (\partial^\mu \phi_1)(-\phi_2) + (\partial^\mu \phi_2)(\phi_1) \\ =& \phi_1 \partial^\mu \phi_2 - \phi_2 \partial^\mu \phi_1 \ . \end{aligned} \tag{8.25}$$

We can check that this current is conserved by evaluating

$$\begin{aligned} \partial_\mu J^\mu =& \partial_\mu(\phi_1)\partial^\mu \phi_2 + \phi_1 \partial_\mu \partial^\mu \phi_2 - \partial_\mu(\phi_2)\partial^\mu \phi_1 - \phi_2 \partial_\mu \partial^\mu \phi_1 \\ =& \phi_1(-m^2\phi_2) - \phi_2(-m^2\phi_1) = 0 \ , \end{aligned} \tag{8.26}$$

where we have used the equations of motion in the second line. Note that the conservation of this current depends crucially on the masses of each field being identical, which is also what makes the Lagrangian symmetric.

We then define the conserved quantity

$$Q = \int d^3x J^0 = \int d^3x \left(\phi_1 \dot{\phi}_2 - \phi_2 \dot{\phi}_1\right) \ . \tag{8.27}$$

We can check that it is conserved by calculating

$$\frac{dQ}{dt} = \int d^3x \left(\dot{\phi}_1 \dot{\phi}_2 + \phi_1 \ddot{\phi}_2 - \dot{\phi}_2 \dot{\phi}_1 - \phi_2 \ddot{\phi}_1\right) \ . \tag{8.28}$$

From the equations of motion we have that

$$\ddot{\phi}_i = \nabla^2 \phi_i - m^2 \phi_i \ . \tag{8.29}$$

This means that

$$\frac{dQ}{dt} = \int d^3x \left[\phi_1 \left(\nabla^2 \phi_2 - m^2 \phi_2\right) - \phi_2 \left(\nabla^2 \phi_1 - m^2 \phi_1\right)\right] \ . \tag{8.30}$$

The m^2 terms cancel immediately. The others also cancel when we integrate both by parts, taking the fields to go to zero at spatial infinity

$$\frac{dQ}{dt} = \int d^3 \left(\nabla \phi_1 \cdot \nabla \phi_2 - \nabla \phi_1 \cdot \nabla \phi_2\right) = 0 \ . \tag{8.31}$$

This confirms that Q is time independent.

Exercise 8.6 Begin with the Lagrangian density for a complex scalar field

$$\mathcal{L} = (\partial_\mu \phi)\partial^\mu \phi^\dagger - m^2|\phi|^2, \tag{8.32}$$

where $|\phi|^2 = \phi^\dagger \phi$. Find the equations of motion for ϕ and $\phi^\dagger$ (treating them as independent fields). Show that

$$\phi = e^{-i\theta}\phi' \ , \quad \phi^\dagger = e^{i\theta}\phi'^\dagger \ , \tag{8.33}$$

is a symmetry of the Lagrangian for any continuously varying θ. Use Noether's theorem to find the conserved current

$$J^\mu = i\left(\phi^\dagger \partial^\mu \phi - \phi \partial^\mu \phi^\dagger\right) \ . \tag{8.34}$$

Show that this current is indeed conserved for fields that satisfy the equations of motion.

You might wonder why the above exercise treats ϕ and $\phi^\dagger$ is independent fields. If we consider that ϕ is made up of two real fields

$$\phi = \phi_1 + i\phi_2 \ , \tag{8.35}$$

then it is clear that we can treat both ϕ_1 and ϕ_2 as independent fields which then define ϕ and $\phi^\dagger$. However, we can simply write

$$\phi_1 = \frac{1}{2}\left(\phi + \phi^\dagger\right) \ , \quad \phi_2 = -\frac{i}{2}\left(\phi - \phi^\dagger\right) \ , \tag{8.36}$$

to see that we can simply think of ϕ and $\phi^\dagger$ as a change of basis from ϕ_1 and ϕ_2. This way of thinking leads to the correct equations for the ϕ and $\phi^\dagger$ in the simplest way and is the typical method for dealing with all complex fields.

We can also use this to obtain the plane wave expansion of a complex scalar field

$$\begin{aligned}\phi =& \phi_1 + i\phi_2 = \int d^3p \left(a_1(\vec{p})e^{-ix^\mu p_\mu} + a_1^\dagger(\vec{p})e^{ix^\mu p_\mu}\right) \\ &+ i\int d^3p \left(a_2(\vec{p})e^{-ix^\mu p_\mu} + a_2^\dagger(\vec{p})e^{ix^\mu p_\mu}\right) \\ =& \int d^3p \left[\left(a_1(\vec{p}) + ia_2(\vec{p})\right)e^{-ix^\mu p_\mu} + \left(a_1^\dagger(\vec{p}) + ia_2^\dagger(\vec{p})\right)e^{ix^\mu p_\mu}\right] \ . \end{aligned} \tag{8.37}$$

We define the two linearly independent operators combinations of the annihilation operators

$$a(\vec{p}) = a_1(\vec{p}) + ia_2(\vec{p}) \ , \quad b(\vec{p}) = a_1(\vec{p}) - ia_2(\vec{p}) \ . \tag{8.38}$$

This means that

$$\phi = \int d^3p \left(a(\vec{p}) e^{-ix^\mu p_\mu} + b^\dagger(\vec{p}) e^{ix^\mu p_\mu} \right) \ . \tag{8.39}$$

In order to understand what these operators do, we remember that in quantum mechanics we often consider processes that start in some initial state $|\text{initial}\rangle$ and end in some final state $|\text{final}\rangle$ by

$$\langle \text{final} | \text{initial} \rangle \ . \tag{8.40}$$

This means that if we are interested in understanding what the ϕ operator does to a system

$$\phi | \text{final} \rangle \ , \tag{8.41}$$

we should consider the conjugate state

$$\langle \text{final} | \phi^\dagger \ , \tag{8.42}$$

where

$$\phi^\dagger = \int d^3p \left(b(\vec{p}) e^{-ix^\mu p_\mu} + a^\dagger(\vec{p}) e^{ix^\mu p_\mu} \right) \ . \tag{8.43}$$

We see that $\phi^\dagger$ creates a particle with $a^\dagger$ or annihilates an antiparticle with b. Similarly, ϕ can create an antiparticle with $b^\dagger$ or annihilate a particle with a.

Now that we understand something of the current given in Eq. (8.34) we might as "What is the conserved quantity associated with this current?" We determine this by expressing the conserved charge

$$Q = \int d^3x \, J^0 = i \int d^3x \left(\phi^\dagger \partial_t \phi - \phi \partial_t \phi^\dagger \right) \ , \tag{8.44}$$

in terms of the creation and annihilation operators that make up ϕ (8.39). To find Q we first note that

$$\partial_t \phi = -i \int d^3p \, \omega_p \left(a(\vec{p}) e^{-ix^\mu p_\mu} - b^\dagger(\vec{p}) e^{ix^\mu p_\mu} \right) \ , \tag{8.45}$$

where ω_p is defined in Eq. (8.9). With this in hand we find

$$\begin{aligned} Q = & \, i \int d^3x \left[\int d^3k \left(b(\vec{k}) e^{-ix^\mu k_\mu} + a^\dagger(\vec{k}) e^{ix^\mu k_\mu} \right) (-i) \int d^3p \, \omega_p \left(a(\vec{p}) e^{-ix^\mu p_\mu} - b^\dagger(\vec{p}) e^{ix^\mu p_\mu} \right) \right. \\ & \left. - \int d^3k \left(a(\vec{k}) e^{-ix^\mu k_\mu} + b^\dagger(\vec{k}) e^{ix^\mu k_\mu} \right) (-i) \int d^3p \, \omega_p \left(b(\vec{p}) e^{-ix^\mu p_\mu} - a^\dagger(\vec{p}) e^{ix^\mu p_\mu} \right) \right] \\ = & \int d^3k \, d^3p \, \omega_p \, d^3x \left\{ e^{-ix^\mu (k_\mu + p_\mu)} \left[b(\vec{k}) a(\vec{p}) - a(\vec{k}) b(\vec{p}) \right) \right] - e^{-ix^\mu (k_\mu - p_\mu)} \left[b(\vec{k}) b^\dagger(\vec{p}) - a(\vec{k}) a^\dagger(\vec{p}) \right) \right] \\ & + e^{ix^\mu (k_\mu - p_\mu)} \left[a^\dagger(\vec{k}) a(\vec{p}) - b^\dagger(\vec{k}) b(\vec{p}) \right) \right] - e^{ix^\mu (k_\mu + p_\mu)} \left[a^\dagger(\vec{k}) b^\dagger(\vec{p}) - b^\dagger(\vec{k}) a^\dagger(\vec{p}) \right) \right] \Big\} \end{aligned} \tag{8.46}$$

We are now ready to use the delta function identity

$$\int_{-\infty}^{\infty} da\, e^{\pm ia(b-b')} = 2\pi\delta(b-b') \ . \tag{8.47}$$

Exercise 8.7

Use the identity in (8.47) to show that

$$Q = 2(2\pi)^3 \int d^3p\, \omega_p \left[a^\dagger(\vec{p})a(\vec{p}) - b^\dagger(\vec{p})b(\vec{p})\right] \ . \tag{8.48}$$

The factors of ω_p and 2π in (8.48) are not important, they can be absorbed into the definitions of $a(\vec{p})$ and $b(\vec{p})$. Essentially, we have obtained number operators

$$N_{\text{particle}} = 2(2\pi)^3 \int d^3p\, \omega_p\, a^\dagger(\vec{p})a(\vec{p}) = \int d^3p\, n_{\text{particle}} \ , \tag{8.49}$$

$$N_{\text{antiparticle}} = 2(2\pi)^3 \int d^3p\, \omega_p\, b^\dagger(\vec{p})b(\vec{p}) = \int d^3p\, n_{\text{antiparticle}} \ , \tag{8.50}$$

which, similar to the number operator that arises in the simple harmonic oscillator (33.96), count up the number from particles and antiparticles, respectively. The difference of these quantities is usually called particle number. Essentially it counts the number of particles in a given field configuration and subtracts the number of antiparticles. An example you may have heard of is baryon number. In nuclear processes baryon number is conserved. If the initial state has some number of baryons N_B and antibaryons $N_{\overline{B}}$ then the baryon number

$$B = N_B - N_{\overline{B}} \ , \tag{8.51}$$

is conserved during the process, even if N_B and $N_{\overline{B}}$ individually change.

Example

Suppose we collider a proton p and an antiproton $\overline{p}$. The initial baryon number is

$$1 - 1 = 0 \ . \tag{8.52}$$

The resulting collision may produce many types of particles. But, as long as baryon number is a good symmetry, then we know that the number of baryons produced must equal the number of antibaryons produced so that

$$N_B - N_{\overline{B}} = 0 \ . \tag{8.53}$$

Note that each of these changes the particle number in the same way. If we again use baryon number as an example, then $\phi^\dagger$ increases the baryon number by one unit. This can occur either by creating a baryon or by destroying an antibaryon. On the other hand, ϕ reduces baryon number by one unit, either by destroying a baryon or by creating an antibaryon.

We can use the stress-energy tensor, as defined in (32.107), to further explore the meaning of these creation and annihilation operators.

Exercise 8.8 Show, using the definition in (32.107), that the stress-energy tensor for a complex scalar field is

$$T^{\mu}{}_{\nu} = \left(\partial^{\mu}\phi^{\dagger}\right)\partial_{\nu}\phi + \left(\partial^{\mu}\phi\right)\partial_{\nu}\phi^{\dagger} - \delta^{\mu}_{\nu}\left[\left(\partial^{\alpha}\phi^{\dagger}\right)\partial_{\alpha}\phi - m^2|\phi|^2\right] \ . \tag{8.54}$$

Using the above result and the definition of the linear momentum carried by the field in (32.109) we find that

$$P^i = \int d^3\, T^{0i} = -\delta^{ij}\int d^3x\left[(\partial_t\phi^{\dagger})\partial_j\phi + (\partial_t\phi)\partial_j\phi^{\dagger}\right] \ . \tag{8.55}$$

Exercise 8.9 Show that the linear momentum of the complex scalar field is given by

$$\begin{aligned} P^i &= 2(2\pi)^3\int d^3p\,\omega_p\,p^i\left[a^{\dagger}(\vec{p})a(\vec{p}) + b^{\dagger}(\vec{p})b(\vec{p})\right] \\ &= \int d^3p\,p^i\left(n_{\text{particle}} + n_{\text{antiparticle}}\right) \ . \end{aligned} \tag{8.56}$$

How do we interpret the above result for the momentum? It is simply that we are add up the three-momentum associated with each possible momentum mode. This summing over all momentum is weighted by the number operators. So, very roughly, we add up the momentum associated with each particle and the momentum associated with each antiparticle. The sum of all these momenta is the total momenta.

Finally, we can consider the total energy as defined in (32.108). We find that

$$E = \int d^3x\, T^{00} = \int d^3x\left[(\partial_t\phi^{\dagger})\partial_t\phi + \delta^{ij}(\partial_i\phi^{\dagger})\partial_j\phi + m^2\phi^{\dagger}\phi\right] \ . \tag{8.57}$$

Exercise 8.10 Using the definition for the energy carried by the field in (32.108) show that

$$\begin{aligned} E &= 2(2\pi)^3\int d^3p\,w_p^2\left[a^{\dagger}(\vec{p})a(\vec{p}) + b^{\dagger}(\vec{p})b(\vec{p})\right] \\ &= \int d^3p\,\omega_p\left(n_{\text{particle}} + n_{\text{antiparticle}}\right) \ . \end{aligned} \tag{8.58}$$

Similar to what we found with the linear momentum we see that the total energy is the sum of the energies associated with each particle mode and each antiparticle mode.

Section 9

The Dirac Equation

We motivated the form of the Klein-Gordon equation by substituting quantum operators $\hat{P}_\mu \to i\partial_\mu$ into the relativistic energy relation

$$p_\mu p^\mu = m^2 \ , \tag{9.1}$$

to produce

$$\begin{aligned} -\partial_\mu \partial^\mu \phi =& m^2 \phi \\ -\eta^{\mu\nu} \partial_\mu \partial_\nu \phi =& m^2 \phi \ . \end{aligned} \tag{9.2}$$

This led to both positive and negative frequency solutions, which forces us to give up the interpretation as a single-particle wave equation. Dirac saw the appearance of positive and negative frequencies as a consequence of the Klein-Gordon equation being second order in time derivatives. This motivated him to look for a relativistic equation with only one time derivative [29].

If we hope to have only one derivative term, then we must include some other Lorentz four-vector to contract with it so that the equation is Lorentz-invariant. Consider

$$i\gamma^\mu \partial_\mu \psi = m\psi \ , \tag{9.3}$$

where γ^μ is left unspecified for the moment. Using this equation twice we find

$$i\gamma^\nu \partial_\nu i\gamma^\mu \partial_\mu \psi = i\gamma^\nu \partial_\nu m\psi = m^2 \psi \ . \tag{9.4}$$

The right-hand side of this equation looks like part of the Klein-Gordon equation.

Exercise 9.1 Show that the left-hand side of Eq. (9.4) can be written as

$$i\gamma^\nu \partial_\nu i\gamma^\mu \partial_\mu \psi = -\frac{1}{2} \left(\gamma^\mu \gamma^\nu + \gamma^\nu \gamma^\mu \right) \partial_\mu \partial_\nu \psi \ . \tag{9.5}$$

This symmetric form of the γ^μs is important, because the pair partial derivatives themselves are symmetric in their indices. Conclude that Eq. (9.4) agrees with the Klein-Gordon equation if

$$\gamma^\mu \gamma^\nu + \gamma^\nu \gamma^\mu = 2\eta^{\mu\nu} \ . \tag{9.6}$$

We know, from the association of quantum operators to the relativistic energy relation, that all relativistic fields should satisfy the Klein-Gordon equation. The above exercise shows that a field that satisfies the Dirac

equation (9.3) also satisfies the Klein-Gordon equation as long as Eq. (9.6) holds. This implies that the four objects γ^0, γ^1, γ^2, and γ^3 satisfy

$$\gamma^0\gamma^0 = \mathbb{I}\ , \quad \gamma^1\gamma^1 = -\mathbb{I}\ , \quad \gamma^2\gamma^2 = -\mathbb{I}\ , \quad \gamma^3\gamma^3 = -\mathbb{I}\ , \tag{9.7}$$

and

$$\gamma^\mu\gamma^\nu = -\gamma^\nu\gamma^\mu \text{ for } \mu \neq \nu\ . \tag{9.8}$$

Exercise 9.2 Show that the requirements in Eqs. (9.7) and (9.8) are satisfied by the four-by-four matrices made of two-by-two blocks

$$\gamma^0 = \begin{pmatrix} \mathbf{0}_2 & \mathbb{I}_2 \\ \mathbb{I}_2 & \mathbf{0}_2 \end{pmatrix}\ , \qquad \gamma^i = \begin{pmatrix} \mathbf{0}_2 & \sigma_i \\ -\sigma_i & \mathbf{0}_2 \end{pmatrix}\ , \tag{9.9}$$

where $\mathbf{0}_2$ is the two-by-two matrix of zeros, $\mathbb{I}_2$ is the identity matrix in two dimensions, and the σ_i are the Pauli spin matrices defined in Eq. (38.39). This is not the only set of matrices which satisfy these requirements. This particular choice is called the Weyl basis.

Though we do not prove it here, there are no matrices of lower dimension than four which satisfy Eqs. (9.7) and (9.8). This means that the minimal objects ψ^α that appear in the Dirac equation are four-dimensional vectors of fields

$$\psi^A(x) = \begin{pmatrix} \psi^1(x) \\ \psi^2(x) \\ \psi^3(x) \\ \psi^4(x) \end{pmatrix}\ , \tag{9.10}$$

where we have used capital Latin letters to denote the four-dimensional Dirac spinor index. It is important to recognize that this four dimensional object is **not** a spacetime four-vector, which we typically index with Greek letters. This form of the field looks like objects that might be familiar from single particle quantum mechanics. When describing a particle that moves through physical space and has nonzero spin the full Hilbert space has the form

$$\mathcal{H}_{\text{space}} \otimes \mathcal{H}_{\text{spin}}\ . \tag{9.11}$$

The field in (9.10) looks like a quantum state that has been expressed in position space and has a four-dimensional Hilbert space to describe the spin of the particle. The following investigation of these objects owes much to the discussions in [10, 14].

Let us assume that the four-dimensional space is associated with spin degrees of freedom. But what representation(s) of spin does it describe? We address this question below, but first write the Dirac equation carefully

$$\left[i\left(\gamma^\mu\right)^A{}_B\,\partial_\mu - m\mathbb{I}^A{}_B\right]\psi^B = 0\ . \tag{9.12}$$

Here we have made explicit that spacetime index (μ) is completely distinct from spinor indices A and B. This is important to remember, especially considering that the spinor indices are rarely written. The γ^μ is a four-vector of operators that act on a four-dimensional, quantum mechanical Hilbert space, the space of Dirac spinors. Typically, these spinor indices are suppressed as in the equations above.

Example

We found in the previous section that the Klein-Gordon equation admits plane-wave solutions. The Dirac equation does too. Consider the Dirac spinor

$$\psi^A(x) = u^A e^{-ix^\mu p_\mu} \ , \tag{9.13}$$

where u^A is a constant spinor, independent of x^μ. In this formula we have suggestively labeled the four-vector contracted with the position to look like the four-momentum. However, it may be best to pretend that this is just some as yet undetermined four vector.

Putting this into the Dirac equation (and suppressing spinor indices) we find

$$(i(-i)\gamma^\mu p_\mu - m\mathbb{I})\, u e^{-ix^\mu p_\mu} = 0 \ , \tag{9.14}$$

which leads to the constraint

$$\left(\not{p} - m\mathbb{I}\right) u = 0 \ . \tag{9.15}$$

Here we have used the notation $\gamma^\mu V_\mu \equiv \not{V}$. In the Weyl basis, and writing the constant spinor as

$$u = \begin{pmatrix} \xi \\ \zeta \end{pmatrix} \ , \tag{9.16}$$

where each of ξ and ζ are two dimensional column vectors, we find that the constraint equation can be written

$$\begin{pmatrix} -m & p_0\mathbb{I}_2 + p^i\sigma_i \\ p_0\mathbb{I}_2 - p^i\sigma_i & -m \end{pmatrix} \begin{pmatrix} \xi \\ \zeta \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \ . \tag{9.17}$$

This can also be written as

$$\left(p_0\mathbb{I}_2 - p^i\sigma_i\right)\xi = m\zeta \ , \qquad \left(p_0\mathbb{I}_2 + p^i\sigma_i\right)\zeta = m\xi \ . \tag{9.18}$$

This implies that

$$\left(p_0\mathbb{I}_2 + p^j\sigma_j\right)\left(p_0\mathbb{I}_2 - p^i\sigma_i\right)\xi = m\left(p_0\mathbb{I}_2 + p^j\sigma_j\right)\zeta = m^2\xi \ . \tag{9.19}$$

Similarly, we find that

$$\left(p_0\mathbb{I}_2 - p^j\sigma_j\right)\left(p_0\mathbb{I}_2 + p^i\sigma_i\right)\zeta = m^2\zeta \ . \tag{9.20}$$

Using the identity in Eq. (38.40) we also find that

$$\begin{aligned}\left(p_0\mathbb{I}_2 + p^j\sigma_j\right)\left(p_0\mathbb{I}_2 - p^i\sigma_i\right) =& p_0^2\mathbb{I}_2 + p_0p^i\sigma_i - p_0p^i\sigma_i - p^ip^j\left(\delta_{ij}\mathbb{I}_2 - i\varepsilon_{ijk}\sigma_k\right)\\ =& \mathbb{I}_2\left(p_0^2 - \vec{p}^2\right) \ . \end{aligned} \tag{9.21}$$

This means we can rewrite Eqs. (9.19) and (9.20) as

$$\left(p_0^2 - \vec{p}^2\right)\xi = m^2\xi \ , \quad \left(p_0^2 - \vec{p}^2\right)\zeta = m^2\zeta \ . \tag{9.22}$$

In other words, our plane wave form is only a solution if $E^2 - \vec{p}^2 = m^2$, where $p_0 = E$.

Exercise 9.3 Find the constraint on the parameters of the other plane wave solution to the Dirac equation

$$\psi(x) = ve^{ix^\mu p_\mu} \ , \tag{9.23}$$

by following the method of the above example.

Exercise 9.4 Show that the Hermitian matrices

$$p_0\mathbb{I}_2 + p^j\sigma_j \ , \quad p_0\mathbb{I}_2 - p^j\sigma_j \ , \tag{9.24}$$

both have eigenvalues

$$\lambda_\pm = p_0 \pm |\vec{p}| \ . \tag{9.25}$$

Argue that both of these eigenvalues are non-negative for all momenta for which $E^2 - \vec{p}^2 = m^2$.

Example Our analysis so far gives physical meaning to the coefficients (components of the four-momentum) that appear in the exponential of the plane waves. What about the two-dimensional spinors ξ and ζ? The exercise above implies that we can define the square roots of the matrices

$$p_0\mathbb{I}_2 + p^j\sigma_j \ , \quad p_0\mathbb{I}_2 - p^j\sigma_j \ . \tag{9.26}$$

This is done by changing to the basis in which the matrices are diagonal, then taking the square root of the diagonal matrices by simply taking the square root of the diagonal elements of the matrices. We then express these new matrices in the original basis. It is not particularly useful to write down the explicit form of these matrices, but they are perfectly well defined. We write them as

$$\sqrt{p_0\mathbb{I}_2 + p^j\sigma_j} \ , \quad \sqrt{p_0\mathbb{I}_2 - p^j\sigma_j} \ . \tag{9.27}$$

Note that

$$\sqrt{p_0\mathbb{I}_2 + p^j\sigma_j}\sqrt{p_0\mathbb{I}_2 - p^j\sigma_j} = \sqrt{m^2\mathbb{I}} = m\mathbb{I} \ . \tag{9.28}$$

Similarly, one finds that

$$\left(p_0\mathbb{I}_2 + p^j\sigma_j\right)\sqrt{p_0\mathbb{I}_2 - p^j\sigma_j} = m\sqrt{p_0\mathbb{I}_2 + p^j\sigma_j} \ , \tag{9.29}$$

and

$$\left(p_0\mathbb{I}_2 - p^j\sigma_j\right)\sqrt{p_0\mathbb{I}_2 + p^j\sigma_j} = m\sqrt{p_0\mathbb{I}_2 - p^j\sigma_j} \ . \tag{9.30}$$

Consider writing

$$\xi = \sqrt{p_0\mathbb{I}_2 + p^j\sigma_j}\chi \ , \tag{9.31}$$

and inserting this into Eq. (9.19). One quickly finds that

$$m\sqrt{p_0\mathbb{I}_2 - p^j\sigma_j}\chi = m\zeta \ . \tag{9.32}$$

This is also consistent with Eq. (9.20). Importantly, the two-dimensional spinor χ is the same for both ξ and ζ. It is also dimensionless and purely numerical. All dependence on the momentum has been separated out. The full plane wave spinor is then

$$u(\vec{p}) = \begin{pmatrix} \sqrt{p_0\mathbb{I}_2 + p^j\sigma_j}\chi \\ \sqrt{p_0\mathbb{I}_2 - p^j\sigma_j}\chi \end{pmatrix} \ , \tag{9.33}$$

where χ is a two-dimensional complex vector, which we take to be normalized

$$\chi^\dagger\chi = 1 \ . \tag{9.34}$$

This spinor is defined completely by the three-momentum $\vec{p}$ because of the condition $p_0 = \omega_p = \sqrt{m^2 + \vec{p}\cdot\vec{p}}$.

This spinor χ can be expressed in terms of some basis. Sometimes spin along the z-direction is chosen. The essential point is that the spinor has two complex dimensions. We might index it as

$$\chi^s \ , \tag{9.35}$$

where $s \in \{1, 2\}$. This means that the plane wave in Eq. (9.33) has two "degrees of freedom." We can write four-dimensional spinor coefficient of the plane wave as $u^s(p)$ where the index refers to the components of χ^s only. Note that the normalization condition is then

$$\chi_r^\dagger\chi^s = \delta_r^s \ . \tag{9.36}$$

Exercise 9.5 Show that for the other plane wave solution to the Dirac equation

$$\psi(x) = v^s(\vec{p}) e^{ix^\mu p_\mu} \ , \tag{9.37}$$

that

$$v^s(\vec{p}) = \begin{pmatrix} \sqrt{p_0 \mathbb{I}_2 + p^j \sigma_j}\Upsilon^s \\ -\sqrt{p_0 \mathbb{I}_2 - p^j \sigma_j}\Upsilon^s \end{pmatrix} \ , \tag{9.38}$$

where Υ^s is a constant, two-dimensional spinor. This shows that this plane wave accounts for another two degrees of freedom in the Dirac field.

Exercise 9.6 Show that the magnitude of these spinors are given by

$$u_r^\dagger(\vec{p}) u^s(\vec{p}) = 2\omega_p \delta_r^s \ , \quad v_r^\dagger(\vec{p}) v^s(\vec{p}) = 2\omega_p \delta_r^s \ . \tag{9.39}$$

Then, show that the spinors also satisfy

$$u_r^\dagger(\vec{p}) v^s(-\vec{p}) = 0 \ , \quad v_r^\dagger(\vec{p}) u^s(-\vec{p}) = 0 \ . \tag{9.40}$$

These analyses of the plane wave solutions motivate writing the four-dimensional Dirac spinor in terms of two two-dimensional Weyl [30] spinors:

$$\psi = \begin{pmatrix} \psi_L(x) \\ \psi_R(x) \end{pmatrix} \ . \tag{9.41}$$

Similar to the plane-wave example above, we find two equations

$$\left(E\mathbb{I}_2 + p^i \sigma_i\right) \psi_L = m\psi_R \ , \quad \left(E\mathbb{I}_2 - p^i \sigma_i\right) \psi_R = m\psi_L \ . \tag{9.42}$$

Note that in the $m = 0$ limit these two equations decouple and we have two independent, two-dimensional spinors. It turns out that ψ_L is in the $(\frac{1}{2}, 0)$ representation of the Lorentz group, which is a spin-1/2 representation. This field is said to be left-handed. The ψ_R field is also spin-1/2, but in the $(0, \frac{1}{2})$ representation of the Lorentz group. Some theories, like QED, have equal interactions for ψ_L and ψ_R. Theories that have different couplings for these fields are said to be **chiral**. We also see from this that a Dirac spinor is in the $\mathbf{\frac{1}{2}} \oplus \mathbf{\frac{1}{2}}$ representation spin.

It is useful to be able to write the left- and right-handed parts of the spinor within the four-dimensional space. To aid in this we make the definition

$$\gamma_5 = \gamma^5 \equiv i\gamma^0\gamma^1\gamma^2\gamma^3 \ . \tag{9.43}$$

One sees the 5 here (which is a label, not an index value) sometimes written as a subscript and sometimes as a superscript. The location has no meaning. This object is used to define projectors into the left-handed and

right-handed subspaces.

Definition The left- and right-handed projections of a Dirac spinor are defined by

$$P_L = \frac{1}{2}\left(\mathbb{I}_4 - \gamma_5\right) \ , \qquad P_R = \frac{1}{2}\left(\mathbb{I}_4 + \gamma_5\right) \ . \tag{9.44}$$

Exercise 9.7 Show that in the Weyl basis that

$$\gamma_5 = \begin{pmatrix} -\mathbb{I}_2 & \mathbf{0}_2 \\ \mathbf{0}_2 & \mathbb{I}_2 \end{pmatrix} \ . \tag{9.45}$$

Then evaluate the action of P_L and P_R on the Dirac spinor

$$\psi = \begin{pmatrix} \psi_L(x) \\ \psi_R(x) \end{pmatrix} \ . \tag{9.46}$$

Subsection 9.1

The Dirac Lagrangian

Now that we have an understanding of the Dirac equation we would like to find a Lagrangian that leads to it. As we do this it is useful to remind ourselves what we require of a Lagrangian. First, it should be either real or Hermitian (which is the operator equivalent of real). Second, it should be a Lorentz scalar. Third, it should lead to the correct equations of motion.

Let's start with our third requirement. It might motivate the (incorrect) guess

$$\mathcal{L}_{\text{incorrect}}(\psi, \psi^\dagger) = \psi^\dagger \left(i\gamma^\mu \partial_\mu - m\mathbb{I}\right)\psi \ . \tag{9.47}$$

Note that this Lagrangian is taken to be a function of independent variables ψ and $\psi^\dagger$ similar to how the Klein-Gordon Lagrangian of a complex scalar was a function of ϕ and $\phi^\dagger$. Then one of the Euler-Lagrange equations is

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \psi^\dagger} &= \partial_\nu \frac{\partial \mathcal{L}}{\partial(\partial_\nu \psi^\dagger)} \\ \left(i\gamma^\mu \partial_\mu - m\mathbb{I}\right)\psi &= 0 \ , \end{aligned} \tag{9.48}$$

which is the expected Dirac equation. Note if we take the Hermitian conjugate of this equation we get

$$-i(\partial_\mu \psi^\dagger)\left(\gamma^\mu\right)^\dagger - m\psi^\dagger = 0 \ . \tag{9.49}$$

But, remembering that the Lagrangian is integrated over all space, $S = \int d^4x\, \mathcal{L}$, we can integrate by parts to rewrite

$$\mathcal{L}_{\text{incorrect}}(\psi, \psi^\dagger) = -\partial_\mu(\psi^\dagger) i\gamma^\mu \psi - m\psi^\dagger\psi \ . \tag{9.50}$$

This leads to the equation of motion

$$-\partial_\mu(\psi^\dagger)i\gamma^\mu - m\psi^\dagger = 0 \ , \tag{9.51}$$

which is only the same as Eq. (9.49) if the γ^μ are Hermitian.

Exercise 9.8 Show that in the Weyl basis that

$$(\gamma^\mu)^\dagger = \gamma^0\gamma^\mu\gamma^0 \ . \tag{9.52}$$

Show that this implies that for $\mu \neq 0$ that the gamma matrices are anti-Hermitian.

The exercise shows that Eq. (9.51) is the wrong equation. In essence, we have found that our guessed Lagrangian is not Hermitian.

Exercise 9.9 Show that

$$\mathcal{L}_{\text{incorrect}}(\psi, \psi^\dagger) = \psi^\dagger \left(i\gamma^\mu\partial_\mu - m\mathbb{I}\right)\psi \ , \tag{9.53}$$

is not Hermitian. You may need to integrate by parts to fully compare.

Not being Hermitian is a problem. Another problem is that this Lagrangian is not a Lorentz invariant. First, remember from Eq. (7.10) that a general Lorentz transformation can be written in the form

$$\Lambda = \exp\left[-(i\theta_i + \beta_i)J_i^+ - (i\theta_i - \beta_i)J_i^-\right] \ , \tag{9.54}$$

where the $J_i^\pm$ are the generators of the two $\mathfrak{su}(2)$ Lie algebras that make up the Lorentz algebra. For the left-handed representation $(\frac{1}{2}, 0)$ we have $J_i^+ = \frac{1}{2}\sigma_i$ and $J_i^- = 0$. Similarly, for the right-handed representation $(0, \frac{1}{2},)$ we have $J_i^- = \frac{1}{2}\sigma_i$ and $J_i^+ = 0$. This means that under Lorentz transformations

$$\psi_L \to e^{\frac{1}{2}(i\theta_i+\beta_i)\sigma_i}\psi_L \ , \qquad \psi_R \to e^{\frac{1}{2}(i\theta_i-\beta_i)\sigma_i}\psi_R \ . \tag{9.55}$$

Now, let us consider the mass term of $\mathcal{L}_{\text{incorrect}}$:

$$m\psi^\dagger\psi = m\left(\psi_L^\dagger, \psi_R^\dagger\right)\begin{pmatrix}\psi_L \\ \psi_R\end{pmatrix} = m\left(\psi_L^\dagger\psi_L + \psi_R^\dagger\psi_R\right) \ . \tag{9.56}$$

Is this invariant under Lorentz transformations?

Exercise 9.10 Show, using Eqs. (38.40) and (38.49) that

$$e^{\frac{1}{2}(-i\theta_i\pm\beta_i)\sigma_i}e^{\frac{1}{2}(i\theta_i\pm\beta_i)\sigma_i} = \mathbb{I}_1\cos\left(\sqrt{\theta^2+\beta^2}\right) \pm i\beta_i\sigma_i\sin\cos\left(\sqrt{\theta^2+\beta^2}\right) \pm 2i\frac{\theta_i\beta_j\varepsilon_{ijk}}{\theta^2+\beta^2}\sigma_k\sin^2\left(\frac{\sqrt{\theta^2+\beta^2}}{2}\right) , \tag{9.57}$$

where $\theta_i\theta_i \equiv \theta^2$ and similar for β^2.

The above exercise shows that

$$m\left(\psi_L^\dagger\psi_L + \psi_R^\dagger\psi_R\right) \not\to m\left(\psi_L^\dagger\psi_L + \psi_R^\dagger\psi_R\right) , \tag{9.58}$$

under general Lorentz transformations. In other words, this not a Lorentz invariant quantity.

So, our guess at the correct Lagrangian fails. It only produces one of the correct equations of motion, it is not Hermitian, and it is not Lorentz invariant. It turns out that all of these problems can be overcome by using the so-called Dirac conjugate field

$$\overline{\psi} \equiv \psi^\dagger\gamma^0 = \left(\psi_R^\dagger, \psi_L^\dagger\right) . \tag{9.59}$$

Exercise 9.11 Show that $\overline{\psi}\psi$ is both Lorentz invariant and Hermitian.

It looks like the mass term is now sorted out. But what about the derivative term? We might hope it has the form

$$\overline{\psi}i\gamma^\mu\partial_\mu\psi . \tag{9.60}$$

It is quite simple to check that this is Hermitian.

Example We are interested in checking that the term in (9.60) is Hermitian. We proceed by first noting that

$$\overline{\psi}^\dagger = \left(\psi^\dagger\gamma^0\right)^\dagger = \left(\gamma^0\right)^\dagger\psi = \gamma^0\psi . \tag{9.61}$$

We then find

$$\begin{aligned}\left(\overline{\psi}i\gamma^\mu\partial_\mu\psi\right)^\dagger &= -i\partial_\mu\left(\psi^\dagger\right)\left(\gamma^\mu\right)^\dagger\gamma^0\psi \\ &= -i\partial_\mu\left(\psi^\dagger\right)\gamma^0\gamma^\mu\gamma^0\gamma^0\psi \\ &= -i\partial_\mu\left(\overline{\psi}\right)\gamma^\mu\mathbb{I}_4\psi \\ &\to\overline{\psi}i\gamma^\mu\partial_\mu\psi .\end{aligned} \tag{9.62}$$

In the last line we have integrated by parts. This shows that this term

is Hermitian (up to vanishing boundary terms).

What about Lorentz invariance? We do not take the time to prove this, but essentially, the γ^μ are a four-vector operator in the spinor space. What this means is that if we take

$$O(\Lambda)^A{}_B \ , \tag{9.63}$$

to be the spinor space operator that represents the spacetime Lorentz transformation $\Lambda^\mu{}_\nu$ then the gamma matrices satisfy

$$O^{-1}(\Lambda)^A{}_B \, (\gamma^\mu)^B{}_C \, O(\Lambda)^C{}_D = \Lambda^\mu{}_\nu \, (\gamma^\nu)^A{}_D \ . \tag{9.64}$$

In other words, enacting the Lorentz transformation in the spinor space is equivalent to enacting it in four dimensional spacetime. Or, the four vector index on the gamma matrices is not just a nice way to label four matrices, the list of matrices really do transform like a four vector. Let us see how these operators apply to the mass term $m\overline{\psi}\psi$ whose Lorentz invariance we have already established. We find

$$\psi^\dagger \gamma^0 \psi \to \psi^\dagger O^\dagger \gamma^0 O \psi = \psi^\dagger \gamma^0 \psi \ , \tag{9.65}$$

where in the last equality we have used the fact that this term is Lorentz invariant. This equality indicates that

$$O^\dagger \gamma^0 = \gamma^0 O^{-1} \ . \tag{9.66}$$

This also makes clear that $O^\dagger \neq O^{-1}$, or in other words that this finite dimensional representation of the Lorentz group is *not unitary*. As discussed in earlier sections, this follows from the boost generators not being Hermitian operators. The power and point of the Dirac conjugate field is that even through $O(\Lambda)$ is not unitary it is true that

$$\overline{\psi} \to \overline{\psi} O^{-1} \ . \tag{9.67}$$

Consequently, one finds that under a Lorentz transformation

$$\begin{aligned} \overline{\psi} i\gamma^\mu \partial_\mu \psi \to & \overline{\psi} O^{-1} i\gamma^\mu \Lambda_\mu{}^\nu \partial_\nu O\psi = \overline{\psi} i O^{-1} \gamma^\mu O \Lambda_\mu{}^\nu \partial_\nu \psi \\ & = \overline{\psi} i \Lambda^\mu{}_\alpha \gamma^\alpha \Lambda_\mu{}^\nu \partial_\nu \psi = \overline{\psi} i \gamma^\alpha \delta^\nu_\alpha \partial_\nu \psi \\ & = \overline{\psi} i \gamma^\mu \partial_\mu \psi \ . \end{aligned} \tag{9.68}$$

Exercise 9.12 Fill in all the steps of the above demonstration that $\overline{\psi} i\gamma^\mu \partial_\mu \psi$ is Lorentz invariant. You may find it useful to write out the spinor indices explicitly. Note also that the spinor and spacetime indices correspond to different

vector spaces. The operators in one space are simply extended to the identity in the other.

The outcome of all this is that we now understand that

$$\mathcal{L} = \overline{\psi}\,(i\gamma^\mu\partial_\mu - m)\,\psi\ , \tag{9.69}$$

is a Hermitian, Lorentz invariant, Lagrangian that produces the Dirac equation. Therefore, we take it to be the Dirac Lagrangian. Note that there is an identity matrix in the spinor space associated with the mass term. This identity matrix is usually left as implicit in textbooks and other writing.

Now that we have this Lagrangian we should see what symmetries, if any, it has. The Dirac field ψ is typically complex so we consider a rephasing of the fields like we did with the complex Klein-Gordon fields

$$\psi = e^{-i\theta}\psi'\ , \qquad \overline{\psi} = e^{i\theta}\overline{\psi}'\ , \tag{9.70}$$

for any constant real number θ. Being constant, this phase can simply be taken through derivatives, and so we find that the Lagrangian is written as

$$\mathcal{L} = \overline{\psi}'\,(i\gamma^\mu\partial_\mu - m)\,\psi'\ . \tag{9.71}$$

This has the same form as the original Lagrangian and leads to the same equations of motion, so it is a symmetry of the theory. As this family of transformations depends on a continuous parameter we expect a conserved quantity due to Noether's theorem. We find the Noether current is

$$\begin{aligned} J^\mu =& \frac{\partial\mathcal{L}}{\partial(\partial_\mu\psi)}\left.\frac{d\psi}{d\theta}\right|_{\theta=0} + \left.\frac{d\overline{\psi}}{d\theta}\right|_{\theta=0}\frac{\partial\mathcal{L}}{\partial(\partial_\mu\overline{\psi})} \\ =& i\overline{\psi}\gamma^\mu(-i)\psi + 0 \\ =& \overline{\psi}\gamma^\mu\psi\ , \end{aligned} \tag{9.72}$$

which, like for complex Klein-Gordon fields, leads to a conserved particle number. Notice that in this calculation one must be careful to link up the spinor indices correctly in the first line.

Exercise 9.13 Show that the current $J^\mu = \overline{\psi}\gamma^\mu\psi$ is conserved ($\partial_\mu J^\mu = 0$) for fields that satisfy the Dirac equation.

Exercise 9.14 Suppose we were to suspect that the Dirac Lagrangian (9.69) is symmetric under the rephasing

$$\psi = e^{-i\theta\gamma_5}\psi'\ , \qquad \overline{\psi} = \overline{\psi}'e^{-i\theta\gamma_5}\ . \tag{9.73}$$

First, check that the above relations are consistent, that is, given the first derive the second. You should begin by using

$$\gamma^\mu \gamma_5 = -\gamma_5 \gamma^\mu \tag{9.74}$$

to prove that

$$e^{-i\theta\gamma_5}\gamma^\mu = \gamma^\mu e^{i\theta\gamma_5} \ . \tag{9.75}$$

Now, check to see if this type of rephasing given in Eq. (9.73) is a symmetry of the Lagrangian.

No matter what you discover, press on and calculate the current associated with transformation. Show, using the usual Noether's theorem methods that the current is

$$\overline{\psi}\gamma^\mu\gamma_5\psi \ . \tag{9.76}$$

Check to see if this current is conserved for fields that satisfy the Dirac equation.

You should find that in general the current is not conserved, but that if $m = 0$ then it is. This is the so-called chiral-symmetry enjoyed by *massless* fermions.

The Dirac equation for massless fermions is

$$\mathcal{L} = i\overline{\psi}\gamma^\mu\partial_\mu\psi \ . \tag{9.77}$$

In the Weyl basis we can write this as

$$\mathcal{L} = i\psi_R^\dagger \left(\mathbb{I}_2\partial_t + \sigma_i\partial_i\right)\psi_R + i\psi_L^\dagger\left(\mathbb{I}_2\partial_t - \sigma_i\partial_i\right)\psi_L \ . \tag{9.78}$$

We see that the left-handed and right-handed fields have completely decoupled from each other. This shows that we can have a theory of only a massless left-handed fermion or only a massless right-handed fermion. For instance, one could describe a massless left-handed fermion by

$$\mathcal{L} = i\psi_L^\dagger\left(\mathbb{I}_2\partial_t - \sigma_i\partial_i\right)\psi_L \ , \tag{9.79}$$

without ever mentioning a right-handed field.

The Dirac fermion field can be expanded in plane waves, similar to what we found for Klein-Gordon fields

$$\psi^A(x) = \int d^3p \sum_{s=1}^{2}\left(u^{sA}\left(\vec{p}\right)a_s\left(\vec{p}\right)e^{-ix^\mu p_\mu} + v^{sA}\left(\vec{p}\right)b_s^\dagger\left(\vec{p}\right)e^{ix^\mu p_\mu}\right) \ . \tag{9.80}$$

Note that the coefficients of the plane waves include creation or annihilation operators and a spinor—u^s or v^s—that depends on the momentum only. Note also that there are four total sets of creation and annihilation

operators a_1, a_2, b_1, and b_2. These correspond to the four degrees of freedom that are described by a Dirac spinor. If we choose the two-dimensional spinors χ and Υ used in Eqs. (9.33) and (9.38) to correspond to spin along the z-direction (as is often the case) then the four degrees of freedom are associated with a spin-up electron, a spin-down electron, a spin-up positron, and a spin-down positron.

Often, the sum over the spins is left implied and the spinor index is not shown explicitly. This leads to

$$\psi(x) = \int d^3p \left(u^s(\vec{p})\, a_s(\vec{p})\, e^{-ix^\mu p_\mu} + v^s(\vec{p})\, b_s^\dagger(\vec{p})\, e^{ix^\mu p_\mu} \right) , \tag{9.81}$$

$$\overline{\psi}(x) = \int d^3p \left(\overline{v}_s(\vec{p})\, b_s(\vec{p})\, e^{-ix^\mu p_\mu} + \overline{u}_s(\vec{p})\, a_s^\dagger(\vec{p})\, e^{ix^\mu p_\mu} \right) , \tag{9.82}$$

where the spinors $\overline{u}$ and $\overline{v}$ are the Dirac conjugates of u and v. We are now ready to calculate the conserved quantity

$$\begin{aligned} Q &= \int d^3x\, J^0 = \int d^3x\, \overline{\psi}\gamma^0\psi = \int d^3x\, \psi^\dagger\psi \qquad (9.83)\\ &= \int d^3p\, d^3k \int d^3x \left[v_r^\dagger(\vec{p}) u^s(\vec{k}) b_r(\vec{p}) a_s(\vec{k}) e^{-ix^\mu(p_\mu+k_\mu)} + v_r^\dagger(\vec{p}) v^s(\vec{k}) b_r(\vec{p}) b_s^\dagger(\vec{k}) e^{-ix^\mu(p_\mu-k_\mu)} \right.\\ &\quad \left. + u_r^\dagger(\vec{p}) u^s(\vec{k}) a_r^\dagger(\vec{p}) a_s(\vec{k}) e^{ix^\mu(p_\mu-k_\mu)} + u_r^\dagger(\vec{p}) v^s(\vec{k}) a_r^\dagger(\vec{p}) b_s^\dagger(\vec{k}) e^{ix^\mu(p_\mu+k_\mu)} \right] . \end{aligned}$$

As in the scalar field case this can be greatly simplified by using the definition of the delta function in (8.47).

Exercise 9.15 Use the definition of the delta function as well as the spinor normalization and orthogonality conditions in (9.39) and (9.40) to show that

$$Q = 2(2\pi)^3 \int d^3p\, w_p \left[b_s(\vec{p}) b_s^\dagger(\vec{p}) + a_s^\dagger(\vec{p}) a_s(\vec{p}) \right] , \tag{9.84}$$

where there is an implied sum over the spin states.

The result in (9.84) is quite similar to what we found for the complex scalar field (8.48), but it also has one significant difference. This is that there is no relative sign between the particle number operator terms and the $b_s(\vec{p})b_s^\dagger(\vec{p})$ terms. However, this operator composed of the antiparticle creation and annihilation operators is not quite the number operator. In order to write it like the number operator we must exchange the order of $b(\vec{p})$ and $b^\dagger(\vec{p})$. In the case of the scalar field we simply assumed that this could be done, that is, we assumed that the creation and annihilation operators behaved like complex numbers and their order did not matter.

It turns out (though I do not prove it) that the components of the spinor fields need to *anticommute*. That is, one can switch the order of

multiplication only by also including a factor of -1. This implies that

$$b_s(\vec{p})b_s^\dagger(\vec{p}) = -b_s^\dagger(\vec{p})b_s(\vec{p}) \ , \tag{9.85}$$

and hence that

$$Q = 2(2\pi)^3 \int d^3p\, w_p \left[a_s^\dagger(\vec{p})a_s(\vec{p}) - b_s^\dagger(\vec{p})b_s(\vec{p})\right] \ , \tag{9.86}$$

which has exactly the same form as the complex scalar field.

It may seem that we have simply changed to rules here in order to get an answer that matches the scalar case. In reality, it is a deep result of quantum field theory that fermionic fields (fields with half-integer spin) must anticommute with all other fermionic fields (they commute with bosonic fields or fields with integer spin). The analysis above is not meant to prove that the fields must anticommute, rather I hope that it shows how the independent requirement that fermions anticommute helps us to understand that the conserved particle number current is quite similar to what we found for scalar fields.

We can produce some more evidence that anticommutation is needed by considering the stress-energy tensor (32.107).

Exercise 9.16 Show that the stress-energy tensor for the Dirac field is given by

$$T^\mu{}_\nu = i\overline{\psi}\gamma^\mu\partial_\nu\psi - \delta^\mu_\nu\left(i\overline{\psi}\not{\partial}\psi - m\overline{\psi}\psi\right) \ . \tag{9.87}$$

From the above result we find that the linear momentum carried by the Dirac field is

$$P^i = \int d^3x\, T^{0i} = i\int d^3x\, \psi^\dagger\partial^i\psi \ , \tag{9.88}$$

while the energy is given by

$$E = \int d^3x\, T^{00} = \int d^3x\left(-i\overline{\psi}\gamma^j\partial_j\psi + m\overline{\psi}\psi\right) \ . \tag{9.89}$$

This result for the energy can be simplified by remembering that the fields that appear are assumed to satisfy the Dirac equation. This means that

$$i\gamma^0\partial_t\psi + i\gamma^j\partial_j\psi = m\psi \quad \Rightarrow \quad i\gamma^j\partial_j\psi = m\psi - i\gamma^0\partial_t\psi \ . \tag{9.90}$$

Using this result we can write the energy as

$$E = i\int d^3x\, \overline{\psi}\gamma^0\partial_t\psi = i\int d^3x\, \psi^\dagger\partial_t\psi \ . \tag{9.91}$$

Exercise 9.17 Show that (remembering the conditions in (9.39) and (9.40) to simplify)

the momentum and energy can be written

$$P^i = 2(2\pi)^3 \int d^3p\, \omega_p \left[a_s^\dagger(\vec{p})a_s(\vec{p}) - b_s(\vec{p})b_s^\dagger(\vec{p})\right] p^i \ , \tag{9.92}$$

and

$$E = 2(2\pi)^3 \int d^3p\, \omega^2 \left[a_s^\dagger(\vec{p})a_s(\vec{p}) - b_s(\vec{p})b_s^\dagger(\vec{p})\right] \ . \tag{9.93}$$

The above exercise shows that by requiring the components of spinor fields to anticommute we obtain an understandable interpretation of the momentum. If we assumed the components behaved like bosonic fields then the momentum of the antiparticles would seem to be opposite to that of particles. That is, an antiparticle moving to the left would have momentum pointed to the right. This is not what we observe in collisions involving antiparticles. In the lab particles and antiparticles obey the same kinematic relations.

Perhaps even more important is the energy result. If the spinor are allowed to commute then there is no ground state, no state of minimum energy. The energy could always be reduced by adding more antiparticles. Again, we do not see this in nature. We observe that nature prefers to minimize energy, but we do not see nature producing an infinite number of antiparticles everywhere in space at all times. Therefore, the requirement that spinor fields anticommute is the consistent choice to model spinor fields that agree with observation:

$$P^i = 2(2\pi)^3 \int d^3p\, \omega_p \left[a_s^\dagger(\vec{p})a_s(\vec{p}) + b_s^\dagger(\vec{p})b_s(\vec{p})\right] p^i \ , \tag{9.94}$$

$$E = 2(2\pi)^3 \int d^3p\, \omega_p \left[a_s^\dagger(\vec{p})a_s(\vec{p}) + b_s^\dagger(\vec{p})b_s(\vec{p})\right] \omega_p \ . \tag{9.95}$$

In fact, the anticommutation behavior of fermions has profound physical effects. You may have heard of the **Pauli exclusion principle** [31] which asserts that no two fermions can be in the same state at the same time. This observed behavior gives rise to the population of energy shells in atoms, for instance. This is a direct consequence of anticommuting creation operators. Suppose that the operator $a_v^\dagger$ "creates" a particular fermionic state $|v\rangle$. That is,

$$a_v^\dagger|\text{vacuum}\rangle = |v\rangle \ . \tag{9.96}$$

Suppose we want to consider the state with two fermions in the state $|v\rangle$. This would be given by

$$a_v^\dagger a_v^\dagger|\text{vacuum}\rangle \ . \tag{9.97}$$

But, because $a_v^\dagger$ anticommutes with itself

$$a_v^\dagger a_v^\dagger = -a_v^\dagger a_v^\dagger \ , \tag{9.98}$$

which means that $a_v^\dagger a_v^\dagger = 0$. In other words, the vector in (9.97) is the zero vector, which is not a physical state. Therefore, there can only be one fermion in a given state. This is not true of bosons, which can have many particles in the same state.

Section 10

Spin-1 Lagrangian

The dynamics of a spin-1 field may be more familiar than those of scalars and spinors. Often the first field theory we study extensively in physics is classical electromagnetism, which we learn contains energy and momentum propagating through the electromagnetic fields. These propagating waves are often described with the four-vector field

$$A^\mu = \begin{pmatrix} A^0 \\ A^i \end{pmatrix} \ , \tag{10.1}$$

where A^0 is often called the electric potential, or scalar potential, while A^i is the vector potential. The first of these is unaffected by spatial rotations while the second rotates in the usual way of a spatial vector. In other words, the first part is in the $\mathbf{0}$ spin representation and the second is in the $\mathbf{1}$ representation. This is a perfect match for the $(\frac{1}{2}, \frac{1}{2})$ representation of the Lorentz group as this has spin $\frac{\mathbf{1}}{\mathbf{2}} \otimes \frac{\mathbf{1}}{\mathbf{2}} = \mathbf{1} \oplus \mathbf{0}$.

Rather than directly beginning with the Lagrangian that gives rise to electrodynamics, in this section we begin with a more general set-up. The Lagrangian for A^μ should be real, or Hermitian, and Lorentz invariant. We also want to produce the correct equations of motion, but for the moment we are supposing that these are unknown.

We expect the Lagrangian to be a function of A_μ and $\partial_\nu A_\mu$. We know that any Lagrangian where the four-vector indices are contracted (like $A_\mu A^\mu$ for instance) is Lorentz invariant. This field is also real, so any combination of the field and its derivatives with real coefficients leads to a real (and so Hermitian) Lagrangian. Let's make a guess, which is incorrect, but hopefully instructive

$$\mathcal{L}_{\text{incorrect}} = \frac{1}{2}\partial_\mu(A_\nu)\partial^\mu A^\nu - \frac{m^2}{2} A_\mu A^\mu \ . \tag{10.2}$$

Superficially, this may look reasonable, but when we expand out the vector

sums we find

$$\begin{aligned}\mathcal{L}_{\text{incorrect}} =& \frac{1}{2}\left(\partial_\mu A^0\right)\partial^\mu A^0 - \frac{m^2}{2}\left(A^0\right)^2 \\ &- \frac{1}{2}\left(\partial_\mu A^1\right)\partial^\mu A^1 + \frac{m^2}{2}\left(A^1\right)^2 \\ &- \frac{1}{2}\left(\partial_\mu A^2\right)\partial^\mu A^2 + \frac{m^2}{2}\left(A^2\right)^2 \\ &- \frac{1}{2}\left(\partial_\mu A^3\right)\partial^\mu A^3 + \frac{m^2}{2}\left(A^3\right)^2 \ . \end{aligned} \tag{10.3}$$

This makes clear that this is simply four copies of Klein-Gordon (which seems more like $\mathbf{0}\oplus\mathbf{0}\oplus\mathbf{0}\oplus\mathbf{0}$) , three of them with the wrong sign for the kinetic energy and mass terms. So, this Lagrangian cannot be what we are looking for.

Instead of guessing, let us try to be more systematic. There are three possible ways to write a Lorentz invariant kinetic energy type term:

$$\left(\partial_\mu A_\nu\right)\partial^\mu A^\nu \ , \qquad \left(\partial_\mu A^\mu\right)\partial_\nu A^\nu \ , \qquad \left(\partial_\mu A_\nu\right)\partial^\nu A^\mu \ . \tag{10.4}$$

Exercise 10.1 Show, by integrating by parts, that two of terms in Eq. (10.4) are equivalent.

The above exercise shows that we only need to include two of the possible derivative terms in (10.4). Therefore we simply parameterize the possible Lagrangians by

$$\mathcal{L} = \pm\frac{1}{2}\left[\left(\partial_\mu A_\nu\right)\partial^\mu A^\nu + c_1\left(\partial_\mu A^\mu\right)\partial_\nu A^\nu - m^2 A_\mu A^\mu\right] \ , \tag{10.5}$$

where we have included the one possible mass term, which we have taken to have opposite sign to the first derivative term in analogy with the Klein-Gordon Lagrangian. In the Lagrangian c_1 is a number that allows for various signs and relative magnitudes for the first two terms. Note that this method relies heavily on Coleman's analysis [15].

Exercise 10.2 Show that the equations of motion corresponding to the Lagrangian in Eq. (10.5) are

$$\left(\partial_\mu\partial^\mu + m^2\right)A^\nu + c_1\partial^\nu\partial_\mu A^\mu = 0 \ . \tag{10.6}$$

As we did with the Klein-Gordon and Dirac equations, we are interested in plane-wave solutions

$$A^\mu = \epsilon^\mu e^{-ix^\mu p_\mu} \ , \tag{10.7}$$

where ϵ^μ is a constant four-vector, often referred to as the **polarization vector** because it determines the spatial orientation of the plane wave.

Using this form in Eq. (10.6) we find

$$\left(-p^2 + m^2\right)\epsilon^\nu(p) - c_1 p^\nu p_\mu \epsilon^\mu(p) = 0 \ . \tag{10.8}$$

This may not look much like progress. To move forward we need to separate the spin-0 and spin-1 parts of the four-vector. The spin-0 part must be a Lorentz scalar. The only such scalar made from ϵ^μ alone is $\epsilon^\mu\epsilon_\mu$, but this clearly involves all parts of the vector, not only the spin-0 part. The only other four-vector at our disposal is p^μ. Making the definition $p \equiv \sqrt{p^\mu p_\mu}$ we can make the dimensionless, Lorentz scalar

$$\epsilon_L \equiv \frac{p_\mu}{p}\epsilon^\mu \ . \tag{10.9}$$

Here the L subscript indicates that this is the Longitudinal part of the polarization vector, the part that is along the momentum direction. The other parts of the polarization vector are those that are transverse to the momentum. We can focus on those parts of the vector by using the projector

$$P^\mu{}_\nu = \delta^\mu_\nu - \frac{1}{p^2}p^\mu p_\nu \ . \tag{10.10}$$

This can be used to project any vector into the three dimensional subspace that is orthogonal to p_μ.

Exercise 10.3 Show that $P^\mu{}_\nu$ satisfies

$$P^\mu{}_\alpha P^\alpha{}_\nu = P^\mu{}_\nu \ , \qquad P^\mu{}_\mu = 3 \ , \qquad p_\mu P^\mu{}_\nu = P^\mu{}_\nu p^\nu = 0 \ . \tag{10.11}$$

This shows that after a vector is projected into this subspace, using the projector more than once is equivalent to using it once. The projector also serves as the metric on this subspace, and the second result shows that the dimension of this subspace is three. The last equalities show that p^μ is not in this subspace.

We can now define the transverse polarization vector by

$$\begin{aligned}\epsilon^\mu_T \equiv P^\mu{}_\nu \epsilon^\nu =& \epsilon^\mu - \frac{1}{p^2}p^\mu p_\nu \epsilon^\nu \\ =& \epsilon^\mu - \frac{1}{p}p^\mu \epsilon_L \ .\end{aligned} \tag{10.12}$$

Note that by construction $p_\mu \epsilon^\mu_T = 0$. In short, we have

$$\epsilon^\mu = \epsilon^\mu_T + \frac{1}{p}p^\mu \epsilon_L \ . \tag{10.13}$$

Note here that ϵ^μ is a four dimensional object and ϵ_L is one dimensional.

Therefore, it must be that ϵ_T^μ describes three dimensional information. Even though it has a four dimensional index, the constraint $p_\mu \epsilon_T^\mu = 0$ means that it is orthogonal to the direction p^μ, meaning that it belongs to a three dimensional space.

Exercise 10.4 Use the decomposition of the polarization vector given in Eq. (10.13) in constraint equation (10.8). By separating the equation into the portion along p^μ and the portion transverse to the momentum, show that

$$(p^2 - m^2)\epsilon_T^\mu = 0 \ , \qquad \left(p^2 - \frac{m^2}{1+c_1}\right)\epsilon_L = 0 \ . \tag{10.14}$$

The results of the above exercise are quite interesting. If we choose $c_1 = 0$ then all the polarizations obey the usual relativistic energy relation $p^2 = m^2$. However, this is not exactly what we want. We do not want the spin-0 part at all, only the spin-1 part. The transverse polarization part clearly obeys the correct energy relation no matter what c_1 is. It looks like by choosing $c_1 = 0$ the longitudinal part would also, but we actually do not want it around at all. How can we remove ϵ_L from the theory?

Though it seems strange, the correct choice is to take $c_1 = -1$. This is equivalent, from Eq. (10.14), to making the "mass" of the longitudinal polarization infinitely large. You might be wondering why that is supposed to be helpful, but consider for a moment the dynamics of a very heavy particle. A particle with infinite mass can never be created, because it takes an infinite energy to do so. Indeed, any effect of this particle is suppressed by its mass. Effectively, an infinite mass particle has no effect on any other particles, it is like it is not there at all. Thus, by giving the longitudinal mode an infinite mass we remove it from the theory. We are left with only the three transverse polarizations, the spin-1 part.

Exercise 10.5 Show that

$$\frac{1}{2}\left[(\partial_\mu A_\nu)\,\partial^\mu A^\nu - (\partial_\mu A^\mu)\,\partial_\nu A^\nu\right] = \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \ , \tag{10.15}$$

where

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \ , \tag{10.16}$$

is sometimes called the **field strength** tensor.

We note that with $c_1 = -1$ the equations of motion in Eq. (10.6) is

$$\left(\partial_\mu\partial^\mu + m^2\right) A^\nu - \partial^\nu \partial_\mu A^\mu = 0 \ . \tag{10.17}$$

If we act on this equation with ∂_ν we find

$$\left(\partial_\mu\partial^\mu + m^2\right) \partial_\nu A^\nu - \partial_\nu\partial^\nu\partial_\mu A^\mu = m^2 \partial_\mu A^\mu = 0 \ . \tag{10.18}$$

For nonzero mass (this caveat is important in the next section), this implies that the constraint

$$\partial_\mu A^\mu = 0 \ , \tag{10.19}$$

must hold. This, in effect, reduces the number of independent fields from four to three. This also allows us to write the original equation of motion as

$$\left(\partial_\mu \partial^\mu + m^2\right) A^\nu = 0 \ . \tag{10.20}$$

While this appears to be simply four copies of the Klein-Gordon equation, remember that these four equations are connected by the constraint in Eq. (10.19).

Exercise 10.6 Show that the equations of motion for the massive spin-1 field can also be written as

$$\partial_\mu F^{\mu\nu} + m^2 A^\nu = 0 \ . \tag{10.21}$$

We are now interested in determining the overall sign of the Lagrangian by requiring the energy to be bounded from below. The Hamiltonian density (which for these theories is also the energy density) is given by

$$\mathcal{H} = \frac{\partial \mathcal{L}}{\partial(\partial_t A_\mu)} \partial_t A_\mu - \mathcal{L} \ . \tag{10.22}$$

Exercise 10.7 Show that

$$\frac{\partial \mathcal{L}}{\partial(\partial_t A_0)} = 0 \ . \tag{10.23}$$

This shows that there is no canonical momentum field to A_0! Go on to show that

$$\frac{\partial \mathcal{L}}{\partial(\partial_t A_i)} = \pm\left(\partial^t A^i - \partial^i A^0\right) = \pm F^{0i} \ , \tag{10.24}$$

where the index i spans the three spatial dimensions.

This result implies that the energy density is

$$\mathcal{H} = \pm\, F^{0i} \partial_t A_i \mp \frac{1}{2}\left[\left(\partial_\mu A_\nu\right)\partial^\mu A^\nu - \left(\partial_\mu A^\mu\right)\partial_\nu A^\nu - m^2 A_\mu A^\mu\right] \ . \tag{10.25}$$

This can be simplified using the constraint equation $\partial_\mu A^\mu = 0$ on the second term in brackets. We also want to write the Hamiltonian in terms of the field A_μ and the conjugate momenta F^{0i}. This entails writing

$$\partial_t A_i = F_{0i} + \partial_i A_0 \ . \tag{10.26}$$

We note also that by using the equations of motion, especially the form in

Eq. (10.21), we can use integration by parts to find

$$\int d^3x\, F^{0i}\partial_i A_0 = -\int d^3x\, A_0 \partial_i F^{0i} = -\int d^3x\, m^2 A_0 A^0 \ . \tag{10.27}$$

The Hamiltonian density then takes the form

$$\begin{aligned}\mathcal{H} &= \pm F^{0i}F_{0i} \pm F^{0i}\partial_i A_0 \mp \frac{1}{2}\left[\frac{1}{2}F^{\mu\nu}F_{\mu\nu} - m^2 A_\mu A^\mu\right] \\ &= \pm\left[F^{0i}F_{0i} - m^2 A_0 A^0 - \frac{1}{4}F^{0i}F_{0i} - \frac{1}{4}F^{i0}F_{i0} - \frac{1}{4}F^{ij}F_{ij} + \frac{1}{2}m^2 A_0 A^0 + \frac{1}{2}m^2 A_i A^i\right] \\ &= \pm\frac{1}{2}\left[F^{0i}F_{0i} - \frac{1}{2}F^{ij}F_{ij} - m^2 A_0 A^0 + m^2 A_i A^i\right] \ . \end{aligned} \tag{10.28}$$

Here we have used that, subject to the constraint in Eq. (10.19) that

$$(\partial_\mu A_\nu)\partial^\mu A^\nu = \frac{1}{2}F_{\mu\nu}F^{\mu\nu} \ , \tag{10.29}$$

and that $F_{0i} = -F_{i0}$. Notice that because the sum over spatial indices comes with a negative sign (from the Minkowski metric) that each term in square brackets of (10.28) is negative. This means the energy density is positive if we choose the lower, negative sign.

The upshot of all this is that the Lorentz invariant, Hermitian Lagrangian that has a positive energy density for fields that satisfy the equations of motion is

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{m^2}{2}A_\mu A^\mu \ . \tag{10.30}$$

This is usually called the Proca Lagrangian, after Alexandru Proca and the resulting equations are the Proca equations. This Lagrangian describes massive spin-1 fields. Such fields do appear in particle physics, but there is special interest in massless vectors, like the photon. We find in the next section that understanding the massless limit takes special care.

As with the other fields we've discussed the massive spin-1 field can be expressed as a sum of plane waves

$$A_\mu = \int d^3p \sum_{s=1}^{3}\left(\epsilon_s^\mu a_s(\vec{p})e^{-ip_\mu x^\mu} + \epsilon_s^{*\mu} a_s^\dagger(\vec{p})e^{ip_\mu x^\mu}\right) \ . \tag{10.31}$$

We see that this vector field is a real field, because $A_\mu^\dagger = A_\mu$. We also see that the sum over the three spin polarizations lead to the three degrees of freedom in this field, or are associated with the three sets of creation and annihilation operators a_1, a_2, and a_3.

Section 11

Massless Spin-1

In this section we consider massless spin-1 fields. This is more subtle than the spin-0 and spin-1/2 cases. In the massive spin-1 case discussed above we used a Lorentz four-vector A^μ to describe the field's three degrees of freedom. This was enforced by the constraint equation $\partial_\mu A^\mu = 0$, but this is only constraining for $m \neq 0$. Equivalently, we were able to give part of the field an "infinite" mass to remove it from the theory. This cannot be done (at least not in the same way) when $m = 0$.

We begin with the spin-1 Lagrangian we developed in the preceding section with $m = 0$

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} \ , \tag{11.1}$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the field strength tensor. Note that this Lagrangian is invariant under the transformation

$$A_\mu(x) \to A'_\mu(x) = A_\mu(x) + \partial_\mu \alpha(x) \ , \tag{11.2}$$

where $\alpha(x)$ is *any* differentiable function.

Exercise 11.1 Show that the transformation in Eq. (11.2) leads to $F_{\mu\nu} = F'_{\mu\nu}$. In other words, the field strength is invariant under such transformations.

This transformation looks like it could lead to a conserved current through Noether's theorem. The transformation preserves the form of the Lagrangian and it depends on a function which can be continuously taken to zero. However, one finds that

$$\left. \frac{dA_\nu}{d\alpha} \right|_{\alpha=0} = 0 \ . \tag{11.3}$$

Consequently, the would-be conserved current is

$$J^\mu = \frac{\partial \mathcal{L}}{\partial(\partial_\mu A_\nu)} \left. \frac{dA_\nu}{d\alpha} \right|_{\alpha=0} = 0 \ . \tag{11.4}$$

In fact, this current is conserved in the sense that $\partial_\mu 0 = 0$, but not in an interesting way. In particular, there is no conserved charge as

$$Q = \int d^3 J^0 = 0 \ . \tag{11.5}$$

What are we to make of these results? The lesson we should take away is that the transformation in Eq. (11.2) is not a physical symmetry of the theory. It is actually tied to our insistence of describing a massless spin-1 particle with a Lorentz four-vector. These symmetries are called **gauge transformations** and do not correspond to a physical change to

the system. Rather, it simply links two mathematically equivalent ways to describe a single system. In other words, the symmetry is in the description (sometimes called a redundancy of the description), not in the physical system that is being described.

We can, however, use this redundancy to understand what is described by the Lagrangian we have written down. The equations of motion that follow from Eq. (11.1) are

$$\partial_\mu F^{\mu\nu} = \partial_\mu \partial^\mu A^\nu - \partial_\mu \partial^\nu A^\mu = 0 \ . \tag{11.6}$$

However, we can simplify these equations considerably.

Example

Consider some massless vector field $\widetilde{A}^\mu$. We then make the gauge transformation

$$\widetilde{A}_\mu = \widehat{A}_\mu + \partial_\mu \widehat{\alpha} \ , \tag{11.7}$$

where

$$\partial_i \partial^i \widehat{\alpha} = \partial_i \widetilde{A}^i \ . \tag{11.8}$$

This implies that

$$\partial_i \widehat{A}^i = \partial_i \widetilde{A}^i - \partial_i \partial^i \widehat{\alpha} = 0 \ . \tag{11.9}$$

Thus, we have used the gauge freedom to pick a vector field that satisfies a specific constraint, in this case that $\partial_i \widehat{A}^i = 0$. Typically what is done is that one chooses a gauge (or particular constraint) that simplifies the problem being solved. The gauge chosen above is often called the Coulomb gauge.

Though we have made a gauge transformation we still have more gauge freedom to exploit. Note that in the gauge we have chosen the equations of motion (11.6) become

$$\partial_\mu \partial^\mu A^\nu - \partial^\nu \partial_0 A^0 = 0 \ . \tag{11.10}$$

Taking $\nu = 0$, we find

$$\partial_\mu \partial^\mu A^0 - \partial^0 \partial_0 A^0 = \partial_i \partial^i A^0 = 0 \ . \tag{11.11}$$

We then make a final gauge transformation

$$A_\mu = \widehat{A}_\mu - \partial_\mu \alpha \ , \tag{11.12}$$

with

$$\partial_i \partial^i \alpha = 0 \ . \tag{11.13}$$

Note that we still have that

$$\partial_i A^i = \partial_i \widehat{A}^i - \partial_i \partial^i \alpha = 0 - 0 = 0 \ , \tag{11.14}$$

so this gauge transformation is contained within the first choice that we made. The fact that we can make another choice within the original restriction is the evidence that we did not use up all of the gauge freedom. This allows us to make the additional choice

$$\partial_0 \alpha = A_0 \ , \tag{11.15}$$

which implies that

$$A_0 = \widehat{A}_0 - \partial_0 \alpha = 0 \ . \tag{11.16}$$

This is a second constraint, additional the first, that $A_0 = 0$. So, we have used the gauge freedom to put two constraints on the four fields that make up A_μ. This means that in general this four vector can only have two degrees of freedom.

We can see what "two degrees of freedom" means by considering a plane wave

$$A^\mu = \epsilon^\mu e^{-ix^\mu p_\mu} \ . \tag{11.17}$$

The $A_0 = 0$ constraint simple becomes $\epsilon^0 = 0$. The $\partial_i A^i = 0$ constraint becomes $p_i \epsilon^i = 0$. If we define the direction of propagation to be in the z direction then $p_i = \delta_i^3 p$ and so $\epsilon^3 = 0$. The remaining vector is confined to a two dimensional subspace. The basis chosen is often

$$\epsilon_R^\mu = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ i \\ 0 \end{pmatrix} \ , \quad \epsilon_L^\mu = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ -i \\ 0 \end{pmatrix} \ . \tag{11.18}$$

These may be familiar as one basis for describing the two polarizations of light. Having two independent polarizations (left- and right-handed polarization in this case) is what two degrees of freedom means.

Taking a step back, we see that the massless vector limit is somewhat subtle. If we insist on using a four vector to describe the massless spin-1 degrees of freedom then we must learn to deal with gauge invariance. It is worth noting that some researchers are developing other ways to describe spin-1 and other fields that do not introduce gauge invariance. However, the methods outlined above are standard, so you should be aware of them.

Subsection 11.1

Connections to E&M

As seen above, massless spin-1 fields are used to describe the massless, photon field. In this section we clarify some connections to the electromagnetism you may be more familiar with. We begin with Maxwell's equations

in Heavyside-Lorentz units with $c = 1$:

$$\nabla \cdot \vec{E} = \rho \ , \qquad \nabla \cdot \vec{B} = 0 \ , \tag{11.19}$$

$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \ , \qquad \nabla \times \vec{B} = \vec{J} + \frac{\partial \vec{E}}{\partial t} \ , \tag{11.20}$$

where ρ is the electric charge density and $\vec{J}$ is the electric current. While the $\vec{E}$ and $\vec{B}$ fields are what we measure, it is often convenient to use the scalar ϕ and vector $\vec{A}$ potentials to analyze a particular system. These are related to the physical fields by

$$\vec{E} = -\nabla\phi - \frac{\partial \vec{A}}{\partial t} \ , \quad \vec{B} = \nabla \times \vec{A} \ . \tag{11.21}$$

Exercise 11.2 Show that for any function $\alpha(t, \vec{x})$ that the transformation

$$\vec{A}' = \vec{A} + \nabla\alpha \ , \quad \phi' = \phi - \frac{\partial \alpha}{\partial t} \ , \tag{11.22}$$

leads to the same physical fields. This is referred to as a gauge transformation of the potentials.

The point of this exercise is that specifying the physical fields does not uniquely determine the potentials. By making a judicious choice of the function α one can often greatly simplify the calculations required to find what the fields are. However, within the realm of classical physics there is no need to consider these potentials as anything more than a mathematical convenience.

However, in single particle, nonrelativistic quantum mechanics it is not the electromagnetic fields that are directly used, but the potentials. The Schrödinger equation becomes

$$i\frac{\partial \psi}{\partial t} = \frac{1}{2m}\left(i\nabla + q\vec{A}\right) \cdot \left(i\nabla + q\vec{A}\right)\psi + q\phi\psi + V\psi \ , \tag{11.23}$$

where q is the electric charge of the particle represented by ψ and V is whatever potential ψ experiences beyond the electric one.

Exercise 11.3 Show that making a gauge transformation of the electromagnetic potentials in Eq. (11.23) leads to an identical Schrödinger equation, but for the field

$$\psi' = e^{-iq\alpha}\psi \ . \tag{11.24}$$

Remember that α is a function of space and time.

The above exercise shows that the unphysical ambiguity of the electromagnetic potentials matches directly onto the unphysical ambiguity of

quantum states. That is, that changing them by an overall phase has no physical effect. So, while the Schrödinger equation depends directly on the potentials rather than the fields, it does so in a gauge invariant way. Even quantum mechanical effects like the Aharonov-Bohm phases that seem to imply a certain reality to the potentials only do so in a gauge invariant way. In a similar way, the quantum field theory of electromagnetism, quantum electrodynamics (QED), is described in terms of the potentials, but leads to gauge invariant experimental predictions.

Before we can write down this quantum theory we need to rewrite Maxwell's equations. The vector notation of Eqs. (11.19) and (11.20) are used to indicate that E&M is covariant under rotations. That is, when a system is rotated in space all the electromagnetic fields associated with it rotate along with it and work in the same way. But, we know that we should expect more. We should be able to write down these equations in a way that is Lorentz covariant, including the effects of changing frames.

Both the electric $\vec{E}$ and magnetic $\vec{B}$ fields are three vectors, so you might expect each of them to be part of some four-vector generalization. In fact, things are a bit more complicated. Remember that under a boost that the components of a four vector V^μ transform as

$$V'^0 = \gamma\left(V^0 + v^i V^i\right) \ , \qquad V'^i = \gamma\left(V^i + V^0 v^i\right) \ , \tag{11.25}$$

where v^i is the relative velocity vector between the two frames. But in Maxwell's theory of electromagnetism, the effects of boosting along v^i changes the electric and magnetic fields according to

$$E'^i = \gamma\left(E^i + \epsilon^{ijk} v_j B_k\right) - (\gamma - 1)\frac{v^i}{v^2} E_j v^j \ , \tag{11.26}$$

$$B'^i = \gamma\left(B^i - \epsilon^{ijk} v_j E_k\right) - (\gamma - 1)\frac{v^i}{v^2} B_j v^j \ . \tag{11.27}$$

The important thing to notice is that components of the electric and magnetic fields are mixed with each other by a boost. This shows that neither are four vectors. Rather, we need an object that contains both fields. It turns out that an antisymmetric tensor in four dimensions has six independent terms, which is just what we need.

The, antisymmetric, field strength tensor is given by

$$F_{\mu\nu} = \begin{pmatrix} 0 & \mathrm{E}_x & \mathrm{E}_y & \mathrm{E}_z \\ -\mathrm{E}_x & 0 & -\mathrm{B}_z & \mathrm{B}_y \\ -\mathrm{E}_y & \mathrm{B}_z & 0 & -\mathrm{B}_x \\ -\mathrm{E}_z & -\mathrm{B}_y & \mathrm{B}_x & 0 \end{pmatrix} \ , \tag{11.28}$$

which we see contains all of both fields. We also note that the dual of this

tensor is defined to be

$$ {}^*F_{\mu\nu} = \frac{1}{2}\varepsilon_{\mu\nu\alpha\beta}F^{\alpha\beta} = \begin{pmatrix} 0 & \mathrm{B}_x & \mathrm{B}_y & \mathrm{B}_z \\ -\mathrm{B}_x & 0 & \mathrm{E}_z & -\mathrm{E}_y \\ -\mathrm{B}_y & -\mathrm{E}_z & 0 & \mathrm{E}_x \\ -\mathrm{B}_z & \mathrm{E}_y & -\mathrm{E}_x & 0 \end{pmatrix} . \tag{11.29} $$

Note that this tensor is not the Hermitian conjugate of $F_{\mu\nu}$. The use of dual here actually relates to the Hodge dual of differential forms, but you do not need to worry about that means to understand what follows. Maxwell's eight equations can now be written in a manifestly Lorentz covariant notation:

$$ \partial_\mu F^{\mu\nu} = J^\nu \ , \qquad \partial_\mu {}^*F^{\mu\nu} = 0 \ . \tag{11.30} $$

Here the four-vector current is given by

$$ J^\mu = \begin{pmatrix} \rho \\ \vec{J} \end{pmatrix} \ , \tag{11.31} $$

or in other words the charge density is the time-component of the four-vector current density. These two quantities can be combined into a Lorentz covariant four-vector because boosting a static field configuration does indeed produce the correct current.

Exercise 11.4 Show that the equations in (11.30) lead to the usual three-vector form of Maxwell's equations.

The current density is a conserved current, meaning that $\partial_\mu J^\mu = 0$. When this is expanded out we find

$$ 0 = \partial_\mu J^\mu = \partial_t \rho + \partial_i J^i \ \Rightarrow \ \partial_t \rho = -\nabla \cdot \vec{J} \ . \tag{11.32} $$

This is the usual continuity equation for electric currents.

The electric and magnetic fields have a more complicated behavior under Lorentz transformations than simple four-vectors. However, the scalar and vector potentials have a quite simple four-vector form

$$ A^\mu = \begin{pmatrix} \phi \\ \vec{A} \end{pmatrix} \ . \tag{11.33} $$

In terms of this four-vector potential the field strength has the simple form used above: $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. Gauge transformations are simply given by

$$ A'_\mu = A_\mu + \partial_\mu \alpha \ , \tag{11.34} $$

which was also the transformation considered in the general discussion of massless spin-1 fields.

Exercise 11.5 Show that for a field strength of the form

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \ , \tag{11.35}$$

that

$$\partial_\mu {}^*F^{\mu\nu} = 0 \ . \tag{11.36}$$

This is similar to taking $\vec{B} = \nabla \times \vec{A}$ and finding that $\nabla \cdot \vec{B} = 0$.

With the field strength defined in terms of the potential in this way, the Lagrangian for electromagnetism is

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} - A_\mu J^\mu \ . \tag{11.37}$$

By simply applying the Euler-Lagrange equations we find $\partial_\mu F^{\mu\nu} = J^\nu$ while the above exercise shows that our definition of the field strength in terms of the vector potential ensures that $\partial_\mu {}^*F^{\mu\nu} = 0$. Therefore, the rather simple expression in Eq. (11.37) encapsulates all of the richness of classical electrodynamics. Of course, being able to write down the Lagrangian is no guarantee that we can find simple equations for the fields in various concepts, but it does highlight some universal aspects of the theory as a whole.

This Lagrangian should have the same gauge invariance discussed above. However, the J^μ term is not obviously gauge invariant. Applying a gauge transformation we find

$$\begin{aligned} -\int d^4x \, A_\mu J^\mu &= -\int d^4x \, \left[A'_\mu + (\partial_\mu \alpha)\right] J^\mu \\ &= \int d^4x \left(-A'_\mu J^\mu + \alpha \partial_\mu J^\mu\right) \\ &= -\int d^4x A'_\mu J^\mu \ , \end{aligned} \tag{11.38}$$

where in the second line we have integrated the second term by parts and in the third line we use the fact that the current is conserved. This last point is crucial. The Lagrangian is only gauge invariant for conserved currents.

PART III

Quantum Electrodynamics

In this section we consider a quantum electrodynamics (QED). This is our first example of a gauge theory and exhibits many of the characteristics of more complicated theories. We also introduce a schematic overview of how such theories are used to make calculations.

Sec. 12. Local Symmetry
Sec. 13. Perturbations
Sec. 14. Running Coupling
Sec. 15. Discrete Symmetries

Section 12

Lagrangian from Local Symmetry

Quantum electrodynamics is the quantum field theory of electrically charged particles, like electrons. As you likely know, electrons are spin-1/2 particles, so we need the Dirac equation to describe the electron field ψ . The Lagrangian is

$$\mathcal{L} = \overline{\psi}\left(i\gamma^\mu\partial_\mu - m\right)\psi \ . \tag{12.1}$$

We saw in Sec. 9 that this Lagrangian is invariant under the field redefintion

$$\overline{\psi}' = e^{i\theta}\overline{\psi} \ , \qquad \psi' = e^{-i\theta}\psi \ , \tag{12.2}$$

where θ is any constant real number. The corresponding conserved current associated with this symmetry was also shown to be

$$J^\mu = \overline{\psi}\gamma^\mu\psi \ . \tag{12.3}$$

This rephasing everywhere by the same constant θ is called a **global symmetry**. This is in contrast to a **local symmetry**, where the rephasing can change in space or time $\theta(x)$. This much more general transformation is

$$\overline{\psi} = \overline{\psi}' e^{i\theta(x)} \ , \qquad \psi = e^{-i\theta(x)}\psi' \ . \tag{12.4}$$

Is the Dirac Lagrangian invariant under such a transformation?

Example

Consider the effect of making the transformation of Eq. (12.4) to the Dirac Lagrangian. We find

$$\begin{aligned}
\mathcal{L} =& \overline{\psi}' e^{i\theta(x)}\left(i\gamma^\mu\partial_\mu - m\right)e^{-i\theta(x)}\psi' \\
=& \overline{\psi}' e^{i\theta(x)}\left(i\gamma^\mu e^{-i\theta(x)}\partial_\mu\psi' + i\gamma^\mu\psi' e^{-i\theta(x)}(-i\partial_\mu\theta(x)) - m e^{-i\theta(x)}\psi'\right) \\
=& \overline{\psi}'\left(i\gamma^\mu\partial_\mu - m\right)\psi' + \overline{\psi}'\gamma^\mu\psi'\partial_\mu\theta(x) \ .
\end{aligned} \tag{12.5}$$

Making this local rephasing is not a symmetry of the Dirac Lagrangian. It produces an extra term associated with the conserved current of a global rephasing.

The extra term produced by the local rephasing in Eq (12.5) looks similar in form to the effect of the gauge transformations discussed in Sec 11. Remember that for a massless spin-1 particle described by A_μ we have the Lagrangian

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - J^\mu A_\mu \ , \tag{12.6}$$

with $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ and J^μ a conserved current . If we make a gauge transformation $A_\mu = A'_\mu + \partial_\mu\alpha(x)$ then we generate the term $-J^\mu\partial_\mu\alpha(x)$,

which looks a lot like the form obtained by the local rephasing.

Exercise 12.1 Show that the QED Lagrangian

$$\mathcal{L}_{\text{QED}} = \overline{\psi}\left(i\partial\!\!\!/ - e\,A\!\!\!/ - m\right)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \ , \tag{12.7}$$

is invariant under the transformations

$$\psi = e^{-i\theta(x)}\psi' \ , \qquad \overline{\psi} = e^{i\theta(x)}\overline{\psi}' \ , \qquad A_\mu = A'_\mu + \frac{1}{e}\partial_\mu\theta(x) \ . \tag{12.8}$$

The exercise above shows that by including a massless spin-1 field with gauge invariance, the Dirac fields can be locally rephased while keeping the full Lagrangian invariant. It is in this sense that requiring invariance under a local symmetry transformation leads to the inclusion of a gauge field, or a field that admits gauge transformations. These massless gauge fields lead to long range forces. We saw in Sec. 11.1 that electrodynamics is the classical theory of a massless spin-1 field. So, we have made a connection between the forces of E&M and the local symmetry of a Lagrangian.

This theory also gives us a specific current $J^\mu = e\overline{\psi}\gamma^\mu\psi$ for the photon field A_μ to couple to. Remember that in the Weyl basis the gamma matrices can be expressed in blocks of two-by-two matrices as

$$\gamma^\mu = \begin{pmatrix} \mathbf{0}_2 & \sigma^\mu \\ \overline{\sigma}^\mu & \mathbf{0}_2 \end{pmatrix} \ , \qquad \text{with} \quad \sigma^\mu = \begin{pmatrix} \mathbb{I}_2 \\ \sigma_i \end{pmatrix} \ , \quad \overline{\sigma}^\mu = \begin{pmatrix} \mathbb{I}_2 \\ -\sigma_i \end{pmatrix} \ . \tag{12.9}$$

Note that the latter vectors are four vectors in spacetime, each component of which is a two-dimensional matrix that contributes to the four-dimensional spinor space the gamma matrices act on. We can also write the four dimensional Dirac spinors in terms of the two-dimensional chiral projections

$$\psi = \begin{pmatrix} \psi_L \\ \psi_R \end{pmatrix} \ , \qquad \overline{\psi} = \psi^\dagger\gamma^0 = \begin{pmatrix} \psi_R^\dagger, & \psi_L^\dagger \end{pmatrix} \ . \tag{12.10}$$

This implies that the conserved current is

$$J^\mu = e\overline{\psi}\gamma^\mu\psi = e\psi_L^\dagger\overline{\sigma}^\mu\psi_L + e\psi_R^\dagger\sigma^\mu\psi_R \ . \tag{12.11}$$

Let us consider a few characteristics of this particle current that couples to the photon. First, the left-handed and right-handed fields are not coupled to each other by this interaction. This means that the photon couples fields of like chirality. Second, the coefficient of the coupling to either field is the same. This means the strength of the interaction between the photon to left-handed fields is the same as the coupling strength to right-handed fields. We refer to this as a vector coupling because left- and right-handed fields are treated in the same way.

The QED Lagrangian is often written as

$$\mathcal{L}_{\rm QED} = \overline{\psi}\left(i\not{D} - m\right)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \ , \tag{12.12}$$

where $\not{D} = \gamma^\mu D_\mu$ and

$$D_\mu = \partial_\mu + ieA_\mu \ , \tag{12.13}$$

is called the **gauge covariant derivative**. The reason for this name is that for

$$\psi = e^{-i\theta(x)}\psi' \ , \quad \overline{\psi} = e^{i\theta(x)}\overline{\psi}' \ , \quad A_\mu = A'_\mu + \frac{1}{e}\partial_\mu\theta(x) \ , \tag{12.14}$$

we find that

$$D_\mu\psi = e^{-i\theta(x)}D'_\mu\psi' \ , \quad \text{for} \quad D'_\mu = \partial_\mu + i\,e\,A'_\mu \ . \tag{12.15}$$

In short, the covariant derivative acting on the field transforms in the same way as the field itself under a local rephasing.

Exercise 12.2 Show that the Lagrangian for scalar QED

$$\mathcal{L} = (D_\mu\phi)^\dagger\, D^\mu\phi - \frac{m^2}{2}|\phi|^2 - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \ , \tag{12.16}$$

is invariant under

$$\phi = e^{-i\theta(x)}\phi' \quad \phi^\dagger = e^{i\theta(x)}\phi^{\dagger\prime} \quad A_\mu = A'_\mu + \frac{1}{e}\partial_\mu\theta(x) \tag{12.17}$$

This is most easily done by first showing that

$$D_\mu\phi = e^{-i\theta(x)}D'_\mu\phi' \ . \tag{12.18}$$

Exercise 12.3 Show that

$$[D_\mu,\, D_\nu]\,\psi = ieF_{\mu\nu}\psi \ . \tag{12.19}$$

Remember to use the derivative product rule. For those familiar with General Relativity this may remind you of how the Reimann tensor, which encodes the strength of the gravitational field, is related to the covariant derivative. For gauge covariant derivatives we see a similar relationship to the strength of the force related to the gauge field.

So far, we have considered only one type of particle that couples to the photon. However, in the standard model there are many particles that interact with the photons, many with different strength or sign to their interactions. Suppose we have two spin-1/2 particles ψ_1 and ψ_2. The

Lagrangian that is invariant under local rephasing of both fields

$$\psi_1 = e^{-iq_1\theta(x)}\psi_1' \ , \qquad \psi_2 = e^{-iq_2\theta(x)}\psi_2' \ , \qquad A_\mu = A_\mu' + \frac{1}{e}\partial_\mu\theta(x) \ , \tag{12.20}$$

is

$$\mathcal{L} = \overline{\psi}_1\left(i\not{D}^{(1)} - m_1\right)\psi_1 + \overline{\psi}_2\left(i\not{D}^{(2)} - m_2\right)\psi_2 - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \ , \tag{12.21}$$

where the two covariant derivatives are given by

$$D_\mu^{(i)} = \partial_\mu + i\,e\,q_i\,A_\mu \ . \tag{12.22}$$

In this Lagrangian the photon couples with the same coupling e to both fields, but each has its own charge q_i. Note that this implies that

$$\overline{\psi}_1\gamma^\mu\psi_2 = e^{i\theta(x)(q_1-q_2)}\overline{\psi}_1'\gamma^\mu\psi_2' \ , \tag{12.23}$$

under the local rephasing. We see that this term does not come back to itself, or that it does not conserve electric charge.

Exercise 12.4

In this problem you will explore a system of two fields ψ_1 and ψ_2 that carry electric charges q_1 and q_2, respectively. The Lagrangian is given in Eq. (12.21), but how do we determine what the coupling e is, and what are the charges?
Suppose, an experimentalist measures the coupling strength of the photon A_μ to ψ_1, and defines its charge to be 1. That is suppose we define

$$\widehat{e} \equiv e\,q_1 \ . \tag{12.24}$$

Making use of the relation

$$\widehat{e}\,\widehat{q}_i = e\,q_i \ , \tag{12.25}$$

find $\widehat{q}_1$ and $\widehat{q}_2$. Suppose a different experimentalist measures coupling of the photon to ψ_2 and defines its charge to be 1. That is, suppose we define

$$\widetilde{e} \equiv e\,q_2 \ . \tag{12.26}$$

Find $\widetilde{q}_1$ and $\widetilde{q}_2$.
In the Standard Model, with electric coupling e, the electron has electric charge $q_e = -1$, the up quark has charge $q_u = 2/3$, and the down quark has charge $q_d = -1/3$. Suppose we redefine the coupling to e' such that $q_d' = 1$. Find q_e' and q_u'.

The exercise above shows that what we call the coupling or charge is, to some degree, only convention. However, once we define one of the fields

to have a particular charge all the others are determined in terms of that choice. The ratios of charges are physical quantities.

Note also that all the local rephasings we have done are simply the multiplication of the fields in question by an element of the Lie group U(1). (This group is composed of all the numbers of the form $e^{i\theta}$.) For this reason one sometimes sees QED referred to as a U(1) gauge theory. In later sections we discuss gauge theories that are related to other Lie groups.

Section 13

Perturbation Theory & Feynman Diagrams

Now that we have an interacting theory that might be related to nature we are interested in making calculations that can be checked against experiment. In Sec. 6.1 it was argued that particular models are tested by comparing their predictions for cross sections with actual experimental data. In principle, all we need to do calculate the cross section for a particular process. However, for nearly every interesting model there is no known way to calculate the cross section exactly. Instead we must calculate perturbatively.

The philosophy of perturbation theory is to write the theory you care about, with Lagrangian $\mathcal{L}_{\text{Total}}$, in terms of several pieces

$$\mathcal{L}_{\text{Total}} = \mathcal{L}_0 + \sum_i g_i \mathcal{L}_i \ , \tag{13.1}$$

where $\mathcal{L}_0$ describes a theory that you can solve exactly and the g_i are dimensionless numbers with $g_i \ll 1$. Every quantity within the theory, like the states that describe a physical particle, are expressed as a series in the g_i. For instance

$$|\text{initial}\rangle = |\text{initial}\rangle_0 + g_i|\text{initial}\rangle_1 + g_i^2|\text{initial}\rangle_2 + \dots \ . \tag{13.2}$$

There are, in fact, certain quantum field theories that we can solve exactly. This means that can be part of what we call $\mathcal{L}_0$. They include the ones we met in the previous part of these notes. A spin-0 particle whose only potential is a mass term, a spin-1/2 particle with no interactions to other particles, and a spin-1 particle with no interactions with other particles. All of these are called free particles, because they do not interact with anything, so nothing can contain them. While we can completely determine these systems, they are not very interesting, because the particles never interact with anything. Most significantly, such a framework does not resemble our Universe, where different particles interact all the time. It does, however, give us a useful starting point to perturb away from.

The general idea is to consider two (or more) free particles approaching each other from far enough away that they would not interact much even in the full theory. We might label this initial state as $|\text{in}\rangle_0$ where the subscript implies the state is in the free theory. These particles interact in some way, producing two (or more) particle that fly way from the interaction point. After long enough time they are sufficiently far from each other that the free particle theory provides a good description. This final state might be labeled as $|\text{out}\rangle_0$.

Recall that the cross section for a specific process is given by (see Sec. 6.1 for velocity and energy dependence)

$$\sigma = \frac{1}{2E_A\, 2E_B\, |v_A - v_B|} \left|\langle \text{out}|\text{in}\rangle\right|^2 \ , \tag{13.3}$$

where A and B refer to the two particles in the initial state. By expanding the states as in (13.2) we find

$$\sigma = \frac{1}{2E_A\, 2E_B\, |v_A - v_B|} \left|{}_0\langle \text{out}|\text{in}\rangle_0 + g_i \left({}_1\langle \text{out}|\text{in}\rangle_0 + {}_0\langle \text{out}|\text{in}\rangle_1\right) + \ldots\right|^2 \ . \tag{13.4}$$

Note that the free particle states are an orthonormal basis for the very big Hilbert space that we are using to describe nature. This means that

$${}_0\langle \text{out}|\text{in}\rangle_0 = 0 \quad \text{for} \quad |\text{in}\rangle_0 \neq |\text{out}\rangle_0 \ . \tag{13.5}$$

That is, this leading part of the cross section is only nonzero when the final state is identical to the initial state. This is typically not what we are interested in. We want to know the cross section for particles to interact, or in other words, to scatter off each other.

In practice, rather than considering the perturbations of the states themselves, we define the **scattering matrix** S by

$$\sigma = \frac{1}{2E_A\, 2E_B\, |v_A - v_B|} \left|{}_0\langle \text{out}|S|\text{in}\rangle_0\right|^2 \ . \tag{13.6}$$

This matrix is constructed from the operators (composed of the free-particle fields) that mediate interactions between free-particle states. In this definition we have implicitly removed the identity matrix part that corresponds to no scattering, the particle just flying past each other. In other words, in the perturbation series

$$S = g_i S_1 + g_i^2 S_2 + \ldots \ , \tag{13.7}$$

the first nonzero term includes a factor of the perturbative couplings.

Some parts of ${}_0\langle \text{out}|S|\text{in}\rangle_0$ only proved kinematical information, which is the same for all models. In order to focus on what distinguishes one

model from another we strip this out. We are left with

$$\sigma = \frac{1}{2E_A\, 2E_B\, |v_A - v_B|} \times \int \left[\prod_{i=1}^{n} \frac{d^4 p_i}{(2\pi)^4} 2\pi \delta(p_i^2 - m_i^2) \right] \times |\mathcal{M}|^2 \, (2\pi)^4 \delta^{(4)} \left(p_A + p_B - \sum_{i=1}^{n} p_i \right) \ , \tag{13.8}$$

where the product index runs over all final state particles (if there are only two we only have p_1 and p_2, for instance). The only remnant of the inner product of the quantum states is $\mathcal{M}$, which is called the Lorentz invariant amplitude. This amplitude must be calculated order by order as a power series

$$\mathcal{M} = g_i \mathcal{M}_1 + \dots \ , \tag{13.9}$$

and, following from the form of the S matrix, is only nonzero when there are nonzero interactions g_i between the particles.

Perhaps the simplest example is the theory of a real scalar field ϕ defined by the Lagrangian

$$\mathcal{L} = \frac{1}{2} \partial_\mu \phi \partial^\mu \phi - \frac{m^2}{2} \phi^2 - \frac{\lambda}{4!} \phi^4 \ . \tag{13.10}$$

In this case the free particle Lagrangian is

$$\mathcal{L}_0 = \frac{1}{2} \partial_\mu \phi \partial^\mu \phi - \frac{m^2}{2} \phi^2 \ , \tag{13.11}$$

and the interactions are contained in

$$\mathcal{L}_{\text{int}} = -\frac{\lambda}{4!} \phi^4 \ . \tag{13.12}$$

Without this interaction this scalar field does not interact with itself. The theory is solvable, but not very interesting. Two ϕ particles would not interact with each other, they cannot scatter off each other. When λ is nonzero then the particles do scatter, but we cannot solve for their dynamics exactly, at least with known techniques. If, however, λ is not too large, then we can treat it as a perturbation of the free particle theory. Each nontrivial amplitude must include at least one factor of λ and so cross sections must have λ^2 dependence.

Another example is QED. The only interaction is between the photon A_μ and fermions f with electric charges q_f, like the electron $q = -1$. The interaction term of the Lagrangian is

$$\mathcal{L}_{\text{int}} = -e q_f \overline{f} \gamma^\mu f A_\mu \ , \tag{13.13}$$

and each cross section in QED end up as power series in

$$\alpha = \frac{e^2}{4\pi} \ , \tag{13.14}$$

which is sometimes called the **fine structure constant**. Low energy measurements of this quantity show that $\alpha \approx 1/137$. Because this number is small, the perturbation series is reliable.

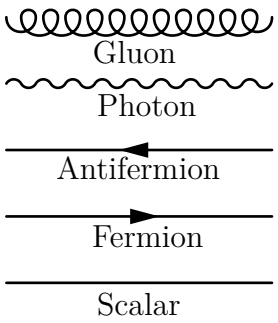


Figure 4. Common choices for free particle lines. The direction of momentum is left to right.

Feynman diagrams are a way to organize the calculation of $\mathcal{M}$ order by order in the perturbation couplings. This is often done in momentum space rather than position space because experimentally we typically deal with states of definite momentum, not definite position. The diagrams have two parts: lines and vertices. The lines correspond to a free particle with a specific momentum. Different types of particles are typically denoted by different types of lines. Some typical choices are shown in Fig. 4, but there are others. The arrows for the fermion lines show the direction of particle number flow. Similar arrows are often used for complex scalars (or vectors), which also have conserved particle number. For an anti-particle the flow of momentum is opposite the flow of particle number.

It probably cannot be overemphasized that neither the parts nor the whole of Feynman diagrams should be interpreted as a physical picture of particle interactions and dynamics. They are not indicating trajectories through spacetime, they are standing in for specific mathematics. For instance, the scalar line in Fig. 4 denotes the propagation of energy and momentum through the scalar field from one spacetime point to another. This can be written as a quantum mechanical amplitude as

$$\langle \text{vacuum} | \phi(x) \phi(y) | \text{vacuum} \rangle \ , \tag{13.15}$$

where a free-particle state is created a point y^μ and is then annihilated at point x^μ. This amplitude can be expressed in terms of the Green function for the Klein-Gordon equation.

Remember (see Sec. 41 for a review) that the Green function related to a differential operator is something like the inverse of the operator. The inverse of the Klein-Gordon equation

$$\left(\partial_\mu \partial^\mu + m^2 \right) \phi(x) \ , \tag{13.16}$$

is easy to see after Fourier transforming the field

$$\phi(x) = \int \frac{d^4 p}{(2\pi)^4} e^{-i p_\mu x^\mu} \widetilde{\phi}(p) \ . \tag{13.17}$$

We find the Klein-Gordon equation becomes

$$\left(p^2 - m^2 \right) \widetilde{\phi}(p) \ . \tag{13.18}$$

The propagator, which corresponds to the lines in Feynman diagrams, is then

$$\langle \text{vacuum}|\phi(x)\phi(y)|\text{vacuum}\rangle = \int \frac{d^4p}{(2\pi)^4} i \frac{e^{-ip_\mu(x^\mu - y^\mu)}}{p^2 - m^2 + i\varepsilon} \ , \tag{13.19}$$

where $0 < \varepsilon \ll 1$ ensures that this Green function produces the appropriate boundary conditions. In what follows, we typically do not write the $+i\varepsilon$ explicitly to make the equations a bit simpler.

The left-hand side of the propagator definition is useful for making a connection to the classical idea of a force. In single-particle quantum mechanics one can calculate the scattering cross section for a particle interacting with some potential $V(\vec{x})$. This result depends on the inner product of the potential between the incoming and outgoing scattering states.

Our form of the propagator above allows us to turn that idea around. It gives us the inner product between all possible states of definite momentum created by the field operators. We can use the form of the propagator infer what potential is created by the exchange of a particle of that type. In short, the propagator above describes the effect of the ϕ field transmitting energy and momentum from one place to another and this can be described by a potential between two objects.

Supposed one object (such as a light electron) has initial four momentum $p^{(i)}_\mu$. It interacts with another object (such as a heavy proton) and leaves with momentum $p^{(f)}_\mu$. The momentum exchanged through the ϕ field is then

$$q_\mu = p^{(f)}_\mu - p^{(i)}_\mu \ . \tag{13.20}$$

The simplest case is to consider elastic scattering, where internal energy of all the particles remains unchanged. We might also assume that the heavy object is so heavy that it is not moved. You can think of this as the infinite mass limit. Then the light object can change its direction by interacting with the heavy object, but not its energy, because the energy of the heavy object (which never moves) does not change. This means that the momentum transfer has the form

$$q^\mu = (0, \vec{q}) \ . \tag{13.21}$$

These are the only types momenta allowed in the propagator for this scenario. This means we are left with the integral

$$-\int \frac{d^3q}{(2\pi)^3} \frac{e^{i\vec{q}\cdot\vec{r}}}{\vec{q}\cdot\vec{q} + m^2} = -\frac{1}{4\pi r} e^{-rm} \ . \tag{13.22}$$

which is evaluated according to the methods discussed above Eq. (41.82).

This is called the Yukawa potential. Notice that it is has very small values for $rm \gg 1$. This means that at distances greater than $1/m$ the

potential is essentially zero, meaning the force is zero also.

Exercise 13.1 Find the Yukawa force by using

$$F(r) = -\frac{dV}{dr} \ . \tag{13.23}$$

is this an attractive force or a repulsive one? What is the force in the $m \to 0$ limit?

The point is that what we call forces are understood to be the quantum exchange of particles as described by the propagator. The exchange of massive particles gives rise to forces that have finite range. Infinite range forces, like electromagnetism and gravity, result from exchanging massless particles.

While the position-space propagator in (13.19) looks quite complicated, the momentum space propagator is proportional to

$$\frac{1}{p^2 - m^2} \ . \tag{13.24}$$

This much simpler object still encodes useful and physical information. For instance, if the four-momentum flowing through the scalar field is much larger than the field's mass then we can write the propagator as

$$\frac{1}{p^2 - m^2} = \frac{1}{p^2\left(1 - \frac{m^2}{p^2}\right)} = \frac{1}{p^2}\left(1 + \frac{m^2}{p^2} + \dots\right) \ . \tag{13.25}$$

At these energies the leading term dominates, which is the propagator of the massless particle. Similarly, at energies much lower than the scalar field mass

$$\frac{1}{p^2 - m^2} = \frac{-1}{m^2\left(1 - \frac{p^2}{m^2}\right)} = \frac{-1}{m^2}\left(1 + \frac{p^2}{m^2} + \dots\right) \ . \tag{13.26}$$

At these low energies the dominant form is not of a propagator at all, but of a $-1/m^2$ interaction.

This basic form of the propagator holds for all types of fields, but those with nonzero spin are modified somewhat. The Dirac equation, which describes spin-1/2 fields, leads to the equation

$$\left(\not{p} - m\mathbb{I}\right) u(p) = 0 \ , \tag{13.27}$$

for plane wave free particles (as opposed to anti-particles). The propagator is essentially the Green function of this operator, which is to say its inverse.

Antiparticles plane waves satisfy

$$\left(\not{p} + m\mathbb{I}\right) v(p) = 0 \ . \tag{13.28}$$

Exercise 13.2

Show that the inverse of

$$\not{p} \mp m\mathbb{I} \ , \tag{13.29}$$

is

$$\frac{\not{p} \pm m\mathbb{I}}{p^2 - m^2} \ . \tag{13.30}$$

You may find it useful to use (9.6) to first show that

$$\not{p}\not{p} = p^2 \ . \tag{13.31}$$

The above exercise shows that the propagators for fermions and anti-fermions are slightly different. Both, however, include the $1/(p^2 - m^2)$ of the scalar propagator, so the spin-1/2 particles propagate in a very similar way. Spin-1 particle plane waves lead to

$$\left(p^2 - m^2\right)\left(\delta^\mu_\nu - \frac{1}{p^2} p^\mu p_\nu\right)\epsilon^\nu(p) = 0 \ . \tag{13.32}$$

This also indicates that the inverse of this operator includes a $1/(p^2 - m^2)$ factor. Unfortunately, the spacetime matrix

$$P^\mu{}_\nu = \delta^\mu_\nu - \frac{1}{p^2} p^\mu p_\nu \ , \tag{13.33}$$

has no inverse. This is because it is a projector, with $P^\mu{}_\nu p^\nu = 0$, or in other words p^μ is an eigenvector with eigenvalue zero. Overcoming this issue is a little subtle. Essentially the theory is modified in a way that produces an invertible matrix, but does not change the physical predictions.

Subsection 13.1

Interaction Vertices

In Feynman diagrams the free particle propagators of various types of particles are represented by lines. These may be draw as straight or curved, but this does not matter, it is up to the whim of the drawer. What does matter is how these lines are connected. Connecting the free particle lines in different ways leads to different types of scattering, different physical predictions. Therefore, it is the interaction terms in the Lagrangian, which lead to specific vertex types in the diagrams, that produce specific physical effects.

Consider the simple scalar theory described by the Lagrangian in (13.10). The free particle ($\mathcal{L}_0$) part of the Lagrangian produces the scalar propaga-

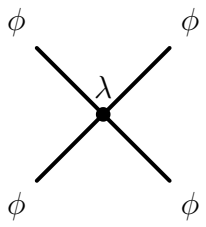


Figure 5. Vertex of self-interacting scalar theory.

tor. The interaction term $\mathcal{L}_I$ produces the vertex shown in Fig. 5. In the figure four scalar lines come together at one vertex. This is because interaction term in the Lagrangian has four powers of the the scalar field, ϕ^4. This vertex can mediate the nontrivial scattering of two ϕ particles. If we take time to flow from the left of the figure to the right, this is interpreted as two particles coming together, interacting with strength λ, and flying apart. This leading interaction can be used to calculate $\mathcal{M}$ at order λ and hence the cross section σ for $\phi\phi \to \phi\phi$ scattering at order λ^2. If λ is small, then this result is a good approximation of the true cross section.

Quantum electrodynamics provides another illustration of these ideas. In QED the coupling is e, and is multiplied by whatever charge the fermion has. The form of the QED interaction (13.13) couples an antifermion to a fermion and a photon. We can represent this interaction by the diagram in Fig. 6. But, it is important to recognize that the three legs of this vertex can be oriented in any way. For instance, the use of this vertex in Fig. 7 is also a correct manifestation of the interaction. As long as one of the fermion arrows points into the vertex and the other points out of the vertex the lines can take any orientation.

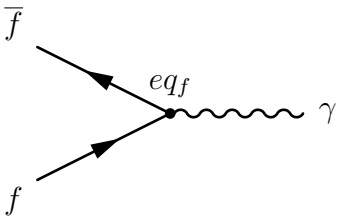


Figure 6. QED vertex with time going from left to right.

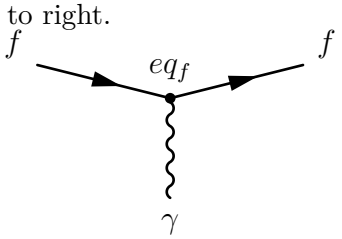


Figure 7. QED vertex with time going from left to right.

The many possibilities for orienting the vertex relates to the truth that it is easy to take Feynman diagrams too seriously. They are a method for organizing a perturbation series, but they are not meant to describe actual particle trajectories through spacetime. Consequently, the orientation of particular lines and vertices does not communicate "what the particle is doing." So, while Feynman diagrams are tremendously useful for making calculations, we must not confuse them with a picture of reality. Indeed, they are not even pictures of what quantum field theory as a framework is describing. Feynman diagrams are simply a useful calculational tool.

As an example, we can consider the process $e^+e^- \to \mu^+\mu^-$, that is to say, suppose we are interested when an electron and a positron annihilate and produce a muon-antimuon pair. (The muon is a particle with the same electric charge as the electron, but whose mass is $m_\mu = 105$ MeV, about 200 times the mass of the electron.) The leading order diagram for this process is given in Fig. 8. What this means is that there is no way to start a diagram with an electron and a positron and end it with a muon and anitmuon that uses fewer QED vertices. In other words, this leading diagram is order e^2 and there is no way to construct a QED diagram at order e. This diagram also allows us to see that there is a qualitative difference between external lines (those that begin and end the diagram) and internal lines, which do not correspond to either an initial state or final state particle.

Figure 8. Leading order (e^2) diagram for $e^+e^- \to \mu^+\mu^-$.

Initial and final state particles are taken to be plane wave solutions of the free particle part of the Lagrangian. This means that, in momentum space, they have a definite momentum. Each external momentum satisfies $p^2 = m^2$ for the specific mass of the particle in question. At each vertex of the diagram momentum is conserved. Note that this ensures to conser-

vation of four-momentum. In fact, other quantities like electric charge and angular momentum are also conserved at every vertex. However, it also has other implications. For instance, what is the value of the momentum carried by the photon line in Fig. 8? We find that

$$p^\mu_{(\gamma)} = p^\mu_{(+)} + p^\mu_{(-)} \ , \tag{13.34}$$

where the $p^\mu_{(\pm)}$ are the four momenta of the electron and positron. This implies that

$$p^2_{(\gamma)} = 2m_e^2 + 2E_{(+)}E_{(-)} - 2\vec{p}_{(+)} \cdot \vec{p}_{(-)} \ . \tag{13.35}$$

Note that $p^2_{(\gamma)}$ is a Lorentz invariant so we can determine its value in any convenient frame. Let us choose the center of momentum frame, in which $-\vec{p}_{(-)} = \vec{p}_{(+)} \equiv \vec{p}$ and hence that $E_{(+)} = E_{(-)} \equiv E$. We then find that

$$p^2_{(\gamma)} = 2m_e^2 + 2E^2 + 2\vec{p}^2 = 4E^2 \ , \tag{13.36}$$

recall that $E^2 = m_e^2 + \vec{p}^2$. The most important point of this is that $p^2_{(\gamma)} > 0$, but a physical external photon always has $p^2 = 0$, because it is massless!

When a particle does not satisfy $p^2 = m^2$, we say that it is off-shell (otherwise we say it is on-shell). The shell in question is simply the surface in four-dimensional space specified by $p^2 = m^2$. So, while the diagram in question seems to communicate that a photon is created which then produces muons, we see that this is an unfamiliar type of photon, one that is never seen as an external particle. It may be more useful to consider that energy and momentum is communicated from the electron field to the muon field through the photon field.

We can better understand the meaning of off-shell particles, sometimes called virtual particles, by considering the propagator given in Eq. (13.19). This function describes the amplitude (related to the probability) for energy and momentum to be transmitted from point y^μ to point x^μ. The amplitude gets large for $p^2 = m^2$, because of the singularity there. The meaning of this is that on-shell propagation has the largest amplitude.

But how much smaller is off-shell propagation, or is there any circumstance when it cannot be neglected? This is determined, in part, by the complex exponential

$$e^{-ip_\mu L^\mu/\hbar} = \cos\frac{p_\mu L^\mu}{\hbar} + i\sin\frac{p_\mu L^\mu}{\hbar} \ , \tag{13.37}$$

where $L^\mu = y^\mu - x^\mu$ and we have written the factors of $\hbar$ explicitly for the moment.

Exercise 13.3 Consider the integral of an exponentially growing function over a finite

domain

$$\int_0^1 dx\, e^x \ . \tag{13.38}$$

What is the numerical values of this integral? Next, modify the integral to

$$\int_0^1 dx\, e^x \sin(ax) \ , \tag{13.39}$$

and evaluate it. About what value of $a > 0$ makes this integral an order of magnitude less than the first integral, what about two orders of magnitude?

The above exercise illustrates a general result. When some function F is multiplied by a sinusoidal function that is oscillating "quickly" then the integral of the function times the sinusoid is reduced. The larger the frequency of oscillation the smaller the integral gets. Here the notion of quickly is important. It must be oscillating on a much shorter length scale that the characteristic length scale of F.

The propagator (13.19) has exactly this form. If we take the momentum to have dimensions of energy E, then for some length L if

$$\frac{EL}{\hbar c} \ , \tag{13.40}$$

is large then the integral is damped. Only when the momentum is right on the pole $p^2 = m^2$ is there any appreciable amplitude. For measurable lengths and energies $EL \gg \hbar c$ so only on-shell particle propagation is observed. However, when we give up the measurement aspect of this analysis other possibilities arise. When two particles interact energy and momentum can be transmitted on very small scales, such that $EL \sim \hbar c$. In this case there can be sizable probabilities for off-shell momenta to be transmitted through a field. So, the essential point is that only on-shell propagation is allowed over arbitrarily long spacetime distances. Off-shell propagation is perfectly physical, but can only occur over short distances or times.

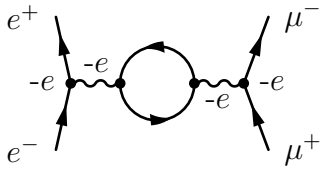


Figure 9. A higher order (e^4) diagram for $e^+e^- \to \mu^+\mu^-$.

As we consider higher order corrections to processes we should again keep in mind that these diagrams are an organizational tool, not a picture of reality. For instance, in Fig. 8 we have the leading order diagram for the $e^+e^- \to \mu^+\mu^-$ process, but this is only the first nontrivial term in an infinite series. One of the next terms is shown in Fig. 9. This diagram is order e^4. As long as the coupling e is smaller than one, this term is suppressed relative to the leading diagram. But, if one wants to calculate with higher precision then this diagram is essential to finding the correct result.

Note that if one tries to take the diagram in Fig. 9 as a description of what is "really going on" then one must construct a quite confusing picture of reality. However, quantum field theory does not require, or even

suggest, that this must be done. A physical measurement of this process can be made and to match theory with experiment one must include as many diagrams as one needs to match the desired precision. But the terms in a perturbation series do not convey all the information about the total process. In fact, while we do not say much more about them in this text, there can be contributions to physical processes that are not captured at all by the perturbation series. It can be that these nonperturbative effects are small and can be safely ignored. There are some situations, however, in which the perturbative contribution vanishes and so the nonperturturbative result gives the entire result.

Section 14

Running Couplings

In the preceding sections we have shown a connection between a Lagrangian and physical cross sections. In this section we reexamine the meaning of the terms in the Lagrangians of interacting theories. That is, what is the connection between the parameter m in a Lagrangian and the physically measured mass of a particle? While the idea that these might not be the same quantity may at first seem strange, it is essential to understanding such theories.

Consider the free Dirac Lagrangian

$$\mathcal{L} = \overline{\psi}\left(i\partial\!\!\!/ - m\right)\psi \ . \tag{14.1}$$

When describing this theory we claimed that m was the mass of the particle in question and this was justified by the plane wave solutions which satisfied the relativistic energy relation, $E^2 = m^2 + \vec{p}^2$. However, in an interacting theory, such as QED

$$\mathcal{L} = \overline{\psi}\left(i\partial\!\!\!/ - \ e\,A\!\!\!/ - m\right)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \ , \tag{14.2}$$

it turns out that the physical mass m_{phys} and physical coupling e_{phys} have a more complicated relationship to the parameters of the Lagrangian.

Consider the fermion mass. Notice that in the absence of interactions we understand this parameter in terms of the fermion propagator

$$\Delta_F(p) = \frac{p\!\!\!/ \pm m}{p^2 - m^2} \ . \tag{14.3}$$

This is clearly a function of the momentum of the particle and the Lagrangian mass parameter. In the free theory the meaning of m is directly connected to a propagating plane wave. The simple diagram for the free particle propagation is given in Fig. 10.

Figure 10. Leading order diagram for fermion propagation.

Figure 11. Higher order diagram contribution to fermion propagation.

But in QED this is merely the first term in an infinite series of diagrams. For instance, the next correction is related to the diagram in Fig. 11. We see that this is an e^2 correction to the first term. Because the process of physical propagation is corrected at this order, it turns out that the physical mass of the particle (as defined by on-shell propagation) is also corrected at e^2

$$m_{\text{phys}} = m + \frac{e^2}{16\pi^2} m_1 + \dots \ , \tag{14.4}$$

where m_1 is some quantity with dimensions of mass that can be evaluated from the diagram. The QFT origin of the factor of $16\pi^2$ is not important to this discussion, but it indicates a good heuristic, that QED calculations are power series in $\alpha/(4\pi)$, where $\alpha = e^2/(4\pi)$.

We emphasize that the physical mass of the particle is a fixed quantity. Equation (14.4) is meant to illustrate that the parameter in the Lagrangian is not quite what we mean by the physical mass of a particle. In QED, where $\alpha/(4\pi) \sim 0.0006$ at low energies, these corrections are small and so the Lagrangian and physical masses are pretty close in value. However, they are distinct objects.

A similar story plays out with the gauge coupling e. Of course, in the non-interacting theory $e = 0$, but we can also find that vertex given in Fig. 12 is only the leading term in a series. The next term is given in Fig. 13. This, as with the mass, makes the association between the Lagrangian parameter and the physical coupling of charged particles to the photon somewhat complicated.

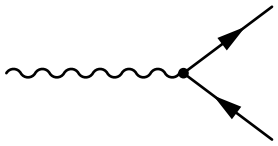


Figure 12. Leading order QED vertex.

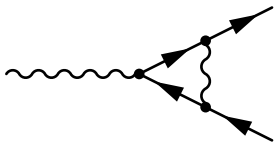


Figure 13. Higher order correction to the QED vertex.

One of the most interesting complications is that the physical coupling ends up being a function of the energy scale $e(Q)$. The way this energy dependence is typically discussed is in terms of off-shell fluctuations of the electron field. The leading diagram for such a fluctuation is given in Fig. 14. Notice that there are no external particles, so this is called a vacuum-to-vacuum process. It is a prediction of quantum field theory that these types of fluctuations are a continual background to the dynamics we observe.

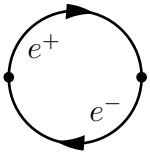


Figure 14. Leading order vacuum fluctuation of the electron field.

Consider this background of particle-antiparticle pair fluctuations near an external charged particle, like an electron. We expect the positron part of each fluctuation to be attracted to the external electron while the electron part of the fluctuation should be repelled. In short, seeing as these fluctuations are happening everywhere all the time, the entire volume of space around the electron should be polarized. This is a striking prediction, that what we think of as empty space is a polarizable medium. What is more striking is that it is borne out experimentally. These polarized fluctuations act as a shield of the electron's true interactions with the photon. At higher and higher energies, we can probe closer and closer distances to the electron and hence measure a different, less shielded, value.

The way this is described in quantum field theory is through the equa-

tion

$$Q\frac{d\alpha}{dQ} = \beta(\alpha) \ , \tag{14.5}$$

where Q is the energy scale. The right-hand side of the equation is called the beta function and can be calculated order by order in perturbation theory. For instance, in QED at leading order one finds

$$\beta(\alpha) = \frac{2\alpha^2}{3\pi} \ . \tag{14.6}$$

Exercise 14.1

Suppose that the value of α is measured at some scale Q_0. Use Eqs. (14.5) and (14.6) to determine that the value of α at some other scale Q is given by

$$\alpha(Q) = \frac{\alpha(Q_0)}{1 - \frac{2}{3\pi}\alpha(Q_0)\ln\frac{Q}{Q_0}} \ . \tag{14.7}$$

If $Q > Q_0$ how does $\alpha(Q)$ compare to $\alpha(Q_0)$?

This is the correct, leading order, result for a single charge particle, but we know that there are many: the electron, muon, tau, and six kinds of quarks. The correct generalization to many fields with charges q_i is

$$\alpha(Q) = \frac{\alpha(Q_0)}{1 - \frac{2q^2(Q)}{3\pi}\alpha(Q_0)\ln\frac{Q}{Q_0}} \ , \tag{14.8}$$

where

$$q^2(Q) = \sum_i q_i^2 \qquad \text{sum over all particles with } m_i < Q \ . \tag{14.9}$$

You may wonder why only particles with $m < Q$ are included in the sum. When considering the vacuum fluctuations of a particular field in real physical space one finds they have a characteristic size, even for noninteracting field. There is only one characteristic of such a particle that can determine this size: the mass of the field. Therefore, the higher the mass of the particle, the smaller the physical size of the vacuum fluctuations. Probing at energies below the mass of a particular particle means using wavelengths that are long enough to always interact with both the particle and antiparticle. These particles do not contribute to charge screening at that energy.

In either case we see that for larger and larger Q the measured coupling becomes larger and larger as well. There is even, in principle, a finite energy scale at which the coupling becomes infinite. Lev Landau, a famous Russian physicist, saw this as evidence that quantum field theory must be corrected by some deeper theory.

Exercise 14.2 Begin with the one-loop equation for the running of α with energy scale Q in Eq. (14.7) and find the scale at which the coupling becomes infinite. At the mass of the electron 0.0005 GeV the value of α is about 1/137. If $\alpha = \infty$ then $1/\alpha = 0$. Set $1/\alpha(Q) = 0$, and solve for the energy scale Q in GeV at which such a singular coupling is reached, this is often called the Landau pole. How does this compare with the Planck scale, which is about 10^{19} GeV?

Subsection 14.1

Renormalization

The notion that masses and couplings are corrected from the Lagrangian values by quantum corrections is closely related to the concept of **renormalization**. When loop corrections are included some quantum field theories only produce corrections to the quantities that appear in the original Lagrangian, whereas other theories produce new types of interactions. The former theories are said to be renormalizable. The connection between the Lagrangian parameters and the physical quantities may be determined by a series of quantum corrections, but the same number of independent quantities (mass, coupling, etc) describe the original Lagrangian and the quantum corrected interactions. Quantum electrodynamics in one such renormalizable theory.

Theories of this type can have validity up to high energy scales because they keep the same form. Such theories have the possibility of being fundamental, that is, of being the ultimate theory of the Nature's structure, because their form persists up to arbitrarily high scales. Nonrenormalizeable theories may be very useful up to some energy scale, but above that scale they must be corrected or completed into a different (presumably renormalizable) theory.

How then can we know that a theory is renormalizable? Specifically, is there any way to know without calculating all quantum corrections? It has been shown that there is, in fact, a fairly simple way of determining whether or not a theory is renormalizeable without calculating even a single quantum effect. However, in order to determine this we need to know the dimension of a given field.

In the "natural" units we are using there is only one dimension: Energy. The dimensions of a Lagrangian or Hamiltonian are exactly energy. The dimensions of a Lagrangian or Hamiltonian density is then energy divided by volume or in other words energy to the fourth power. We might denote this as

$$[\mathcal{L}] = 4 \ . \tag{14.10}$$

The dimension of a given field is then determined by the derivative terms in the Lagrangian. For instance, the Lagrangian for a spin-0 field

contains the term

$$\mathcal{L} \supset \frac{1}{2}\partial_\mu\phi\partial^\mu\phi \ . \tag{14.11}$$

Each derivative has dimensions of inverse length, which is equivalent to energy. This leads to

$$4 = [\mathcal{L}] = 2\,[\partial_\mu] + 2\,[\phi] = 2 + 2\,[\phi] \ . \tag{14.12}$$

In other words the field ϕ has dimensions of energy, or $[\phi] = 1$.

Exercise 14.3

Using methods similar to those above show that the dimension of a spin-1/2 field is

$$[\psi] = \frac{3}{2} \ , \tag{14.13}$$

and the dimension of a spin-1 field is

$$[A_\mu] = 1 \ . \tag{14.14}$$

Seeing that we now know the dimension of each type of field we can determine the dimension of the coefficients of different combinations of fields. As a simple example consider the term

$$\mathcal{L} \supset c\phi^2 \ . \tag{14.15}$$

The field ϕ has dimension 1 so ϕ^2 has dimension 2. Because the Lagrangian density must have dimension 4, we see that c must have dimension 2. This is why this term is written

$$\mathcal{L} \supset \frac{m^2}{2}\phi^2 \ . \tag{14.16}$$

Exercise 14.4

Find the dimension of the coefficient c in the term

$$\mathcal{L} \supset c\overline{\psi}\psi \ , \tag{14.17}$$

where ψ is a spin-1/2 field. What about the term

$$\mathcal{L} \supset c\phi\overline{\psi}\psi \ , \tag{14.18}$$

where ϕ is a spin-0 field?

We now state, without proof, the defining characteristic of renormalizable theories. Theories in which the coefficient of every term in the Lagrangian has non-negative dimension are renormalizeable. For instance, in QED the interaction term is

$$eqA_\mu\overline{\psi}\gamma^\mu\psi \ . \tag{14.19}$$

The fields have a combined dimension of 4, so the coefficient is dimensionless. Because this is non-negative it is a renormalizeable interaction. Enrico Fermi's first theory of beta decay involved four fermion fields

$$G_F \overline{\psi}_n \psi_p \overline{\psi}_\nu \psi_e \ . \tag{14.20}$$

The fields have a combined dimension of 6, so the dimension of G_F is -2. This interaction is non-renormalizeable. Consequently, we expect that it is an approximation of a more fundamental, renormalizable theory. In this case we know what the deeper theory is (though we have not reached it in this text yet). The true interaction between the four fermions is mediated by a massive spin-1 particle, the W boson.

Section 15

Discrete Symmetries

We have made considerable effort to understand the consequences of symmetries. From Noether's theorem we learned that there is a connection between symmetries of the Lagrangian and conserved quantities. This holds for all such symmetries that are parameterized by a continuous parameter. However, there are also symmetries that are not of this form. In particular, there can be discrete symmetries, those whose set of symmetry operations is finite.

Perhaps the simplest example is a Z_2 symmetry. Recall the Klein-Gordon Lagrangian

$$\mathcal{L} = \frac{1}{2}\partial_\mu(\phi)\partial^\mu\phi - \frac{m^2}{2}\phi^2 \ . \tag{15.1}$$

This Lagrangian is invariant under $\phi \to -\phi$. Because this is a free theory there is no interesting consequence, so consider the interacting theory

$$\mathcal{L} = \frac{1}{2}\partial_\mu(\phi)\partial^\mu\phi - \frac{m^2}{2}\phi^2 - \frac{\lambda}{4!}\phi^4 \ , \tag{15.2}$$

where the factor of 4! is conventional. This interacting theory is also invariant under $\phi \to -\phi$.

One consequence of this symmetry is that any physical process can only change the number of ϕ particles by two units. Note that this is something like a weaker version of conserved particle number. We can understand this as follows. Any initial state has some fixed number of ϕ particles and under $\phi \to -\phi$ this state either changes sign or does not. The final state process should have the same sign because all of the interactions involve even numbers of ϕs and so do not change the sign. So, one might consider $\phi\phi \to \phi\phi$ scattering or the $\phi\phi \to \phi\phi\phi\phi$ production of additional particles. But, the process $\phi\phi \to \phi\phi\phi$ must vanish.

Note that this would not be the case if the theory included a ϕ^3 term in the Lagrangian. In this case there would be no such symmetry and one can find nonzero cross sections for process that change the number of particles by one unit. So, if one was experimentally probing a scalar field theory and found a nonzero cross section for $\phi\phi \to \phi\phi\phi$ then you would know that theory could not be modeled by a Lagrangian with $\phi \to -\phi$.

Recall that symmetries in quantum mechanics are represented by unitary operators

$$|\Psi\rangle \to |\Psi'\rangle = U|\Psi\rangle \ , \tag{15.3}$$

so that inner products are preserved

$$\langle\Phi|\Psi\rangle \to \langle\Phi'|\Psi'\rangle = \langle\Phi|U^\dagger U|\Psi\rangle = \langle\Phi|\Psi\rangle \ . \tag{15.4}$$

We then find that for an operator O

$$O|\Psi\rangle \to O'|\Psi'\rangle = UO|\Psi\rangle \tag{15.5}$$

$$= O'U|\Psi\rangle \ . \tag{15.6}$$

The top equality holds because $O|\Psi\rangle$ is itself a vector and the second equality uses the transformation law of $|\Psi\rangle$ alone. These two results imply that

$$O' = UOU^\dagger \ , \tag{15.7}$$

which is the transformation law for operators.

Remember that the field ϕ is an operator in our quantum model. For the Z_2 symmetry above we have

$$U(z_2)\phi U(z_2)^\dagger = -\phi \ . \tag{15.8}$$

The above symmetry is an internal one, having to do with the particles and not with spacetime. However, there are also discrete symmetries (parity and time-reversal) that have to do with the structure of spacetime. Charge conjugation is a related idea that has to do with the relationships between particles and antiparticles. We discuss these different discrete symmetries in the following subsections.

Subsection 15.1

Parity

We first consider parity, which takes three-vectors to their negative

$$\vec{v} \to -\vec{v} \ . \tag{15.9}$$

Note that Newton's 2nd law is invariant under parity transformations

$$\vec{F} = m\vec{a} \;\rightarrow\; -\vec{F} = m(-\vec{a}) \; . \tag{15.10}$$

This shows that parity is a symmetry of at least some familiar physical theories. Note also that acting with parity twice simply brings a vector back to itself. That is, twice parity is the identity operation.

Within the quantum Hilbert space we define the parity operator as P. It seems reasonable that the action of parity on a scalar field ϕ might at most change it by only a multiplicative factor

$$\phi \xrightarrow{P} P\phi P^\dagger = \eta_P \phi \; . \tag{15.11}$$

Acting with parity twice should be equivalent to do doing nothing, so we have that

$$\phi \xrightarrow{P} PP\phi P^\dagger P^\dagger = P\eta_P \phi P^\dagger = \eta_P^2 \phi \; , \tag{15.12}$$

must be equal to ϕ itself. This shows that only $\eta_P = \pm 1$ are allowed. These two possibilities are used to categorize two different types of spin-0 fields. As spin-0 field with $\eta_P = 1$ is called a scalar, which is also sometimes used for a generic spin-0 field. In contrast, a spin-0 field with $\eta_P = -1$ is called a **pseudoscalar**. Note that the Klein-Gordon Lagrangian applies to both scalar and pseudoscalar fields.

Next, we consider spin-1 fields A_μ. Each of the four components of this field is multiplied by an η_P-like factor under parity. Because acting with parity twice should be equivalent to doing nothing we find that each of these is ± 1. But how can related the factors of different components? Let us consider how the Lagrangian is affected by parity transformations. In doing so, we note that spatial derivatives change sign under parity but time derivatives do not

$$\partial_t \xrightarrow{P} \partial_t \; , \qquad \partial_i \xrightarrow{P} -\partial_i \; . \tag{15.13}$$

Exercise 15.1

Show that the kinetic part of the spin-1 Lagrangian

$$\begin{aligned} -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} &= -\frac{1}{2} \partial^\mu A^\nu \left(\partial_\mu A_\nu - \partial_\nu A_\mu \right) \\ &= -\frac{1}{2} \left[\partial^0 A^i \left(\partial_0 A_i - \partial_i A_0 \right) + \partial^i A^0 \left(\partial_i A_0 - \partial_0 A_i \right) + \partial^i A^j \left(\partial_i A_j - \partial_j A_i \right) \right] \; , \end{aligned} \tag{15.14}$$

is preserved if under parity

$$A^0 \xrightarrow{P} \eta_P A^0 \; , \qquad A^i \xrightarrow{P} -\eta_P A^i \; , \tag{15.15}$$

with $\eta_P^2 = 1$.

The above exercise shows that the time and spatial components of

the four-vector must transform oppositely under parity in order for the Lagrangian to be invariant. Spin-1 fields with $\eta_P = 1$ are called vectors, while those with $\eta_P = -1$ are called axial-vectors. The spatial components of axial-vectors do not change under parity. This is similar to quantities like the angular momentum

$$\vec{L} = \vec{r} \times \vec{p}\,, \tag{15.16}$$

which do not change sign under parity.

Finally, we turn to the spin-1/2 field ψ. Just like with the spin-1 field we find that not every component of the field transforms in the same way, because the Lagrangian would not be invariant. If we suppose that $\psi \xrightarrow{P} \eta_P \psi$ then the Lagrangian transforms like

$$\begin{aligned}\mathcal{L} =& \overline{\psi}\left(i\partial\!\!\!/ - m\right)\psi \\ =& \overline{\psi}\left(i\gamma^0\partial_0 + i\gamma^i\partial_i - m\right)\psi \\ \xrightarrow{P}& \overline{\psi}\eta_P^\dagger\left(i\gamma^0\partial_0 - i\gamma^i\partial_i - m\right)\eta_P\psi\,.\end{aligned} \tag{15.17}$$

If η_P is just a number then there is no choice that corresponds to an invariant Lagrangian. So, instead of a simple number we take η_P to be a matrix in the spinor space.

For a matrix η_P we have that $\psi^\dagger \xrightarrow{P} \psi^\dagger\eta_P^\dagger$. This means, using $\gamma^0\gamma^0 = \mathbb{I}_4$, that

$$\overline{\psi} = \psi^\dagger\gamma^0 \xrightarrow{P} \psi^\dagger\eta_P^\dagger\gamma^0 = \psi^\dagger\gamma^0\gamma^0\eta_P^\dagger\gamma^0 = \overline{\psi}\gamma^0\eta_P^\dagger\gamma^0\,. \tag{15.18}$$

We are now ready to consider how the Dirac Lagrangian transforms. We find

$$\begin{aligned}\mathcal{L} =& \overline{\psi}\left(i\gamma^0\partial_0 + i\gamma^i\partial_i - m\right)\psi \\ \xrightarrow{P}& \overline{\psi}\gamma^0\eta_P^\dagger\gamma^0\left(i\gamma^0\partial_0 - i\gamma^i\partial_i - m\right)\eta_P\psi\,.\end{aligned} \tag{15.19}$$

We find three conditions for this last result to be equal to the original Lagrangian:

$$\gamma^0\eta_P^\dagger\gamma^0\gamma^0\eta_P = \gamma^0\,,\quad \gamma^0\eta_P^\dagger\gamma^0\eta_P = \mathbb{I}_4\,,\quad \gamma^0\eta_P^\dagger\gamma^0\gamma^i\eta_P = -\gamma^i\,. \tag{15.20}$$

Exercise 15.2 Show that each of the relations in Eq. (15.20) is satisfied for $\eta_P = \pm\gamma^0$, at least in the Weyl basis of the gamma matrices.

From the above exercise we find that under parity a Dirac spinor transforms like $\psi \xrightarrow{P} \pm\gamma^0\psi$ and so $\overline{\psi} \xrightarrow{P} \pm\overline{\psi}\gamma^0$. In this case the differences in sign are not given different names. This is because we never see a field by itself in a Lagrangian, it is alway part of a **fermion bilinear**. Something that is bilinear is linear in two things. For instance, the mass term involves

$\overline{\psi}\psi$ which is linear in ψ and $\overline{\psi}$. Under parity this goes to

$$\overline{\psi}\psi \xrightarrow{P} P\overline{\psi}\psi P^\dagger = P\overline{\psi}P^\dagger P\psi P^\dagger = (\pm)^2\overline{\psi}\gamma^0\gamma^0\psi = \overline{\psi}\psi \ . \tag{15.21}$$

Here in the second equality we have used the fact that for unitary operators $P^\dagger P = \mathbb{I}$. Because this bilinear does not change sign under parity is called a scalar bilinear. Contrast that with

$$\begin{aligned}\overline{\psi}\gamma_5\psi \xrightarrow{P} & P\overline{\psi}\gamma_5\psi P^\dagger = P\overline{\psi}P^\dagger\gamma_5 P\psi P^\dagger \\ & =(\pm)^2\overline{\psi}\gamma^0\gamma_5\gamma^0\psi = -\overline{\psi}\gamma_5\gamma^0\gamma^0\psi = -\overline{\psi}\gamma_5\psi \ .\end{aligned} \tag{15.22}$$

Note that here the operator P passes through γ_5 because it is a constant matrix. It is only after the P operator acts on the field ψ that a spinor matrix η_P is produced. Because $\overline{\psi}\gamma_5\psi$ does change sign under parity it is called a pseudoscalar bilinear. Similarly, one can show that

$$\overline{\psi}\gamma^0\psi \xrightarrow{P} \overline{\psi}\gamma^0\psi \ , \quad \overline{\psi}\gamma^i\psi \xrightarrow{P} -\overline{\psi}\gamma^i\psi \ . \tag{15.23}$$

This implies that $\overline{\psi}\gamma^\mu\psi$ transforms like vector, with the time component not changing and the spatial components changing sign. Consequently, $\overline{\psi}\gamma^\mu\psi$ is called a vector bilinear.

Exercise 15.3 Show that $\overline{\psi}\gamma^\mu\gamma_5\psi$ transforms like an axial-vector under parity.

In the Weyl basis we can also determine the effect of a parity transformation on the chiral fermions. The parity transformation is then

$$\psi = \begin{pmatrix}\psi_L \\ \psi_R\end{pmatrix} \xrightarrow{P} \pm\gamma^0\psi = \pm\begin{pmatrix}\psi_R \\ \psi_L\end{pmatrix} \ . \tag{15.24}$$

This shows that under parity left-handed fields get taken to right-handed fields and vice versa. Note that this implies that the Lagrangian for a single chirality

$$\mathcal{L} = i\psi_L^\dagger\left(\mathbb{I}_2\partial_t - \sigma_i\partial_i\right)\psi_L \ , \tag{15.25}$$

is not invariant under parity. In other words, parity is not a symmetry of the theory of a single massless fermion.

Subsection 15.2

Charge Conjugation

The classical electrostatic force between two objects with electric charges q_1 and q_2 is (in Heavyside-Lorentz units)

$$\vec{F} = \frac{e^2 q_1 q_2}{4\pi r^3}\vec{r} \ . \tag{15.26}$$

Note that this force remains the same if $q_1 \to -q_1$ and $q_2 \to -q_2$. That is, it looks like at least some theories are symmetric under all charges are changed to their opposite. This is the essence of charge conjugation.

But what does it mean to change the charge of a field, or what is the field configuration with opposite charge? Recall from Eq. (8.34) that the conserved particle number for a complex scalar field is

$$Q = i \int d^3x \left(\phi^\dagger \partial_t \phi - \phi \partial_t \phi^\dagger \right) \ . \tag{15.27}$$

From this formula it is simple to see that if we take $\phi \to \phi^\dagger$ and $\phi^\dagger \to \phi$ that $Q \to -Q$. Therefore, we can change the charge of a field configuration by switching to the Hermitian conjugate of the field.

More generally, we expect the action of the charge conjugation operator C to be proportional to the Hermitian conjugate field

$$C\phi C^\dagger = \eta_C \phi^\dagger \ . \tag{15.28}$$

By taking the Hermitian conjugate of this equation we find

$$\left(C\phi C^\dagger \right)^\dagger = \left(C^\dagger \right)^\dagger \phi^\dagger C^\dagger = C\phi^\dagger C^\dagger = \eta_C^* \phi \ . \tag{15.29}$$

Like parity, the action of charge conjugation twice should be equivalent to doing nothing. This leads us to evaluate

$$CC\phi C^\dagger C^\dagger = C\eta_C \phi^\dagger C^\dagger = \eta_C C\phi^\dagger C^\dagger = \eta_C \eta_C^* \phi = |\eta_C|^2 \phi \ . \tag{15.30}$$

In order to have this equal to ϕ we see that $\eta_C = e^{i\theta_C}$ for some real number θ_C.

For a real scalar field $\phi = \phi^\dagger$ so $\eta_C = \pm 1$. In fact, there is nothing that changes in this analysis for a real or complex spin-1 field. For instance, a real vector field A_μ must satisfy

$$CA_\mu C^\dagger = \pm A_\mu \ . \tag{15.31}$$

It is straightforward to see that the Klein-Gordon Lagrangian and spin-1 Lagrangians are preserved by these transformations.

As usual, things are a little more complicated for spin-1/2 fields. In fact, we need to include something that was mentioned when we discussed the Dirac Lagrangian in Sec. 9.1. Unlike scalar fields and the components of vector fields, the components of spinor fields anticommute. That is

$$\psi_1^A \psi_2^B = -\psi_2^B \psi_1^A \ . \tag{15.32}$$

The ultimate justification for why this is springs from quantum field theory, and beyond the scope of this text. However, we did find in Sec. 9.1 that

anticommuting spinor fields ensured that the energy was bounded from below and that the momentum of antiparticles matched experiment.

A second complication is that unlike spin-0 and spin-1 fields taking the Hermitian conjugate of a spinor changes what type of object is being considered. That is, a spinor field ψ is typically written with index up ψ^A. But this means that $\psi^\dagger$ has its index down ψ^*_A. This means that

$$C\psi^A C^\dagger \neq \eta_C \psi^*_B \ . \tag{15.33}$$

We must be a little more careful to ensure that C, which operates in the infinite dimensional Hilbert space of quantum states, preserves the spinor index structure of the spin-1/2 field.

Our starting point is that under charge conjugation $\psi \to \eta_C \psi^*$ where η_C is a matrix in spinor space, similar to what we considered for parity. In other words,

$$C\psi^A C^\dagger = (\eta_C)^A{}_B \, \psi^{*B} \ . \tag{15.34}$$

Because it is tedious to keep track of all the spinor indices we typically suppress them as

$$C\psi C^\dagger = \eta_C \psi^* \ . \tag{15.35}$$

Taking the Hermitian conjugate of this equation we find

$$C\psi^\dagger C^\dagger = \psi^T \eta_C^\dagger \ . \tag{15.36}$$

This, in turn, leads to

$$\overline{\psi} \xrightarrow{C} \psi^T \eta_C^\dagger \gamma^0 \ . \tag{15.37}$$

We are now ready to consider the transformation of the Dirac Lagrangian. We begin with the mass term $m\overline{\psi}\psi$. Here the mass, as a real constant, has no effect. We find that the bilinear of fermions transforms as

$$\overline{\psi}\psi \xrightarrow{C} \psi^T \eta_C^\dagger \gamma^0 \eta_C \psi^* \ . \tag{15.38}$$

This does not look particularly helpful. We improve things by recognizing that these objects are both numbers in spinor space, and that the transpose of a number is equal to itself. This means by taking the transpose of these products of spinors and matrices we may be able to write this mess in something more like what we started with. However, in doing so we end up moving fermion fields past one another and because the components of spinors anticommute this produces a minus sign difference.

Therefore we have

$$\overline{\psi}\psi \xrightarrow{C} \psi^T \eta_C^\dagger \gamma^0 \eta_C \psi^* = \left(\psi^T \eta_C^\dagger \gamma^0 \eta_C \psi^*\right)^T = (-1)\psi^\dagger \eta_C^T (\gamma^0)^T \eta_C^* \psi \ . \tag{15.39}$$

Here the minus sign comes from anticommuting the spinor fields. In the Weyl basis we have that γ^0 is symmetric, so $(\gamma^0)^T = \gamma^0$. Putting this all

together we find

$$\overline{\psi}\psi \xrightarrow{C} (-1)\overline{\psi}\gamma^0\eta_C^T\gamma^0\eta_C^*\psi \ . \tag{15.40}$$

For this term in the Lagrangian to come back to itself we require

$$\gamma^0\eta_C^T\gamma^0\eta_C^* = -\mathbb{I}_4 \tag{15.41}$$

which is (by multiplying by γ^0 on the left of both sides of the equation and taking the transpose) equivalent to

$$\gamma^0 = -\eta_C^\dagger\gamma^0\eta_C \ . \tag{15.42}$$

Exercise 15.4 Show that Eq. (15.42) is satisfied for

$$\eta_C = -i\gamma^2 \ , \tag{15.43}$$

for γ^2 in the Weyl basis. Conclude that in this basis, under charge conjugation we have

$$\psi \xrightarrow{C} -i\gamma^2\psi^* \ , \qquad \overline{\psi} \xrightarrow{C} i\psi^T\gamma^0\gamma^2 \ . \tag{15.44}$$

Example By construction we have ensured that $\overline{\psi}\psi \to \overline{\psi}\psi$ under charge conjugation. However, it is useful to see how to calculate the transformation of other fermion bilinears. For instance, we find

$$\begin{aligned}\overline{\psi}\gamma^5\psi \xrightarrow{C}&(-i)i\psi^T\gamma^0\gamma^2\gamma_5\gamma^2\psi^* = (-1)\psi^T\gamma^0\gamma_5\gamma^2\gamma^2\psi^* = (-1)\psi^T\gamma^0\gamma_5(-\mathbb{I}_4)\psi^* \\ &= \left(\psi^T\gamma^0\gamma_5\psi^*\right)^T = -\psi^\dagger(\gamma_5)^T(\gamma^0)^T\psi = \psi^\dagger\gamma^0\gamma_5\psi \\ &= \overline{\psi}\gamma_5\psi \ .\end{aligned} \tag{15.45}$$

Here we have used the fact that γ_5 is symmetric in the Weyl basis and that it anticommutes with all other gamma matrices. We have also used that the components of the spinor field anticommute. All in all we have found that this bilinear is also invariant under charge conjugation.

Exercise 15.5 Show that under charge conjugation

$$\overline{\psi}\gamma^\mu\psi \xrightarrow{C} -\overline{\psi}\gamma^\mu\psi \ , \qquad \overline{\psi}\gamma^\mu\gamma_5\psi \xrightarrow{C} \overline{\psi}\gamma^\mu\gamma_5\psi \ . \tag{15.46}$$

The results of the above exercise imply that under charge conjugation

$$\overline{\psi}i\gamma^\mu\partial_\mu\psi \xrightarrow{C} -\partial_\mu(\overline{\psi})i\gamma^\mu\psi \ . \tag{15.47}$$

In the Dirac Lagrangian we can then integrate by parts to show that the derivative term is invariant under charge conjugation. This finishes our

demonstration that the Dirac equation is invariant under charge conjugation. We can also use these results to learn something about QED. In order for QED to be invariant under charge conjugation we need

$$\overline{\psi}\gamma^\mu\psi A_\mu \xrightarrow{C} \overline{\psi}\gamma^\mu\psi A_\mu \ . \tag{15.48}$$

Because $\overline{\psi}\gamma^\mu\psi \xrightarrow{C} -\overline{\psi}\gamma^\mu\psi$ the only way to ensure that QED is invariant under charge conjugation is for the photon field to satisfy $A_\mu \xrightarrow{C} -A_\mu$.

The photon field being odd under charge conjugation has physical consequences. Any QED process that only involves external photons is zero if the number of photons is odd. Such a process would changes sign under charge conjugation, but QED as a theory is invariant under charge conjugation. This result is typically known as Furry's theorem.

Under charge conjugation the left- and right-handed chiral projections transform as

$$\psi = \begin{pmatrix} \psi_L \\ \psi_R \end{pmatrix} \xrightarrow{C} -i\gamma^2\psi^* = \begin{pmatrix} -i\sigma_2\psi_R^* \\ i\sigma_2\psi_L^* \end{pmatrix} \ . \tag{15.49}$$

As we saw with parity, this shows that the Lagrangian for a single chirality, such as

$$\mathcal{L} = i\psi_L^\dagger \left(\mathbb{I}_2 - \sigma_i\partial_i\right)\psi_L \ , \tag{15.50}$$

is not invariant under charge conjugation. However, if we act with both parity and charge conjugation then

$$\psi = \begin{pmatrix} \psi_L \\ \psi_R \end{pmatrix} \xrightarrow{CP} \mp\gamma^0 i\gamma^2\psi^* = \mp \begin{pmatrix} i\sigma_2\psi_L^* \\ -i\sigma_2\psi_R^* \end{pmatrix} \ . \tag{15.51}$$

Exercise 15.6

Show that the Lagrangian for a left-handed Weyl spinor, Eq. (15.50), is invariant under the CP transformation given above

$$\psi_L \xrightarrow{CP} \mp i\sigma_2\psi_L^* \ . \tag{15.52}$$

You may need to integrate by parts to show this invariance.

The above exercise shows that massless fermion Lagrangians may not be invariant under C and P separately, but are invariant under the combined CP transformation. It is in this sense that CP is a more fundamental symmetry operation.

We have seen that QED is CP invariant, but not every model is. Consider the Lagrangian interactions between two fermions f_1, f_2 and a real scalar ϕ

$$\lambda\phi\overline{f}_1 f_2 + \lambda^*\phi\overline{f}_2 f_1 \ . \tag{15.53}$$

Note that we must include both terms to ensure that the Lagrangian is

Hermitian. Under CP the scalar transforms as

$$\phi \xrightarrow{CP} \pm\phi \ . \tag{15.54}$$

The fermions transform as

$$f_i \xrightarrow{CP} \mp i\gamma^0\gamma^2 f_i^* \ , \quad \overline{f}_i \xrightarrow{CP} \pm i f^T\gamma^0\gamma^2\gamma^0 \ . \tag{15.55}$$

This leads to

$$\lambda\phi\overline{f}_1 f_2 \xrightarrow{CP} \pm\lambda\phi\overline{f}_2 f_1 \ . \tag{15.56}$$

This leads to

$$\lambda\phi\overline{f}_1 f_2 + \lambda^*\phi\overline{f}_2 f_1 \xrightarrow{CP} \pm\lambda^*\phi\overline{f}_1 f_2 \pm \lambda\phi\overline{f}_2 f_1 \ . \tag{15.57}$$

Which shows that this interaction term is only CP invariant if ϕ is a scalar (rather than a pseudoscalar) and λ is a real coupling.

Subsection 15.3

Time Reversal

Much of the time-reversal analysis is similar to the previous sections, but there is one big difference. While parity and charge conjugation are represented by unitary operators, time reversal cannot be.

Consider two quantum states $|\text{initial}\rangle$ and $|\text{final}\rangle$ that correspond to some initial configuration in time and some final configuration in time. We are typically interested in the overlap of these two states

$$\langle\text{final}|\text{initial}\rangle. \tag{15.58}$$

Time reversal T should take these to $U|\text{final}\rangle$ and $U|\text{initial}\rangle$ such that

$$\langle\text{final}|\text{initial}\rangle \xrightarrow{T} \langle\text{initial}|\text{final}\rangle = \langle\text{final}|\text{initial}\rangle^* \ , \tag{15.59}$$

or

$$\langle\text{final}|\text{initial}\rangle \xrightarrow{T} \langle\text{final}|U^\dagger U|\text{initial}\rangle = \langle\text{final}|\text{initial}\rangle^* \ . \tag{15.60}$$

This last equality is what defines an **antiunitary** operator.

Eugene Wigner proved that quantum mechanical symmetries can either be represented by an operator that is unitary and linear, or antiunitary and antilinear. An antilinear operator is one for which

$$U\left(\alpha|\psi_1\rangle + \beta|\psi_2\rangle\right) = \alpha^* U|\psi_1\rangle + \beta^* U|\psi_2\rangle \ . \tag{15.61}$$

In other words, when a constant is "pulled through" an antilinear operator it must be complex conjugated.

Antilinearity produces an important difference in the analysis of how field fields transform, including scalar fields. If we begin with

$$T\phi T^{-1} = \eta_T \phi \ , \tag{15.62}$$

for some constant η_T then acting with time reversal twice leads to

$$TT\phi T^{-1}T^{-1} = T\eta_T \phi T^{-1} = \eta_T^* T\phi T^{-1} = |\eta_T|^2 \phi \ . \tag{15.63}$$

Since we expect reversing time twice is equivalent to doing nothing at all our constraint is that η_T is a phase

$$\eta_T = e^{-i\theta_T} \ . \tag{15.64}$$

Recall that for parity a similar analysis led to $\eta_P = \pm 1$. In the case of time reversal we find that $\eta_T = \pm 1$ only for real scalars.

The analysis of spin-1 fields is quite similar to parity. Under time reversal derivatives transform as

$$\partial_t \xrightarrow{T} -\partial_t \ , \qquad \partial_i \xrightarrow{T} \partial_i \ . \tag{15.65}$$

This is opposite to parity, but leads to very similar results.

Exercise 15.7

Show that the kinetic part of the spin-1 Lagrangian

$$\begin{aligned} -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} &= -\frac{1}{2}\partial^\mu A^\nu \left(\partial_\mu A_\nu - \partial_\nu A_\mu\right) \\ &= -\frac{1}{2}\left[\partial^0 A^i \left(\partial_0 A_i - \partial_i A_0\right) + \partial^i A^0 \left(\partial_i A_0 - \partial_0 A_i\right) + \partial^i A^j \left(\partial_i A_j - \partial_j A_i\right)\right] \ , \end{aligned} \tag{15.66}$$

is preserved if under time reversal

$$A^0 \xrightarrow{T} -\eta_T A^0 \ , \qquad A^i \xrightarrow{T} \eta_T A^i \ , \tag{15.67}$$

with $\eta_T^2 = 1$.

As with other discrete symmetries, the more tedious case is for spin-1/2 fields. We begin with

$$\psi \xrightarrow{T} \eta_T \psi \ , \qquad \overline{\psi} \xrightarrow{T} \overline{\psi}\gamma^0 \eta_T^\dagger \gamma^0 \ , \tag{15.68}$$

where η_T is take to be a matrix in the four-dimensional spinor space. In the second relation we have used the fact that $\gamma^0\gamma^0 = \mathbb{I}_4$. We can then consider the action of time reversal on the Dirac Lagrangian. In doing so we must be careful to take the complex conjugate of all constants, like

gamma matrices

$$\begin{aligned}\mathcal{L} =&\overline{\psi}\left(i\gamma^0\partial_t + i\gamma^i\partial_i - m\right)\psi \\ \xrightarrow{T}&\overline{\psi}\gamma^0\eta_T^\dagger\gamma^0\left[-i(\gamma^0)^*\partial_{-t} - i(\gamma^i)^*\partial_i - m\right]\eta_T\psi \\ =&\overline{\psi}\gamma^0\eta_T^\dagger\gamma^0\left[i\gamma^0\partial_t - i(\gamma^i)^*\partial_i - m\right]\eta_T\psi\ . \end{aligned} \tag{15.69}$$

In the last line we have used the fact that γ^0 is a real matrix in the Weyl basis.

Exercise 15.8 Show that in the Weyl basis of gamma matrices that

$$\eta_T = \gamma^1\gamma^3\ , \tag{15.70}$$

ensures that the Dirac Lagrangian is invariant under time reversal.

Exercise 15.9 Determine the behavior of the fermion bilinears under time reversal

$$\overline{\psi}\psi \xrightarrow{T}\overline{\psi}\psi\ , \qquad \overline{\psi}\gamma_5\psi \xrightarrow{T}\overline{\psi}\gamma_5\psi\ , \tag{15.71}$$

$$\overline{\psi}\gamma^0\psi \xrightarrow{T}\overline{\psi}\gamma^0\psi\ , \qquad \overline{\psi}\gamma^i\psi \xrightarrow{T} -\overline{\psi}\gamma^i\psi\ , \tag{15.72}$$

$$\overline{\psi}\gamma^0\gamma_5\psi \xrightarrow{T}\overline{\psi}\gamma^0\gamma_5\psi\ , \qquad \overline{\psi}\gamma^i\gamma_5\psi \xrightarrow{T} -\overline{\psi}\gamma^i\gamma_5\psi\ . \tag{15.73}$$

The combination of parity, charge conjugation, and time reversal or CPT (the order does not matter) is a symmetry of all known models of physical systems. It has been proven that Lorentz invariant Lagrangians that produce Hermitian Hamiltonians are CPT symmetric. Looking back at these last few sections it is clear that the fermion bilinears have the most complicated transformation laws of the fields we have considered. Yet, it is not hard to check that each bilinear is invariant under CPT.

Exercise 15.10 Show that each of the fermion bilinears considered in this section is invariant under the combined CPT transformation.

PART

IV

Quantum Chromodynamics

Quantum electrodynamics is a remarkable theory. It has amazing agreement with experimental predictions and although it is in many respects quite simple, yet it is quite deep. Quantum chromodynamics (QCD) builds upon that foundation and predicts new phenomena regarding the strong nuclear force. There is much that we are not able to cover about this theory in the short introduction below, but hopefully it gives you a strong starting point for further study and understanding.

Sec. 16. Nonabelian
Sec. 17. SU(2)
Sec. 18. QCD

Section 16

Nonabelian Local Symmetry

Quantum electrodynamics is known as a U(1) gauge theory or sometimes as an abelian gauge theory. The structure of the Lagrangian

$$\mathcal{L}_{\text{QED}} = \overline{\psi}\left(i\not{D} - m\right)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{16.1}$$

with $D_\mu = \partial_\mu + ieqA_\mu$ is invariant under a local U(1) symmetry. The transformations are given by

$$\psi \to e^{-iq\theta(x)}\psi \quad A_\mu \to A_\mu - \frac{1}{e}\partial_\mu\theta(x) \ , \tag{16.2}$$

where in general $\theta(x)$ depends on the three dimensions of space and on time. The is dependence on spacetime is why this is referred to as a local symmetry. It is referred to as a U(1) symmetry because every $e^{-i\theta}$ is an element of the group U(1). In other words, the action of the local symmetry is an element of the group U(1).

The Lie group U(1) is abelian, it does not matter what order group elements are multiplied in $e^{-i\theta_1}e^{-i\theta_2} = e^{-i\theta_2}e^{-i\theta_1}$. However, most Lie groups G are nonabelian, changing the order of multiplication changes the result $U_1U_2 \neq U_2U_1$. The group elements can also be made local, functions of space and time $U(x)$. This leads to the question, can a QED-like theory be generalized to a local, nonabelian symmetry group.

We begin by considering a field ψ (we continue to consider a spin-1/2 field though this is not essential) that transforms by the action of the elements U of some group G

$$\psi \to U\psi \ . \tag{16.3}$$

In effect, this means that ψ is the element of a vector space and U is given by a unitary representation of the group element in the vector space. In this section we use lower case Latin indices to denote the components of ψ^i in this vector space. For instance, quantum chromodynamics is tied to an SU(3) symmetry group. Fields in the fundamental representation have the form

$$\psi(x) = \begin{pmatrix} \psi^1(x) \\ \psi^2(x) \\ \psi^3(x) \end{pmatrix} \ . \tag{16.4}$$

Note that for a spin-1/2 field, like a quark, each of these three components is a four-dimensional Dirac spinor. For these types of fields the group elements U are three-by-three unitary matrices with determinant equal to one.

In general we express the symmetry transformation with index notation

$$\psi^i \to U^i{}_j(x)\psi^j \ . \tag{16.5}$$

What about $\overline{\psi}$? The Hermitian conjugation process means that if ψ^i has its index up (like a column vector) then $(\psi^\dagger)_i$ has index down, like a row vector. It transforms in the representation conjugate to the representation of ψ^i. This leads to

$$\overline{\psi}_i \to \overline{\psi}_j U^{\dagger j}{}_i(x) \ . \tag{16.6}$$

Note here that the γ^0 related to $\overline{\psi}$ has no effect upon this transformation because the four-dimensional spinor space is completely distinct from the symmetry space. In what follows we do not write the explicit dependence of U on the coordinates but these should be taken to be local symmetry transformations.

Example

We can now evaluate how $\overline{\psi}\psi$ transforms under local symmetry transformations of the type outlined above. We find

$$\overline{\psi}_i\psi^i \to \overline{\psi}_j U^{\dagger j}{}_i U^i{}_k \psi^k \ . \tag{16.7}$$

Because U is a unitary operator it is true that

$$U^{\dagger j}{}_i U^i{}_k = \delta^j_k \ . \tag{16.8}$$

This implies that

$$\overline{\psi}_i\psi^i \to \overline{\psi}_j \delta^j_k \psi^k = \overline{\psi}_i\psi^i \ . \tag{16.9}$$

Therefore, this inner product in the symmetry space is invariant under symmetry transformations.

The above example shows that the mass term $m\overline{\psi}_i\psi^i$ in the Dirac Lagrangian is invariant under global or local symmetry transformations of general Lie groups. As with QED the derivative term is more subtle. The derivative of the field transforms like

$$\partial_\mu\psi^i \to \partial_\mu\left(U^i{}_j\psi^j\right) = U^i{}_j\partial_\mu\psi^j + \partial_\mu\left(U^i{}_j\right)\psi^j \ . \tag{16.10}$$

We need to construct a gauge covariant derivative $(D_\mu)^i{}_j\,\psi^j$ such that when $\psi^i \to U^i{}_j\psi^j$ that

$$(D_\mu\psi)^i \to U^i{}_j\,(D_\mu\psi)^j \ . \tag{16.11}$$

As with the covariant derivative in QED we need the covariant derivative of the field to transform like the field itself. Taking motivation from QED

we consider operators of the form

$$(D_\mu)^i{}_j = \mathbb{I}^i{}_j \partial_\mu + ig\,(A_\mu)^i{}_j \ . \tag{16.12}$$

Here we have introduced a matrix of photon-like fields $(A_\mu)^i{}_j$. In QED these fields were real. In a more general gauge theory this matrix of fields must be Hermitian.

With this starting point we determine how this covariant derivative transforms. We find

$$\begin{aligned}(D_\mu\psi)^i &= \left[\mathbb{I}^i{}_j\partial_\mu + ig(A_\mu)^i{}_j\right]\psi^j \\ &\to \left[\mathbb{I}^i{}_j\partial_\mu + ig(A_\mu)^i{}_j\right]U^j{}_k\psi^k \\ &= \left[U^i{}_k\partial_\mu + \partial_\mu(U^i{}_k) + ig(A_\mu)^i{}_jU^j{}_k\right]\psi^k \\ &= U^i{}_\ell\left[\mathbb{I}^\ell{}_k\partial_\mu + U^{-1\ell}{}_j\partial_\mu(U^j{}_k) + igU^{-1\ell}{}_m(A_\mu)^m{}_jU^j{}_k\right]\psi^k \ .\end{aligned} \tag{16.13}$$

Or, suppressing all the symmetry space indices

$$D_\mu\psi \to U\left[\partial_\mu + U^{-1}\partial_\mu(U) + igU^{-1}A_\mu U\right]\psi \ . \tag{16.14}$$

If we want this to equal $\left(\partial_\mu + igA'_\mu\right)\psi'$ then the gauge fields must transform like

$$A_\mu = U^{-1}A'_\mu U - \frac{i}{g}U^{-1}\partial_\mu U \ . \tag{16.15}$$

Exercise 16.1 From Eq. (16.15) show that

$$A'_\mu = UA_\mu U^{-1} + \frac{i}{g}\partial_\mu(U)U^{-1} \ . \tag{16.16}$$

We would like to see Eq. (16.16) as a generalization of the gauge transformations of the QED photon. This can be done most easily if we use the fact that our symmetry group element $U(x)$ is a representation of an element of a Lie group. This means we can write

$$U(x) = e^{-i\theta^a(x)T^a} \ , \tag{16.17}$$

where the T^a are basis elements of of the associated Lie algebra. Note that we take the standard practice of repeated indices implying sums. In this case they are sums over the algebra indices. The metric (or inner product) on this space is real and diagonal (at least for the groups we focus on), so we keep all the indices up for simplicity. The basis elements satisfy the commutation relations

$$\left[T^a, T^b\right] = if^{abc}T^c \ , \tag{16.18}$$

for structure constants f^{abc}. These can always be chosen to be completely

antisymmetric. For unitary U the T^a are Hermitian. If the U have determinant equal to one (as is the case for the SU(3) symmetry of QCD) then the T^a are traceless. This follows from the identity

$$\mathrm{Det}\left(e^M\right) = e^{\mathrm{Tr}(M)} \ . \tag{16.19}$$

Exercise 16.2 Using the connection between Lie group and Lie algebra given by Eq. (16.17) show that for $\theta^a \ll 1$ that Eq. (16.15) becomes

$$A'_\mu = A_\mu - i\theta^a \left[T^a,\, A_\mu\right] + \frac{1}{g} T^a \partial_\mu \theta^a + \mathcal{O}(\theta^2) \ . \tag{16.20}$$

The above exercise shows that for small θ^a the A_μ transformation law given in Eq. (16.15) can be written as

$$A_\mu \to A_\mu - i\theta^a \left[T^a,\, A_\mu\right] + \frac{1}{g} T^a \partial_\mu \theta^a + \mathcal{O}(\theta^2) \ . \tag{16.21}$$

Remember that we have suppressed the symmetry space indices, but A_μ is a matrix in that space. In the abelian case (QED) the gauge field is not a matrix and so the commutator term is zero. Therefore, we can see that (16.21) as a generalization of the photon transformation from QED

$$A_\mu \to A_\mu + \frac{1}{e} \partial_\mu \theta(x) \ . \tag{16.22}$$

The implication of this nonabelian gauge field transformation law is that for a *global* symmetry transformation (the thetas all constant) we have

$$A_\mu \to e^{-i\theta^a T^a} A_\mu e^{i\theta^a T^a} \ . \tag{16.23}$$

This is the transformation law for an object in the adjoint representation of the Lie group, as defined in Eq. (39.22). The adjoint representation of a group is always defined because it is the representation on the vector space in which the Lie algebra itself is defined. This means that we can express the matrix $\left(A_\mu\right)^i{}_j$ in terms of the basis matrices $\left(T^a\right)^i{}_j$ of the Lie algebra, because these are a basis for the vector space Lie algebra is defined in. More concisely, we have

$$\left(A_\mu\right)^i{}_j = A^a_\mu \left(T^a\right)^i{}_j \ , \tag{16.24}$$

where the A^a_μ are not matrices in the symmetry space. This is useful because it allows us to separate the Lorentz spacetime index A^a_μ from the symmetry space indices $\left(T^a\right)^i{}_j$. Typically these symmetry space indices are suppressed: $A_\mu = A^a_\mu T^a$.

Exercise 16.3 Using the fact that $UU^{-1} = \mathbb{I}$ show that

$$\partial_\mu U^{-1} = -U^{-1}\partial_\mu(U)U^{-1} \ . \tag{16.25}$$

Then, use this result to show that the transformation law for $A_\mu^\dagger$ is

$$A_\mu^\dagger \to U A_\mu^\dagger U^{-1} + \frac{i}{g}\partial_\mu(U)U^{-1} \ . \tag{16.26}$$

This is the same transformation law as A_μ. Therefore, if we begin with a Hermitian A_μ then gauge transformations are to another Hermitian matrix.

Thus far, we have determined how to make the

$$\overline{\psi}\left(i\not{D} - m\right)\psi \ , \tag{16.27}$$

part of the Lagrangian invariant under local symmetry transformations. This includes the term

$$g\overline{\psi}T^a\gamma^\mu\psi A_\mu^a \ . \tag{16.28}$$

The precise form of T^a depends on what representation of the symmetry group the field ψ transforms under. It is this terms that gives rise to the interaction shown in Fig. 15, similar to the QED vertex between photons and charged fermions. Note that a different fermion field $\widehat{\psi}$ might couple by way of a different representation which changes how it couples to the gauge field, but every coupling includes the same coupling g. This illustrates that different representations are analogous to different charges q in the QED case.

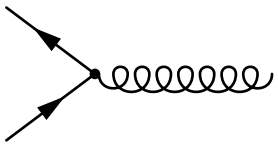


Figure 15. Vertex for QCD interaction between a gauge field and a fermion.

While we now understand the interactions between the gauge fields A_μ and the fermion ψ we have not yet determined the rest of the Lagrangian, the part that determines how the gauge fields propagate through spacetime. In QED this was given by

$$-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} \ , \tag{16.29}$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. But what would the field strength be for more general symmetry transformations. Taking inspiration from Eq. (12.19), we define the field strength tensor in terms of the commutator of the covariant derivatives

$$[D_\mu, D_\nu]\,\psi = igF_{\mu\nu} \ . \tag{16.30}$$

In the present case this field strength is a matrix in the symmetry space.

Exercise 16.4 Using Eq. (16.30) show that

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + ig[A_\mu, A_\nu] \ . \tag{16.31}$$

Then, using the transformation law for A_μ given in Eq. (16.16) show that under local symmetry transformations that

$$F_{\mu\nu} \to U F_{\mu\nu} U^{-1} \ . \tag{16.32}$$

The result of the above exercise shows that under local symmetry transformations that

$$F_{\mu\nu}F^{\mu\nu} \to U F_{\mu\nu} U^{-1} U F^{\mu\nu} U^{-1} = U F_{\mu\nu} F^{\mu\nu} U^{-1} \ . \tag{16.33}$$

This is Lorentz invariant, because all the spacetime indices come in contracted pairs, but not invariant under the local symmetry. What is taken to appear in the Lagrangian is

$$\mathrm{Tr}\left[F_{\mu\nu}F^{\mu\nu}\right] \ . \tag{16.34}$$

Due to the cyclicity of the trace, under local symmetry transformations

$$\begin{aligned} \mathrm{Tr}\left[F_{\mu\nu}F^{\mu\nu}\right] &\to \mathrm{Tr}\left[U F_{\mu\nu} F^{\mu\nu} U^{-1}\right] \\ &= \mathrm{Tr}\left[U^{-1} U F_{\mu\nu} F^{\mu\nu}\right] = \mathrm{Tr}\left[F_{\mu\nu}F^{\mu\nu}\right] \ . \end{aligned} \tag{16.35}$$

This is both Lorentz invariant and invariant under symmetry transformations.

Using $A_\mu = A^a_\mu T^a$ we can separate the spacetime and symmetry indices of the fields strength. We find that

$$\begin{aligned} F_{\mu\nu} =& T^a \left(\partial_\mu A^a_\nu - \partial_\nu A^a_\mu\right) + i g A^b_\mu A^c_\nu [T^b, T^c] \\ =& T^a \left(\partial_\mu A^a_\nu - \partial_\nu A^a_\mu - g f^{bca} A^b_\mu A^c_\nu\right) \ . \end{aligned} \tag{16.36}$$

Because the structure constants are completely antisymmetric we have $f^{bca} = -f^{bac} = f^{abc}$. So, we see that the field strength is in the adjoint representation $F_{\mu\nu} = F^a_{\mu\nu} T^a$ with

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu - g f^{abc} A^b_\mu A^c_\nu \ . \tag{16.37}$$

The typical choice for the Lie algebra basis elements in the fundamental representations is to rescale them such that

$$\mathrm{Tr}\left[T^a T^b\right] = \frac{1}{2}\delta^{ab} \ . \tag{16.38}$$

This leads to the Lagrangian term

$$-\frac{1}{2}\mathrm{Tr}\left[F_{\mu\nu}F^{\mu\nu}\right] = -\frac{1}{2} F^a_{\mu\nu} F^{b\mu\nu} \mathrm{Tr}\left[T^a T^b\right] = -\frac{1}{4} F^a_{\mu\nu} F^{a\mu\nu} \ . \tag{16.39}$$

This is the correct generalization of the QED result. The careful reader may note that this has only been shown for the fundamental representation. It can be shown [32] that this is, in fact, the correct for any representation. The final result may be understood as applying to any representation because it is only defined in terms of the structure constants, which apply to all representations.

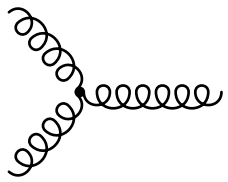


Figure 16. Vertex for three gauge field interactions.

Unlike the QED version of the field strength term in the Lagrangian, when there are nonabelian symmetries there are also interactions among the gauge fields. When we expand this Lagrangian term we find

$$\begin{aligned}\frac{1}{4}F^a_{\mu\nu}F^{a\mu\nu} =& -\frac{1}{4}\left(\partial_\mu A^a_\nu - \partial_\nu A^a_\mu\right)\left(\partial^\mu A^{a\nu} - \partial^\nu A^{a\mu}\right) \\ &+ g f^{abc} A^b_\mu A^c_\nu \partial^\mu A^{a\nu} - \frac{g^2}{4} f^{abc} f^{ade} A^b_\mu A^c_\nu A^{d\mu} A^{e\nu} \end{aligned} \tag{16.40}$$

The terms on the top line look exactly like those of QED and describe the propagation of the gauge fields through spacetime. The terms on the second line, however, are interaction terms between three (see Fig. 16) or four (see Fig. 17) gauge fields. One difference to note is that the term

$$g f^{abc} A^b_\mu A^c_\nu \partial^\mu A^{a\nu} \ , \tag{16.41}$$

is only invariant under parity if the gauge fields are vectors, rather than axial vectors.

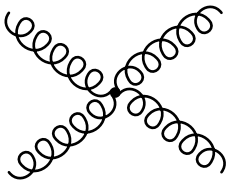


Figure 17. Vertex for four gauge field interactions.

These interactions between gauge fields are the most significant difference between abelian gauge theories like QED and nonabelian gauge theories like QCD. Essentially, these show that while the photon interacts with charged particles it is electrically neutral itself. In contrast, the gluons of QCD are themselves charged under the force that they communicate. This plays a role some of the ways nonabelian gauge theories can be qualitatively different than abelian ones.

Section 17

A Simple SU(2) Example

The previous section covered a lot of ground and was somewhat abstract. In this section we look at the simplest nonabelian gauge theory, one with SU(2) symmetry. As outlined in Sec. 38, the usual basis elements for the Lie algebra $\mathfrak{su}(2)$ are the Pauli matrices divided by two

$$T^1 = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \ , \quad T^2 = \frac{1}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \ , \quad T^3 = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \ . \tag{17.1}$$

The structure constants for the algebra are simply given by the Levi-Civita symbol ε^{abc}.

Exercise 17.1 Show that

$$\mathrm{Tr}\left[T^a\,T^b\right] = \frac{1}{2}\delta^{ab}\ , \tag{17.2}$$

and

$$\left[T^a,\,T^b\right] = i\varepsilon^{abc}T^c\ . \tag{17.3}$$

Suppose that a fermionic field is in the fundamental representation of SU(2)

$$\psi = \begin{pmatrix}\psi^1\\ \psi^2\end{pmatrix}\ . \tag{17.4}$$

Then the interaction of the fermion with the gauge fields is

$$-g\overline{\psi}T^a\gamma^\mu\psi A^a_\mu\ . \tag{17.5}$$

Because there are three basis elements of the algebra there are also three real gauge fields A^1_μ, A^2_μ, and A^3_μ. The matrix valued gauge field is then

$$T^aA^a_\mu = \frac{1}{2}\begin{pmatrix}A^3_\mu & A^1_\mu - iA^2_\mu\\ A^1_\mu + iA^2_\mu & -A^3_\mu\end{pmatrix}\ . \tag{17.6}$$

It is easy to see that this matrix is Hermitian and traceless, as expected.

The commutator appearing in the field strength is found, in the fundamental representation, to be

$$\begin{aligned}&[A_\mu,\,A_\nu] = \\ &\frac{i}{2}\begin{pmatrix}A^1_\mu A^2_\nu - A^1_\nu A^2_\mu & A^2_\mu A^3_\nu - A^2_\nu A^3_\mu - i(A^3_\mu A^1_\nu - A^3_\nu A^1_\mu)\\ A^2_\mu A^3_\nu - A^2_\nu A^3_\mu + i(A^3_\mu A^1_\nu - A^3_\nu A^1_\mu) & -A^1_\mu A^2_\nu + A^1_\nu A^2_\mu\end{pmatrix}\\ &= iT^1\left(A^2_\mu A^3_\nu - A^2_\nu A^3_\mu\right) + iT^2\left(A^3_\mu A^1_\nu - A^3_\nu A^1_\mu\right) + iT^3\left(A^1_\mu A^2_\nu - A^1_\nu A^2_\mu\right)\ .\end{aligned} \tag{17.7}$$

The three components of the field strength matrix, in any representation, are

$$\begin{aligned}F^1_{\mu\nu} =&\partial_\mu A^1_\nu - \partial_\nu A^1_\mu - g\left(A^2_\mu A^3_\nu - A^2_\nu A^3_\mu\right)\ ,\\ F^2_{\mu\nu} =&\partial_\mu A^2_\nu - \partial_\nu A^2_\mu - g\left(A^3_\mu A^1_\nu - A^3_\nu A^1_\mu\right)\ ,\\ F^3_{\mu\nu} =&\partial_\mu A^3_\nu - \partial_\nu A^3_\mu - g\left(A^1_\mu A^2_\nu - A^1_\nu A^2_\mu\right)\ .\end{aligned} \tag{17.8}$$

Suppose some field ϕ transforms in the triplet representation

$$\phi = \begin{pmatrix}\phi^1\\ \phi^2\\ \phi^3\end{pmatrix} \tag{17.9}$$

The three-dimensional representation of SU(2), which is also the adjoint representation, is given in Eq. (39.100). When interacting with a field in

the triplet representation the gauge fields are written as

$$A_\mu^a T^a = \begin{pmatrix} A_\mu^3 & \frac{1}{\sqrt{2}}(A_\mu^1 - iA_\mu^2) & 0 \\ \frac{1}{\sqrt{2}}(A_\mu^1 + iA_\mu^2) & 0 & \frac{1}{\sqrt{2}}(A_\mu^1 - iA_\mu^2) \\ 0 & \frac{1}{\sqrt{2}}(A_\mu^1 + iA_\mu^2) & -A_\mu^3 \end{pmatrix} . \tag{17.10}$$

Fields in larger dimensional representations of the symmetry group couple to the three gauge fields through the larger dimensional matrices appropriate to that representation.

Exercise 17.2 Find the form of $[A_\mu, A_\nu]$ in the triplet representation defined above. Show that the components are of the field strength are still given by Eq. (17.8) in this representation.

The relatively simple structure constants ε^{abc} of SU(2) also offer an opportunity to understand the gauge boson self-interactions in the Lagrangian, see Eq. (16.40). The three gauge-boson interaction has the form

$$g\varepsilon^{abc} A_\mu^b A_\nu^c \partial^\mu A^{a\nu} . \tag{17.11}$$

The complete antisymmetry of the structure constants ensures that only one of each type of boson can appear. For example we could have

$$g\varepsilon^{123} A_\mu^2 A_\nu^3 \partial^\mu A^{1\nu} . \tag{17.12}$$

The four gauge-boson interaction is given by

$$-\frac{g^2}{4}\varepsilon^{abc}\varepsilon^{ade} A_\mu^b A_\nu^c A^{d\mu} A^{e\nu} = -\frac{g^2}{4}\left(\delta^{bd}\delta^{ce} - \delta^{be}\delta^{cd}\right) A_\mu^b A_\nu^c A^{d\mu} A^{e\nu} . \tag{17.13}$$

SECTION 18

QCD from SU(3)

Quantum chromodynamics is a gauge theory based on a local SU(3) symmetry. Some details of this symmetry group are given in Sec. 39.1. The gauge fields are called gluons. As there are eight generators of Lie algebra, or eight basis elements for the vector space the Lie algebra is defined on, there are eight gluon fields. The Lagrangian can be written as

$$\mathcal{L}_{\text{QCD}} = -\frac{1}{4} G_{\mu\nu}^a G^{a\mu\nu} + \sum_q \overline{\psi}_q \left(i\not{D} - m\right) \psi_q , \tag{18.1}$$

where $G_{\mu\nu}^a$ is the field strength tensor composed of the gluon fields and D_μ is the covariant derivative appropriate to whatever representation of SU(3)

the fermion ψ_q transforms under. The index a on $G^a_{\mu\nu}$ runs from one to eight, corresponding to the eight gluon fields.

In addition to the gluons are the quarks fields ψ_q. These come in six "flavors" that have have somewhat whimsical names: up, down, strange, charm, bottom, and top. They are spin-1/2 particles in the fundamental, or **3** representation of SU(3), see Sec. 39.1 for more information of the representations of SU(3). They are sometimes expressed schematically (also somewhat whimsically) as

$$\psi = \begin{pmatrix} r \\ g \\ b \end{pmatrix} \ , \tag{18.2}$$

with r referring to red, g referring to green, and b referring to blue. However, we never speak about the red up quarks, for example, because the symmetry of theory implies that we cannot tell red from blue. The conjugate fields, sometimes loosely called the antiquark fields, $\overline{\psi}$ transform in the $\overline{\mathbf{3}}$ or conjugate representation, see Eq. (39.3) for further discussion, of the fundamental representation of SU(3). In general, ψ and $\overline{\psi}$ are in related but distinct representations.

The combination of a quark and antiquark belong to the $\mathbf{3} \otimes \overline{\mathbf{3}} = \mathbf{1} \oplus \mathbf{8}$ representation. We showed in Eq. (16.9) that the inner product $\overline{\psi}_i \psi^i$ is unchanged by symmetry transformations. This is the singlet **1** part of the product of the two fields.

The other representation **8** is the adjoint representation of SU(3). We can express it as

$$\psi^i \overline{\psi}_j - \frac{1}{3} \delta^i_j \psi^k \overline{\psi}_k \ , \tag{18.3}$$

This is not a form that we typically find in a Lagrangian, but this object has the transformation law

$$\begin{aligned} \psi^i \overline{\psi}_j - \frac{1}{3} \delta^i_j \psi^k \overline{\psi}_k \rightarrow & U^i{}_m \psi^m \overline{\psi}_n U^{\dagger n}{}_j - \frac{1}{3} \delta^i_j U^k{}_m \psi^m \overline{\psi}_n U^{\dagger n}{}_k \\ = & U^i{}_m \psi^m \overline{\psi}_n U^{\dagger n}{}_j - \frac{1}{3} U^i{}_k U^{\dagger k}{}_j \psi^m \overline{\psi}_n U^{\dagger n}{}_k U^k{}_m \\ = & U^i{}_m \psi^m \overline{\psi}_n U^{\dagger n}{}_j - \frac{1}{3} U^i{}_k U^{\dagger \ell}{}_j \delta^k_\ell \psi^m \overline{\psi}_n \delta^n_m \\ = & U^i{}_m \left(\psi^m \overline{\psi}_n - \frac{1}{3} \delta^m_n \psi^k \overline{\psi}_k \right) U^{\dagger n}{}_j \ , \end{aligned} \tag{18.4}$$

of the adjoint representation, just like the gauge fields. This allows us to learn something about the gluons from this combination of fields. It has

the matrix form

$$\begin{pmatrix} r \\ g \\ b \end{pmatrix} \left(\overline{r},\, \overline{g},\, \overline{b}\right) - \frac{1}{3}\left(r\overline{r} + g\overline{g} + b\overline{b}\right) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} r\overline{r} & r\overline{g} & r\overline{b} \\ g\overline{r} & g\overline{g} & g\overline{b} \\ b\overline{r} & b\overline{r} & b\overline{b} \end{pmatrix} - \frac{1}{3}\left(r\overline{r} + g\overline{g} + b\overline{b}\right) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} . \tag{18.5}$$

The significant feature of this matrix is that each term has the form of color times anticolor. So, inasmuch as we think of the quarks as having three different colors the eight gluons are made of color-anticolor combinations. This agrees with our earlier understanding that the gluons carry color charge, but also somewhat extends it.

Subsection 18.1

Confinement

The fact that the gluons are themselves charged under the force they convey is quite significant. One illustration of this significance is how it affects the running of the QCD coupling. Similar to QED, we define

$$\alpha_s \equiv \frac{g_s^2}{4\pi} , \tag{18.6}$$

where g_s is the QCD coupling, or strong coupling. The label strong comes from the connection to the strong nuclear force.

This coupling is a function of energy scale. At leading order in perturbation theory the equation that determines this energy dependence is

$$\alpha_s(Q) = \frac{\alpha_s(Q_0)}{1 + \frac{\alpha_s(Q_0)}{2\pi}\left(11 - \frac{2}{3}n_f\right)\ln\frac{Q}{Q_0}} , \tag{18.7}$$

where n_f is the number of fermion fields in the fundamental representation of SU(3) with mass below Q. This equation is quite similar to the Eq. (14.8), which governs the running of the electric coupling of QED. A particularly salient difference is the factor of 11 that enters with a different sign than the fermion contributions. This factor of 11 comes from the gluon self-interactions, for which there is no QED analogue.

Exercise 18.1 Find the number of fermions fields required for the effect of the fermions in Eq. (18.7) to overwhelm the gluon term. If the gluon term dominates does the coupling get larger or smaller when the energy scale Q increases? What about if the fermion term is larger than the gluon term?

The six known quark fields are not enough to overcome the effects of the gluon contribution to the running of the QCD coupling. Unlike QED, this means that at as energies get larger α_s gets smaller. This is referred to as asymptotic freedom. At large enough energies the strong force coupling is small, so the quarks are effectively free from QCD interactions.

The other side of this behavior is that as energies get lower the QCD coupling becomes larger and larger. At some point it becomes so large that the perturbation theory used to calculate the running breaks down. In such a situation we must turn to some other way of analyzing the theory. One nonperturbative way of understanding gauge theories like QCD is to consider a discrete spacetime lattice of points, instead of a continuum, and describe the theory numerically on the lattice, this is known as Lattice Gauge Theory.

Using lattice gauge theory it has been shown that at low energies, where perturbative QCD points to a diverging coupling between quarks and gluons, there is an effective potential between a quark and an antiquark [33]

$$V_{\rm QCD}(r) \approx \sigma r \ . \tag{18.8}$$

These works further show that we can think of quark-antiquark pair as connected by a tube or string of color flux [34, 35]. We can use the potential to find the force between the quarks in the usual way

$$F = -\frac{d}{dr} V_{\rm QCD}(r) = -\sigma \ . \tag{18.9}$$

This is a constant attractive force that acts on the quarks no matter how far away they are from each other. This is why σ in Eq. (18.8) is called the string tension. Note that as the quarks are pulled further and further apart the potential energy increases, which we think of as being stored in the string. If the quarks are separated by a distance L then the potential energy from QCD is

$$V_{\rm QCD} = \sigma L \ . \tag{18.10}$$

The total energy of this configuration is the potential energy plus the mass energy of the antiquark $\overline{q}_1$ and the quark q_2

$$E_1 = \sigma L + m_1 + m_2 \ , \tag{18.11}$$

where we are allowing the two fields to have different masses. Let's compare this configuration to one in which a section from the middle of flux tube has been removed. To do this we must produce a particle-antiparticle pair of quarks q and $\overline{q}$ so that our two strings go from $\overline{q}_1$ to q and $\overline{q}$ to q_2. These two sections of string are separated by a distance $\Delta L \sim 1/m_q$. The distance of separation is inexact, but as discussed when describing the QED vacuum polarization the characteristic size of vacuum fluctuations is given

by the inverse of the particle mass. The energy of this configuration is

$$\begin{aligned} E_2 =& \sigma\left(L-\Delta L\right)+m_1+m_2+2m_q \\ \approx& \sigma\left(L-\frac{1}{m_q}\right)+m_1+m_2+2m_q\ . \end{aligned} \tag{18.12}$$

We expect whichever of E_1 and E_2 is lower to be the configuration preferred by nature, the one with lowest energy. Note that

$$E_1-E_2=\sigma\Delta L-2m_q\ , \tag{18.13}$$

is positive if

$$\sigma\Delta L>2m_q\ , \tag{18.14}$$

which is approximately equal to

$$\sigma>2m_q^2\ . \tag{18.15}$$

From lattice calculations [36] in SU(3) gauge theory it has been determined that $\sqrt{\sigma}=485\pm 6$ MeV, and this is taken to be about the energy scale when QCD changes phase. It is similar to the value one finds for the scale, sometimes called $\Lambda_{\rm QCD}$, when the perturbative running of α_s becomes infinite. The result in (18.15) implies that if there are quarks with masses below about 350 MeV then $E_1>E_2$, which means it is energetically favorable for these light quarks to be pair produced when the flux tube between then is longer than about $1/\sqrt{\sigma}\sim 4\times 10^{-16}$ meters. In other words, in such a scenario quarks would be confined to bound states of a few 10^{-16} meter size. Lattice calculations confirm of this preference for quark pair creation [37].

Experimentally, we know of three quarks(up, down, and strange) with masses below $\sqrt{\sigma}$ which is why we cannot find isolated quarks. If two quarks are in a bound state and we try to pull them apart to examine just one of the quarks then we are adding energy into the system. In order to see an isolated quark we add enough energy to pair create new quarks which bind to the original pair, making two bound states. These quark bound states are generically called **Hadrons** and are found to have a size on the order of a few 10^{-16} meters, in agreement with the lattice QCD calculation of the string tension. For instance, the proton is typically taken to have a size of 8.4×10^{-16} meters.

What kind of bound states can occur? The most significant quality of a QCD bound state is that all of the color interactions are bound within the state. So, from the outside QCD is hidden. In other words, the full bound state must be in the singlet, or trivial, or **1** representation of SU(3). A more involved discussion of the representations of SU(3) is given in Sec. 39.1 but here we simply claim results. The interested reader can look to that section

for more detail.

When constructing the SU(3) invariant Lagrangian, see Eq. (16.9), we found that $\overline{\psi}_i\psi^i$ is in the trivial representation of SU(3). We can see this from representation theory by noting that the tensor product of something in the fundamental representation **3** (like ψ^i) and something in the conjugate representation of the fundamental $\overline{\mathbf{3}}$ (like $\overline{\psi}_i$) obeys

$$\mathbf{3} \otimes \overline{\mathbf{3}} = \mathbf{8} \oplus \mathbf{1} \ . \tag{18.16}$$

When we sum over the group indices $\overline{\psi}_i\psi^i$ we select only the **1** part. Thus, we expect there to be bound states of quarks and antiquarks. Many such bound states have been discovered and studied, and in general they are called **Mesons**. The pions, $\pi^\pm$ and π^0, are the lightest known mesons.

What about the combination of two quarks? This corresponds to $\mathbf{3} \otimes \mathbf{3} = \mathbf{6} \oplus \overline{\mathbf{3}}$, which has no singlet **1** part. Consequently, we do not expect to find QCD bound states of two quarks. Similarly, two antiquarks leads to $\overline{\mathbf{3}} \otimes \overline{\mathbf{3}} = \overline{\mathbf{6}} \oplus \mathbf{3}$ which, as might be expected, does not allow for bound states of this type.

We might next consider three particles. Two quarks and an antiquark is characterized by $\mathbf{3} \otimes \mathbf{3} \otimes \overline{\mathbf{3}} = \mathbf{15} \oplus \overline{\mathbf{6}} \oplus \mathbf{3} \oplus \mathbf{3}$. This does not include a **1** so there is no possible bound state of this type. Three quarks, however, leads to

$$\mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3} = \mathbf{10} \oplus \mathbf{8} \oplus \mathbf{8} \oplus \mathbf{1} \ . \tag{18.17}$$

The singlet part can be obtained by using the definition of the determinant discussed around Eq. (35.73). Using this relation we note that

$$\varepsilon_{ijk}U^i{}_\ell U^j{}_m U^k{}_n = \mathrm{Det}(U)\varepsilon_{\ell mn} \ , \tag{18.18}$$

where all these indices correspond to the SU(3) symmetry space. Note that a matrix in the fundamental representation of SU(3) has determinant equal to one, this is what "special"—the S of SU(3)—means for matrices. This means that under a symmetry transformation three quark fields transform as

$$\begin{aligned} \varepsilon_{ijk}\psi^i\psi^j\psi^k &\to \varepsilon_{ijk}U^i{}_\ell\psi^\ell U^j{}_m\psi^m U^k{}_n\psi^n \\ &= \varepsilon_{\ell mn}\psi^\ell\psi^m\psi^n \ . \end{aligned} \tag{18.19}$$

The bound states of composed of three quarks are called **Baryons** and the proton and neutron are two examples.

Exercise 18.2 Argue, from representation theory, that the combination of three antiquarks should also allow a bound state. You may use the fact that both the **8** and **1** representations are real, meaning that their conjugate representations equivalent to the original representation. That is $\overline{\mathbf{8}} = \mathbf{8}$, for

example. Show the explicit transformation law of three antiquark bound states to confirm this.

There is another QCD bound state that does not involve any quarks. The combination of two gluons transform in the representation

$$\mathbf{8} \otimes \mathbf{8} = \mathbf{27} \oplus \mathbf{10} \oplus \overline{\mathbf{10}} \oplus \mathbf{8} \oplus \mathbf{8} \oplus \mathbf{1} \ . \tag{18.20}$$

Importantly, there is a singlet contribution to this. This implies that there is a combination of two gluons that is a color singlet. These gluon bound states are called **glueballs**. Adding a third gluon produces two $\mathbf{8}\otimes\mathbf{8}$ factors in the direct sum, each of which contain a singlet part. Consequently, there are glueballs that are combinations of more than two gluons. In general glueballs are simply bound states that are not composed of quarks.

The force due to the SU(3) gauge bosons is the true strong nuclear force. The gluons are massless, just like the photon. But, while the photon's influence is formally infinite, the reach of the gluons is confined to length scales set by $1/\sqrt{\sigma} \sim 4 \times 10^{-16}$ meters. However, one often hears of the strong nuclear force keeping the nucleus together, rather than keeping the protons and neutrons together.

It turns out that both of these ideas are correct. There is a vestige of the strong nuclear force that attracts nucleons, meaning both protons and neutrons, to each other. This can be effectively modeled as the exchange of pions. The pions are the lightest QCD mesons, and are pseudoscalars. Like the proton and neutron the up and down quarks dominate their internal structure. Unlike gluons, the pions have masses of about 140 MeV. Pion exchange leads to Yukawa potential

$$V(r) = -\frac{g_\pi^2}{4\pi}\frac{e^{-rm_\pi}}{r} \ , \tag{18.21}$$

between the nucleons, where g_π is the coupling between each pion and the neucleons. For distances $r > 1/m_\pi \sim 1.4 \times 10^{-15}$ we see that this potential is exponentially suppressed, making it only effective over short distances. However, these distances are a little larger than $1/\sqrt{\sigma}$.

Of course, there are more bound states of the up and down quarks than just the pions. The next lightest states are the three rho particles, ρ^0 and $\rho^\pm$, which are vectors with masses of about 775 MeV. Also, the omega ω is vector bound state with mass of about 783 MeV. Quantum field theory calculations show that the exchange of a vector between two particles (rather than a particle and an antiparticle) produces the Yukawa-like potential

$$V(r) = \frac{g_V^2}{4\pi}\frac{e^{-rm_V}}{r} \ , \tag{18.22}$$

where g_V and m_V are the coupling and mass of the vector particle to the

nucleons. While the form of this potential is the same, the sign is crucially different.

Exercise 18.3 Assuming that g_π and the g_V's are all the same size calculate the force between nucleons using

$$F = -\frac{dV}{dr} \ . \tag{18.23}$$

Where is the force attractive and where is it repulsive. Where is the force zero?

SUBSECTION 18.2

Partons, Fragmentation, and Showers

When the idea that objects like protons and pions were built from more elementary constituents was first being explored the generic name given to the constituents was "partons." For instance, it is sometimes said that the proton is composed of a two up quarks and one down quark. The labels up and down refer to two **Flavors** of quark. This rather silly label has nothing to do with culinary flavors, but simply means that there are several types of quarks that we would like to distinguish. They have the whimsical names up, down, strange, charm, bottom, and top. We discuss them in more detail below.

The up quark has electric charge of 2/3 and the down quark has electric charge -1/3, meaning that two ups and a down has total charge of 1, just like a proton. But the simple model of a proton made up of three quarks is not quite correct. These days the proton is characterized by **parton distribution functions** or PDFs. These functions $f_q^h(x)$ describe, for a hadron h with some momentum, what fraction x of the total momentum is carried by particles q of some specific type. The proton is characterized, at least in part, by two relations

$$\int_0^1 dx \, [f_u^p(x) - f_{\overline{u}}^p(x)] = 2 \ , \qquad \int_0^1 dx \, \left[f_d^p(x) - f_{\overline{d}}^p(x)\right] = 1 \ . \tag{18.24}$$

This indicates that the total momentum carried by up quarks minus the momentum carried by antiup quarks is two, and a similar relation for the down- and antidown quarks.

Note that these relations allow for anti-quarks to make up some aspects of the proton. The fact is that the inside of a hadron is a huge nonpertubative mess of strongly interacting quarks and gluons. Indeed, quite a lot of the proton momentum fraction is carried by the gluons. For the proton we say that the two up quarks and the down quark are the **valence quarks**. These carry a lot of the momentum. The other quark fields are called **sea quarks**. They play a nonzero role in the proton, but are subdominant

to the valence quarks. They make up part of the sea of nonperturbative strong interactions.

The PDFs are very important to making predictions at hadron colliders and other experiments. Using the tools of perturbative quantum field theory we can often calculate the leading order cross section for two quarks colliding to produce some final state, $qq \to X$. But the quantity we need is the collision of two hadrons (typically protons or antiprotons) into that final state, $pp \to X$. The proton PDFs allow us to use the partonic cross section to find the hadronic cross section

$$\sigma(pp \to X) \sim \sum_q \sum_{\bar{q}} \int \frac{dx}{x} f_q^p(x) f_{\bar{q}}^p(x) \hat{\sigma}(q\bar{q} \to X) \ . \tag{18.25}$$

The details of the exact relation do not concern us in this work.

On the other side of things, suppose we produce a Z boson (we learn more about this object below) in some experiment. Much of the time this Z decays into quarks, but of course we do not detect the individual quarks produced. What is actually detected is a number of different types of hadrons. Going from initial quarks to numbers of hadrons is characterized by a quark **fragmentation function**.

The actual process of going from initial quark to final hadrons includes a lot of nonperturbative effects. But there are also significant portions of the process for which perturbative calculations are helpful. If we imagine the quarks are produced at an energy scale well above $\sqrt{\sigma}$ then α_s is small enough for perturbation theory to be a good guide. But by the time we get to the final state, on-shell, bound states we have passed through the QCD phase transition.

Figure 18. Initial, off-shell, quark state in parton shower.

Getting from the initial perturbative process with two, for example, quarks to the final state of many hadrons is accomplished, in part, by what is called a **parton shower**. We begin, as characterized in Fig. 18 by a quark field. However, because it is not a final state particle it is allowed to be very off-shell. This means that if the quark field has mass m_q and the momentum flowing through the line is p^μ, then $p^2 > m_q$. The quark field can move toward being on-shell by reducing its energy through radiating a gluon, as shown in Fig. 19, which may also be off-shell.

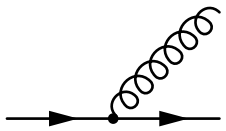


Figure 19. Off-shell quark radiating a gluon.

Now there are two off-shell partons (the quark and gluon) that can move toward begin on-shell by shedding energy. These include the quark radiating off another gluon, or the gluon radiating a gluon, see Fig. 20, or the gluon becoming a quark-antiquark pair as in Fig. 21. Then each of the produced partons can continue the splitting process, all of them getting closer and closer to on-shell and lower energy.

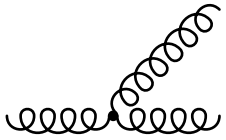


Figure 20. Off-shell gluon radiating a gluon.

In total all these effects produce a shower of partons, as shown in Fig. 22. Of course, when the energy becomes low enough, close to $\sqrt{\sigma}$ perturbation theory becomes a poor approximation of the dynamics. At this point a hadronization model is used to group the endpoints of the

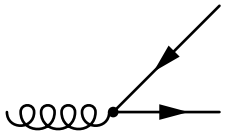


Figure 21. Off-shell

shower into hadrons.

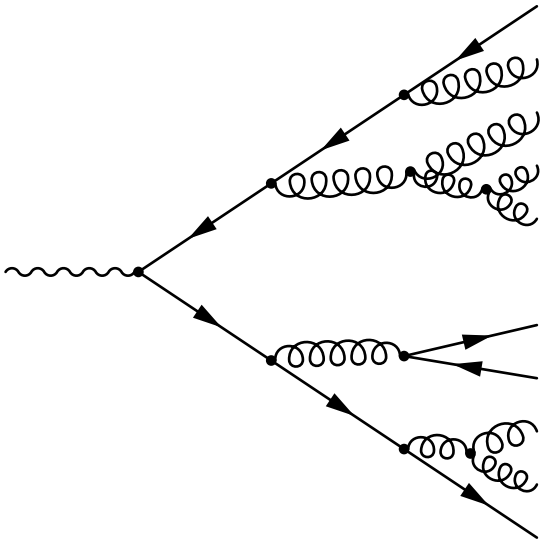


Figure 22. Schematic parton shower.

While this has all been somewhat schematic, these processes have been quite successful in simulating experimental events. One more piece of information can aid our understanding. The process of gluon and quark splitting have collinear singularities. That is, the amplitudes become very large when the two final state particles have momenta in the same direction. This implies that it is most likely for these radiated particles to have momenta close to the original off-shell quarks. Experimentally, we do find that the hadrons that result from quark production are organized into **jets** of particles, grouped around some direction.

PART

V

Electroweak Symmetry Breaking

In this part of the text we introduce the dynamics responsible for the weak nuclear force. We also outline the origin of QED, not as a fundamental symmetry but the remnant of a larger symmetry $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ that is spontaneously broken. This process involves what has come to be called the Higgs mechanism.

Sec. 19. $\mathrm{SU}(2)_L$
Sec. 20. Higgs
Sec. 21. $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$
Sec. 22. Beta Decay

Section 19

Nonabelian $\mathrm{SU}(2)_L$

Some of the basics of a nonablelian SU(2) of symmetry were introduced in Sec. 17. In this section we focus on understanding what is different about SU(2)$_L$. The subscript L in this case is a reminder that this local symmetry only involves left-handed fermions (at least within the Standard Model). The notion of left- and right-handed spin-1/2 fields was introduced in Sec. 9. Using the Weyl basis for the γ^μ we saw that a four-dimensional Dirac spinor field can be written as

$$\psi = \begin{pmatrix} \psi_L \\ \psi_R \end{pmatrix} \ , \tag{19.1}$$

where ψ_L and ψ_R are two-dimensional Weyl spinors. Of course, these two-dimensional fields can be described in the four-dimensional space by simply filling in the remaining components with zeros. Therefore, the same notation, for instance

$$\psi_L \equiv P_L \psi \ , \tag{19.2}$$

is often used to describe four-dimensional fields that only contain the left-handed part of a field. Here we have used the chirality projector defined in Eq. (9.44).

Unlike the couplings we found in QED

$$qe\overline{\psi}\gamma^\mu\psi A_\mu \ , \tag{19.3}$$

and QCD

$$g_s\overline{\psi}T^a\gamma^\mu\psi A^a_\mu \ , \tag{19.4}$$

the gauge bosons of SU(2)$_L$ only couple to left-handed fermions

$$g\overline{\psi}_L t^a\gamma^\mu\psi_L A^a_\mu \ , \tag{19.5}$$

where the t^a are the generators of the SU(2)$_L$ gauge group. We see that the structure is much like that of the SU(3) of QCD, except for the chirality specific coupling.

In Sec. 15.1 we found that under parity transformations the Weyl fermions transform as

$$\psi_L \xrightarrow{P} \pm\psi_R \ , \quad \psi_R \xrightarrow{P} \pm\psi_L \ . \tag{19.6}$$

Therefore, under a parity transformation

$$g\overline{\psi}_L t^a\gamma^\mu\psi_L A^a_\mu \xrightarrow{P} \eta_P g\overline{\psi}_R t^a\gamma^\mu\psi_R A^a_\mu \ , \tag{19.7}$$

where $\eta_P = \pm 1$ corresponds to the A^a_μ being either vectors or axial vectors. We saw above that the structure of the nonabelian field strength requires

that the gauge fields to be vectors ($\eta_P = 1$), but we keep track of this formal possibility for the moment. Importantly, there are no couplings of the SU(2)$_L$ gauge bosons to right-handed fields. This shows that parity is not a symmetry of this theory. In fact, this theory is often said to be maximally parity violating [7, 14, 15].

This terminology can be understood in the following way. Suppose a theory has couplings

$$gc_L\overline{\psi}_L t^a\gamma^\mu\psi_L A^a_\mu + gc_R\overline{\psi}_R t^a\gamma^\mu\psi_R A^a_\mu \ , \tag{19.8}$$

in terms of its chiral projections. Before it was known that the spin-1 fields of the related to the weak force were not axial vectors there were two possible ways for the interactions to be parity invariant. This is perhaps most obvious by rewriting the interaction in terms of

$$c_L = c_+ + c_- \ , \qquad c_R = c_+ - c_- \ , \tag{19.9}$$

to find

$$gc_+\left(\overline{\psi}_L t^a\gamma^\mu\psi_L + \overline{\psi}_R t^a\gamma^\mu\psi_R\right)A^a_\mu + gc_-\left(\overline{\psi}_L t^a\gamma^\mu\psi_L - \overline{\psi}_R t^a\gamma^\mu\psi_R\right)A^a_\mu \ . \tag{19.10}$$

Under parity this becomes

$$gc_+\eta_P\left(\overline{\psi}_L t^a\gamma^\mu\psi_L + \overline{\psi}_R t^a\gamma^\mu\psi_R\right)A^a_\mu - \eta_P gc_-\left(\overline{\psi}_L t^a\gamma^\mu\psi_L - \overline{\psi}_R t^a\gamma^\mu\psi_R\right)A^a_\mu \ . \tag{19.11}$$

This shows that the c_+ terms is parity invariant when $\eta_P = 1$ (vector A^a_μ) and the c_- part is invariant under parity only for $\eta_P = -1$ (axial-vector A^a_μ).

By expressing these couplings as

$$c_+ = \frac{c_L + c_R}{2} \ , \qquad c_- = \frac{c_L - c_R}{2} \ , \tag{19.12}$$

we see that the two parity preserving options are when $c_L = \pm c_R$. We might then define the deviation from the parity preserving case by

$$V_1(c_L,\, c_R) = \frac{\big||c_L| - |c_R|\big|}{|c_L| + |c_R|} \ . \tag{19.13}$$

One can show that the structure of the weak interactions, $c_R = 0$, does maximize this measure of deviation.

Once we know that the gauge fields are vector (rather than axial-vectors) then the complete c_- interaction term is parity violating. In this case we might simply define the deviation from a parity preserving theory

by

$$V_2(c_L, c_R) = \left|\frac{c_-}{c_+}\right| \ . \tag{19.14}$$

This measure is not maximized by the $c_R = 0$ form of the weak force.

Exercise 19.1 Find the values of $|c_L|$ and $|c_R|$ that minimize and maximize V_1 and the values of c_L and c_R that maximize and minimize V_2. How do these extreme values compare to the $c_R = 0$ form of the $\mathrm{SU}(2)_L$ gauge theory of the Standard Model? What about the $c_R = c_L$ form of QED and QCD?

Exercise 19.2 Show that even though the $\mathrm{SU}(2)_L$ interactions are maximally parity violating, they preserve the discrete CP symmetry as long as

$$A^a_\mu \xrightarrow{C} -A^a_\mu \ , \tag{19.15}$$

similar to the photon field.

Aside from this parity violating characteristic the SU(2) theory has many similarities to the SU(3). It has a running coupling that is asymptotically free and confining at low energies. However, we do not see any $\mathrm{SU}(2)_L$ bound states experimentally. It turns out that this is because this symmetry is **spontaneously broken** at the scale of about 100 GeV, well above its confining scale. This process occurs through what has come to be called the Higgs mechanism.

Section 20

The Higgs Mechanism

The Higgs mechanism [38–41] is a property of scalar field theories charged under a gauge field. To understand how it works we need to introduce several concepts and ideas. These are built up in the following subsections.

Subsection 20.1

Spontaneous Symmetry Breaking

The first crucial ingredient is **spontaneous symmetry breaking** in scalar field theories.

Definition 2 A system exhibits **spontaneous symmetry breaking** when the symmetry of the lowest energy state (ground state or vacuum) is less than the full symmetry of the theory.

To understand this definition we consider the lowest energy state, or ground state, of a scalar field theory. This is often referred to as the vacuum state of the theory. Consider the Lagrangian for a real scalar field

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - U(\phi) \ , \tag{20.1}$$

where $U(\phi)$ is called the potential for ϕ. The Euler-Lagrange equations lead to

$$\partial_\mu\partial^\mu\phi = -\frac{dU}{d\phi} \ , \tag{20.2}$$

which looks something like Newton's second law. The energy contained in the field is given by

$$E = \int d^3x \left[|\dot{\phi}|^2 + |\nabla\phi|^2 + U(\phi)\right] \ . \tag{20.3}$$

How can the energy be minimized? First, we assume that $U(\phi)$ has a finite minimum. If this were not the case then the field ϕ could continue to go to lower and lower values, which would mean $\dot{\phi}$ could grow without bound. We do not see any evidence of such strange behavior in nature. This motivates assuming that $U(\phi)$ has a minimum, which we may as well define to be zero.

Having made this assumption we see that each term in the energy (20.3) is positive. Therefore, the only way to minimize the energy is to minimize each term. First, if the field is static (not changing in time) then $\dot{\phi} = 0$. Second, if the field is constant in space the $\nabla\phi = 0$. The potential term is minimized when the value of ϕ minimizes $U(\phi)$. In short, the vacuum of the theory is for the field to come to rest and the same value everywhere in space at the minimum of $U(\phi)$.

Example

As an example, we consider the potential

$$U(\phi) = \frac{m^2}{2}\phi^2 + \frac{\lambda}{4}\phi^4 \ , \tag{20.4}$$

for a real scalar field ϕ. Note that as $\phi \to \infty$ that $U(\phi) \to \pm\infty$ depending on whether λ is positive or negative. Because we want $U(\phi)$ to have a finite minimum we take $\lambda > 0$. Notice that this potential is invariant under the discrete symmetry $\phi \to -\phi$.

To find the extrema of the potential we determine where the derivative vanishes

$$\frac{dU}{d\phi} = 0 = m^2\phi + \lambda\phi^3 = \phi\left(m^2 + \lambda\phi^2\right) \ . \tag{20.5}$$

Clearly, one solution is $\phi = 0$. With m real and $\lambda > 0$ there are no other real solutions. To determine if $\phi = 0$ is a maximum or minimum

we consider the second derivative

$$\frac{d^2U}{d\phi^2} = m^2 + 3\lambda\phi^2 \ . \tag{20.6}$$

At $\phi = 0$ this is simply $m^2 > 0$, which implies that $\phi = 0$ is a minimum. In summary, $\phi = 0$ is the vacuum configuration of the scalar field. This vacuum field configuration preserves the discrete symmetry of the potential because $-0 = 0$. Because the symmetry of the vacuum matches the full symmetry of the theory there is **no** spontaneous symmetry breaking.

The example above is a completely classical analysis. For such a scenario the quantum field theory would be of small, quantum, fluctuations φ around this classical background state. This classical vacuum is also the average value of the quantum field, or the **Vacuum Expectation Value** (VEV) denoted by $\langle\phi\rangle$. In other words, the quantum theory takes $\phi = \langle\phi\rangle + \varphi$ and considers the dynamics of φ. In the previous example $\langle\phi\rangle = 0$ so the potential for φ is identical to the potential for ϕ.

Exercise 20.1 For the scalar potential

$$U(\phi) = \frac{m^2}{2}\phi^2 + \frac{\lambda}{4}\phi^4 \ , \tag{20.7}$$

set $\phi = \langle\phi\rangle + \varphi$ and find the coefficients of each power of φ. Then set $\langle\phi\rangle = 0$ and compare these coefficients to the ϕ coefficients in $U(\phi)$.

The preceding exercise shows that for the potential in question the mass of the field describing the quantum fluctuations around the vacuum has the same mass as the classical scalar field in question. However, this is not always the case, which is crucial to the Higgs mechanism.

Example Consider the potential

$$U(\phi) = -\frac{\mu^2}{2}\phi^2 + \frac{\lambda}{4}\phi^4 \ , \tag{20.8}$$

for a real scalar field ϕ. As with the previous example we require $\lambda > 0$ to ensure the potential has a finite minimum. Note also that what we usually think of the mass term is negative. As in the former example this potential is also symmetric under $\phi \to -\phi$.

The extrema of the potential are determined by

$$\frac{dU}{d\phi} = 0 = \phi\left(-\mu^2 + \lambda\phi^2\right) \ . \tag{20.9}$$

This has three solutions, $\phi = 0$ and $\phi = \pm\mu/\sqrt{\lambda}$. We determine whether these are maxima or minima by considering the second derivative

$$\frac{d^2U}{d\phi^2} = -\mu^2 + 3\lambda\phi^2 \ . \tag{20.10}$$

If $\phi = 0$ this is negative, showing that there is a local maximum at $\phi = 0$. This point cannot be taken as a vacuum for the theory. For $\phi = \pm\mu/\sqrt{\lambda}$ the second derivative is

$$\left.\frac{d^2U}{d\phi^2}\right|_{\pm\mu/\sqrt{\lambda}} = -\mu^2 + 3\lambda\frac{\mu^2}{\lambda} = 2\mu^2 \ . \tag{20.11}$$

This is a positive quantity so both of these field values correspond to a local minimum, and as it happens they are both the global minimum as well. It is important to notice that each of these vacuum states are related to the other by $\phi \to -\phi$, but also that neither, on its own, preserves the discrete symmetry of the potential. In seeking the minimum energy configuration, the vacuum, the scalar field must "choose" one of the two minimum. This is an example of spontaneous symmetry breaking.

Exercise 20.2 For the potential

$$U(\phi) = -\frac{\mu^2}{2}\phi^2 + \frac{\lambda}{4}\phi^4 \ , \tag{20.12}$$

set $\phi = \langle\phi\rangle + \varphi$ and find the coefficients of each power of φ. Then set $\langle\phi\rangle = \pm\mu/\sqrt{\lambda}$ and compare these coefficients to the ϕ coefficients in $U(\phi)$.

In the above exercise one finds that the mass of the quantum fluctuation field φ is $\sqrt{2}\mu$. This agrees exactly with Eq. (20.11), the value of the second derivative of the potential evaluated at the vacuum gives the squared mass of fluctuations around that vacuum. This explains why we can have scalar particles with positive mass even when the classical potential has a term that looks like a negative mass.

It is important to understand that only scalar fields can have a nonzero VEV in a Lorentz invariant theory. If a spin-1/2 or spin-1 field have a nonzero VEV then there would be a special direction in space associated with the angular momentum (spin) of the VEV. This would change when using a Lorentz transformation to compare one frame with another. Therefore, only spin-0 fields can get a VEV.

Subsection 20.2

Breaking Continuous Symmetries

Up to this point we have only considered explored the spontaneous breaking of a discrete symmetry. However, the Higgs mechanism pertains to

continuous symmetries. Perhaps the simples continuous symmetry is the rephasing symmetry of a scalar field theory

$$\mathcal{L} = \partial_\mu \phi^\dagger \partial^\mu \phi - U(\phi^\dagger \phi) \ . \tag{20.13}$$

Note that we have specified that the potential is a function of $\phi^\dagger\phi \equiv |\phi|^2$ to ensure that

$$\phi \to \phi' = e^{-i\theta}\phi \ , \qquad \phi^\dagger \to \phi'^\dagger = \phi^\dagger e^{i\theta} \ , \tag{20.14}$$

is a symmetry of the Lagrangian.

Exercise 20.3 Show that the Lagrangian in (20.13) leads to the same conserved charge Q we found for complex Klein-Gordon fields.

Example Suppose that

$$U(\phi^\dagger\phi) = m^2\phi^\dagger\phi + \lambda\left(\phi^\dagger\phi\right)^2 \equiv m^2|\phi|^2 + \lambda|\phi|^4 \ . \tag{20.15}$$

In order to have a finite minimum for the potential we require $\lambda > 0$. This is now a potential in two real fields. One way to express them is in terms of a linear combination

$$\phi(x) = \frac{1}{\sqrt{2}}\left(X(x) + iY(x)\right) \ . \tag{20.16}$$

The potential then takes the form

$$U(X,Y) = \frac{m^2}{2}\left(X^2+Y^2\right) + \frac{\lambda}{4}\left(X^2+Y^2\right)^2 \tag{20.17}$$

To find the critical points we evaluate

$$\frac{\partial U}{\partial X} = 0 = m^2 X + \lambda X\left(X^2+Y^2\right) \ , \tag{20.18}$$

$$\frac{\partial U}{\partial Y} = 0 = m^2 Y + \lambda Y\left(X^2+Y^2\right) \ . \tag{20.19}$$

The only solution to these equations with $m^2 > 0$ and $\lambda > 0$ is $X = Y = 0$.

Is this point a maximum, a minimum, or a saddle point? For this we must construct the Hessian matrix, constructed from the second deriva-

tives of U

$$M = \begin{pmatrix} \frac{\partial^2 U}{\partial X^2} & \frac{\partial^2 U}{\partial X \partial Y} \\ \frac{\partial^2 U}{\partial X \partial Y} & \frac{\partial^2 U}{\partial Y} \end{pmatrix} = \begin{pmatrix} m^2 + \lambda(3X^2 + Y^2) & 2\lambda XY \\ 2\lambda XY & m^2 + \lambda(X^2 + 3Y^2) \end{pmatrix} . \tag{20.20}$$

The determinant of this matrix is

$$\begin{aligned} \text{Det}(M) =& m^4 + 4m^2\lambda\left(X^2 + Y^2\right) + 3\lambda^2\left(X^2 + Y^2\right)^2 \\ =& \left[m^2 + \lambda\left(X^2 + Y^2\right)\right]\left[m^2 + 3\lambda\left(X^2 + Y^2\right)\right] . \end{aligned} \tag{20.21}$$

Which at $X = Y = 0$ is simply $m^4 > 0$. At this same point the top left component of M is $m^2 > 0$ which implies the matrix is positive definite and hence that this point is a minimum. This minimum trivially preserves the rephasing symmetry because $0e^{-i\theta} = 0$. Therefore this symmetry is not spontaneously broken.

We point out that there is another, and somewhat simpler, way to find these same results. We can write ϕ in terms of two real fields in such a way as to resemble the symmetry group in question. Specifically, we write

$$\phi(x) = \frac{1}{\sqrt{2}} r(x) e^{-i\theta(x)/f} , \tag{20.22}$$

where f is a constant with dimensions of mass. In this case $\phi^\dagger\phi = r^2/2$ so the potential is simply

$$U(r) = \frac{m^2}{2} r^2 + \frac{\lambda}{4} r^4 . \tag{20.23}$$

This is exactly the one dimensional potential (20.4) analyzed in the previous section.

Exercise 20.4 Work through the process shown in the above example for the potential

$$U(\phi^\dagger\phi) = -\mu^2|\phi|^2 + \lambda|\phi|^4 . \tag{20.24}$$

Is the point $\phi = \phi^\dagger = 0$ a maximum or a minimum? You should now find a continuous family of extrema determined by

$$\frac{\mu^2}{\lambda} = X^2 + Y^2 . \tag{20.25}$$

What is the value of the determinant of the Hessian matrix? What does this imply about this family of extrema?

We do find spontaneous symmetry breaking for the potential

$$U(\phi^\dagger\phi) = -\mu^2|\phi|^2 + \lambda|\phi|^4 \ . \tag{20.26}$$

This is most easily see by using the parameterization

$$\phi(x) = \frac{1}{\sqrt{2}} r(x) e^{-i\theta(x)/f} \ , \tag{20.27}$$

of the scalar field. Then the potential becomes

$$U(r) = -\frac{\mu^2}{2} r^2 + \frac{\lambda}{4} r^4 \ , \tag{20.28}$$

which is exactly the same form as was discussed above (20.8). This implies that $r(x)$ has vacua at $\pm\mu/\sqrt{\lambda}$ and undergoes spontaneous symmetry breaking.

However, $\theta(x)$ does not appear in the potential and consequently has a VEV of zero. To determine the dynamics about the vacua we assume the form

$$\phi = \frac{v+h}{\sqrt{2}} e^{-i\theta/f} \ , \tag{20.29}$$

where

$$v = \sqrt{2}\langle\phi\rangle = \frac{\mu}{\sqrt{\lambda}} \ . \tag{20.30}$$

The potential can be written as

$$U(h) = -\frac{v^2\mu^2}{4} + \frac{2\mu^2}{2} h^2 + \lambda v h^3 + \frac{\lambda}{4} h^4 \ . \tag{20.31}$$

This shows that the mass of the h field is $m_h = \sqrt{2}\mu$. We can drop the constant term because it does not affect the dynamics of the theory (neglecting gravity) and express the potential as

$$U(h) = \frac{m_h^2}{2} h^2 + \sqrt{\frac{\lambda}{2}} m_h h^3 + \frac{\lambda^4}{4} h^4 \ . \tag{20.32}$$

Notice that a signal of the spontaneous symmetry breaking is that the coefficient of the cubic term is completely determined by the mass and quartic coupling.

So far, we have not written out the derivative term. It is straightforward to show that

$$\partial_\mu \frac{v+h}{\sqrt{2}} e^{-i\theta/f} = e^{-i\theta/f} \frac{1}{\sqrt{2}} \left(\partial_\mu h - i \frac{v+h}{f} \partial_\mu \theta \right) \ . \tag{20.33}$$

This leads to

$$\partial_\mu \phi \partial^\mu \phi = \frac{1}{2}\partial_\mu h \partial^\mu h + \frac{v^2 + 2vh + h^2}{2f^2}\partial_\mu \theta \partial^\mu \theta \ . \tag{20.34}$$

The derivative for h is exactly as it should be for a real scalar field. To ensure that θ has the correct derivative term we must take $f = v$. Then we have

$$\partial_\mu \phi \partial^\mu \phi = \frac{1}{2}\partial_\mu h \partial^\mu h + \frac{1}{2}\partial_\mu \theta \partial^\mu \theta + \frac{h}{v}\partial_\mu \theta \partial^\mu \theta + \frac{h^2}{2v^2}\partial_\mu \theta \partial^\mu \theta \ . \tag{20.35}$$

So, while θ does not appear in the potential it does couple to the h field. These couplings are different than others we have discussed in that they depend not on θ but its derivatives. In particular, if we shifted θ by $\theta \to \theta' = \theta + c$ for some constant c then this only changes ϕ by a phase, which does not affect the system. The derivative couplings of θ are the manifestation of this shift symmetry.

Let us pause for a moment to understand this system. The potential, and the theory, have a global U(1) symmetry associated with rephasing the scalar field. However, the vacuum state is one point of a ring of equal energy states. Therefore, while all the vacua are in some sense equivalent any particular vacuum does not have the U(1) symmetry, it is spontaneously broken. In this less symmetric ground state we find two real scalar particles. One corresponds to the radial direction in field space. This radial mode is massive. The other scalar has not potential and no mass.

The massless spin-0 partial is called a Nambu-Goldstone boson [42, 43]. In this section we have seen that the spontaneous breaking a single continuous symmetry leads to the existence of one massless Nambu-Goldstone. This a particular example of a more general result. If some continuous global symmetry G whose Lie algebra has M basis elements is spontaneously broken to a smaller symmetry group H which has a Lie algebra with N basis elements then $M-N$ massless Nambu-Goldstone bosons exist in the theory.

This analysis also applies to symmetries that are not exact. If there is an approximate symmetry that is spontaneously broken then light bosons are produced, rather than massless ones. These are sometimes called pseudo-Nambu-Goldstone bosons, or pNGBS. The masses of these scalars are controlled by the degree explicit symmetry breaking. That is, a symmetry that is more nearly exact produces lighter bosons when spontaneously broken than a symmetry that exhibits larger explicit violations. A symmetry that is barely there at all produces nearly no reduction in the mass of the spin-0 field.

Subsection 20.3

Breaking Gauged Symmetries

The previous subsection outlines how spontaneously breaking continuous global symmetries leads to massless scalar bosons. Experimentally, we have seen no evidence of such massless particles. Why then are spending such effort to understand this effect? The answer is that we are not so interesting in the spontaneous breaking of global symmetries, but of local symmetries.

Consider the theory of a single complex scalar field charged under a local U(1) symmetry. The full Lagrangian is

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + (D_\mu\phi)^\dagger D^\mu\phi - U(\phi,\phi^\dagger) \ , \tag{20.36}$$

where $D_\mu = \partial_\mu + iqeA_\mu$ and the potential is

$$U(\phi,\phi^\dagger) = -\mu^2|\phi|^2 + \lambda|\phi|^4 \ . \tag{20.37}$$

We found in the previous subsections that this potential leads to a spontaneous breaking of the U(1) symmetry. As in those sections we take express the scalar field as

$$\phi(x) = \frac{v+h(x)}{\sqrt{2}}e^{-i\theta(x)/v} \ . \tag{20.38}$$

Nothing changes about how the potential appears in this parameterization. In particular, we note that θ does not appear in the potential.

The derivative terms are somewhat different. We find

$$D_\mu\phi = \frac{e^{-i\theta/v}}{\sqrt{2}}\left(\partial_\mu h - i\frac{v+h}{v}\partial_\mu\theta + iqe(v+h)A_\mu\right) \ . \tag{20.39}$$

This form is quite significant, because we can make the gauge transformation

$$A_\mu \to A'_\mu = A_\mu - \frac{1}{eqv}\partial_\mu\theta \ , \tag{20.40}$$

to find

$$D_\mu\phi = \frac{e^{-i\theta/v}}{\sqrt{2}}\left(\partial_\mu h + iqe(v+h)A'_\mu\right) \ . \tag{20.41}$$

The full derivative term is

$$\begin{aligned}(D_\mu\phi)^\dagger D^\mu\phi =& \frac{e^{i\theta/v}}{\sqrt{2}}\left(\partial_\mu h - iqe(v+h)A_\mu\right)\frac{e^{-i\theta/v}}{\sqrt{2}}\left(\partial^\mu h + iqe(v+h)A^\mu\right)\\ =& \frac{1}{2}\partial_\mu h\partial^\mu h + \frac{qe}{2}(v+h)^2 A_\mu A^\mu \ .\end{aligned} \tag{20.42}$$

This shows that we can always make a gauge transformation to eliminate θ from the derivative terms. Once this is done we have completely removed θ from the Lagrangian. But this is a bit unsettling. The scalar

field had two real degrees of freedom, along with the two degrees of freedom of the massless vector field. Now we seem to have only one spin-0 degree of freedom. At the same time a mass have been given to the gauge field:

$$\frac{q^2e^2v^2}{2}A_\mu A^\mu = \frac{m_A^2}{2}A_\mu A^\mu \ . \tag{20.43}$$

A massive spin-1 field has three degrees of freedom, so this is where the missing scalar degree of freedom has gone. The colorful language used is that the gauge field "ate" the field that would have been a Nambu-Goldstone boson. The point is that the Higgs mechanism moves a degree of freedom from the scalar field to the gauge field. A process like this is what gives mass to the some of the gauge fields associated with $\mathrm{SU}(2)_L$.

One example of spontaneous breaking of a gauge symmetry is superconductivity. Within a superconductor the free electrons pair up into spin-0 states. This collection of paired electrons can be modeled with a complex scalar field whose quanta have charge -2 and of course couple to the electron. In some situations this scalar field develops a VEV, spontaneously breaking the symmetry within the superconductor and giving a mass to the photon. This photon mass is the origin of many superconductor characteristics.

For instance, the **London penetration depth** λ_L describes the attenuation of magnetic fields within a superconductor

$$\vec{B}(z) = B_0 e^{-z/\lambda_L} \ , \tag{20.44}$$

taking the z direction to be perpendicular to the surface of the conductor. In terms of the mass m_γ that the photon has within the superconductor one finds that [44]

$$\lambda_L = \frac{1}{m_\gamma} \ . \tag{20.45}$$

Section 21

Electroweak Symmetry Breaking

In this section we show how the Higgs mechanism occurs within the Standard Model of particle physics. The gauge symmetry that is broken is $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$. The scalar field, the Higgs field, that gets a vacuum expectation value is

$$H = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \ , \tag{21.1}$$

which is in the fundamental representation of $\mathrm{SU}(2)_L$. For now, we leave its $\mathrm{U}(1)_Y$ charge unspecified, and denote it y_H. Note that each component of

this symmetry space two-vector is a complex scalar field. The superscripts we use are explained below.

The potential for this field is

$$V(H^\dagger, H) = -\mu^2 |H|^2 + \lambda |H|^4 \ , \tag{21.2}$$

where $|H|^2 = H^\dagger H$.

Exercise 21.1

Write the components of the Higgs field in terms of real and imaginary parts $\phi^+ = \frac{1}{\sqrt{2}}(X^+ + iY^+)$ and $\phi^0 = \frac{1}{\sqrt{2}}(X^0 + iY^0)$. Write the Higgs potential (21.2) in terms of these fields. Find the extrema and determine if they are minima, maxima, or saddle points. You should find a collection of degenerate vacua that satisfy

$$|H|^2 = \frac{\mu^2}{2\lambda} \equiv \frac{v^2}{2} \ . \tag{21.3}$$

The above exercise shows that there is a family of vacua, all with the same energy, that satisfy Eq. (21.3). Because the theory has an $\mathrm{SU}(2)_L$ symmetry we can always choose to describe the vacuum state with the field configuration

$$\langle H \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \ . \tag{21.4}$$

Similar to the complex scalar parameterization (20.38) used in previous sections, we would like to write the Higgs field in a way that includes the form of the symmetry group elements. In the U(1) case a field $r(x)/\sqrt{2}$ that gets the VEV $v/\sqrt{2}$ was multiplied by a field in the form of a U(1) element $e^{-i\theta(x)/v}$. We need to show the $\mathrm{SU}(2)_L$ equivalent

$$H = U(x) \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ r(x) \end{pmatrix} \ , \tag{21.5}$$

accounts for all possible Higgs field configurations.

Exercise 21.2

Any element of the fundamental representation of SU(2) can be expressed as

$$U(x) = e^{-i\frac{\sigma_a}{2}\frac{\pi_a(x)}{v}} \ , \tag{21.6}$$

where the σ_a are the Pauli matrices matrices are a basis and the $\pi_a(x)$ are real functions.

First, show that

$$\pi_a \sigma_a = \begin{pmatrix} \pi_3 & \pi_1 - i\pi_2 \\ \pi_1 + i\pi_2 & -\pi_3 \end{pmatrix} \equiv M \ . \tag{21.7}$$

The exponential of this matrix is defined in terms of a power series in M. Of course, $M^0 = \mathbb{I}$ and $M^1 = M$. Show that $M^2 \propto \mathbb{I}$ and determine the proportionality constant. You may find it convenient to use the notation

$$\vec{\pi}^{\,2} = \pi_1^2 + \pi_2^2 + \pi_3^2 \ . \tag{21.8}$$

Argue from this result that

$$M^n = \begin{cases} \vec{\pi}^{\,n}\mathbb{I}_2 & n \text{ even} \\ \vec{\pi}^{\,n-1}M & n \text{ odd} \end{cases} \ . \tag{21.9}$$

Use the fact that

$$U(x) = \sum_{n=0}^{\infty} \frac{1}{n!}\left(\frac{-i}{2v}\right)^n M^n \ , \tag{21.10}$$

to show that

$$U(x) = \mathbb{I}_2 \cos\left(\frac{|\vec{\pi}|}{2v}\right) - \frac{i\sigma_a\pi_a}{|\vec{\pi}|}\sin\left(\frac{|\vec{\pi}|}{2v}\right) \ . \tag{21.11}$$

From the above exercise we see that any element of the fundamental representation of $\mathrm{SU}(2)_L$ can be written as

$$U(x) = \begin{pmatrix} \cos\alpha - i\hat{\pi}_3\sin\alpha & -(\hat{\pi}_2 + i\hat{\pi}_1)\sin\alpha \\ (\hat{\pi}_2 - i\hat{\pi}_1)\sin\alpha & \cos\alpha + i\hat{\pi}_3\sin\alpha \end{pmatrix} \tag{21.12}$$

where

$$\alpha = \frac{|\vec{\pi}|}{2v} \ , \qquad \hat{\pi}_a = \frac{\pi_a}{|\vec{\pi}|} \ . \tag{21.13}$$

This leads to

$$\begin{pmatrix} \cos\alpha - i\hat{\pi}_3\sin\alpha & -(\hat{\pi}_2 + i\hat{\pi}_1)\sin\alpha \\ (\hat{\pi}_2 - i\hat{\pi}_1)\sin\alpha & \cos\alpha + i\hat{\pi}_3\sin\alpha \end{pmatrix} \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ r(x) \end{pmatrix} = \frac{r(x)}{\sqrt{2}}\begin{pmatrix} -(\hat{\pi}_2 + i\hat{\pi}_1)\sin\alpha \\ \cos\alpha + i\hat{\pi}_3\sin\alpha \end{pmatrix} \ , \tag{21.14}$$

which can be compared to the definition of H in Eq. (21.1). By making the associations

$$\phi^+ = -r(x)\frac{\sin\alpha}{\sqrt{2}}(\hat{\pi}_2 + i\hat{\pi}_1) \ , \qquad \phi^0 = \frac{r(x)}{\sqrt{2}}\left(\cos\alpha + i\hat{\pi}_3\sin\alpha\right) \ , \tag{21.15}$$

we see that we can always write

$$H = U(x)\frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ r(x) \end{pmatrix} \ . \tag{21.16}$$

This parameterization is very useful. For instance, we see that

$$H^\dagger H = \frac{1}{2}\,(0\,,r)\,U^\dagger U \begin{pmatrix} 0 \\ r \end{pmatrix} = \frac{r^2}{2}\ , \tag{21.17}$$

which implies the potential (21.2) has the form

$$V(r) = -\frac{\mu^2}{2} r^2 + \frac{\lambda}{4} r^4\ . \tag{21.18}$$

This leads to the usual vacuum expectation value v and our expansion $r = v + h$.

Next, we consider the $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ covariant derivative D_μ acting on H. This is

$$D_\mu H = (\partial_\mu + igA_\mu + ig' y_H B_\mu)\, U \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ r \end{pmatrix}\ , \tag{21.19}$$

where the $A_\mu = \frac{\sigma_a}{2} A^a_\mu$ are the $\mathrm{SU}(2)_L$ gauge fields and B_μ is the $\mathrm{U}(1)_Y$ gauge field. Note that each Lie group is associated with a distinct coupling, g for $\mathrm{SU}(2)_L$ and g' for $\mathrm{U}(1)_Y$. It is also important to understand that B_μ acts proportional to the identity in the $\mathrm{SU}(2)_L$ symmetry space. This allows us to rewrite the covariant derivative as

$$D_\mu H = U \left(\partial_\mu + igU^{-1} A_\mu U + U^{-1}\partial_\mu U + ig' y_H B_\mu\right) \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ r \end{pmatrix}\ . \tag{21.20}$$

But, according to form of nonabelian gauge transformations (16.15) this is simply

$$D_\mu H = U \left(\partial_\mu + igA'_\mu + ig' y_H B_\mu\right) \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ r \end{pmatrix}\ . \tag{21.21}$$

This means that in $(D_\mu H)^\dagger\, D^\mu H$ all of $U(x)$ disappears from the Lagrangian, that is the three would-be Nambu-Goldstone bosons have been eaten.

In fact, the preceding analysis is not quite what we want. We could have used gauge transformations of B_μ to remove a degree of freedom. In general, a linear combination of B_μ and A^3_μ can remove a degree of freedom.

The simplest way to see this is to write the covariant derivative as

$$\begin{aligned}\frac{1}{\sqrt{2}}D_\mu\begin{pmatrix}0\\ r\end{pmatrix} =&\frac{1}{\sqrt{2}}\left(\partial_\mu + i\frac{g}{2}\sigma^a A^a_\mu + ig' y_H B_\mu\right)\begin{pmatrix}0\\ r\end{pmatrix}\\ =&\frac{1}{\sqrt{2}}\begin{pmatrix}\partial_\mu + i\left(g' y_H B_\mu + \frac{g}{2}A^3_\mu\right) & \frac{ig}{2}\left(A^1_\mu - iA^2_\mu\right)\\ \frac{ig}{2}\left(A^1_\mu + iA^2_\mu\right) & \partial_\mu + i\left(g' y_H B_\mu - \frac{g}{2}A^3_\mu\right)\end{pmatrix}\begin{pmatrix}0\\ r\end{pmatrix}\\ =&\frac{1}{\sqrt{2}}\begin{pmatrix}\frac{ig}{2}\left(A^1_\mu - iA^2_\mu\right)r\\ \partial_\mu r + i\left(g' y_H B_\mu - \frac{g}{2}A^3_\mu\right)r\end{pmatrix}\ . \end{aligned} \tag{21.22}$$

While this may not look like an improvement, it leads to

$$\begin{aligned}&(D_\mu H)^\dagger D^\mu H = \left[\frac{1}{\sqrt{2}}D_\mu\begin{pmatrix}0\\ r\end{pmatrix}\right]^\dagger \frac{1}{\sqrt{2}}D_\mu\begin{pmatrix}0\\ r\end{pmatrix} = \\ &\frac{1}{2}\partial_\mu r\partial^\mu r + \frac{r^2}{2}\frac{g^2}{4}\left(A^1_\mu + iA^2_\mu\right)\left(A^{1\mu} - iA^{2\mu}\right) + \frac{r^2}{2}\left(g' y_H B_\mu - \frac{g}{2}A^3_\mu\right)\left(g' y_H B^\mu - \frac{g}{2}A^{3\mu}\right)\ .\end{aligned} \tag{21.23}$$

This form suggests that certain combinations of the gauge fields can be usefully grouped together. We define

$$W^\pm_\mu = \frac{1}{\sqrt{2}}\left(A^1_\mu \mp iA^2_\mu\right)\ , \qquad Z_\mu = \frac{1}{\sqrt{g^2 + 4y_H^2 g'^2}}\left(gA^3_\mu - 2g' y_H B_\mu\right)\ , \tag{21.24}$$

and find

$$(D_\mu H)^\dagger D^\mu H = \frac{1}{2}\partial_\mu r\partial^\mu r + r^2\frac{g^2}{4}W^+_\mu W^{-\mu} + \frac{r^2}{2}\frac{g^2 + 4y_H^2 g'^2}{4}Z_\mu Z^\mu\ . \tag{21.25}$$

We show below that $y_H = 1/2$. In order to provide more useful reference formulae we use this value going forward. Therefore, when we expand $r(x)$ around its vacuum expectation value, $r = v + h$, we find the mass terms

$$+m_W^2 W^+_\mu W^{-\mu} + \frac{m_Z^2}{2}Z_\mu Z^\mu\ , \tag{21.26}$$

where

$$m_W = \frac{gv}{2}\ , \qquad m_Z = \frac{v\sqrt{g^2 + g'^2}}{2}\ . \tag{21.27}$$

Thus, we see that these massive vectors ate the Nambu-Goldstone bosons. The complex spin-1 field $W^\pm_\mu$ eats two and the real spin-1 field Z_μ eats one.

Remember that we started with four massless gauge fields and have only found three that are massive. One massless combination of the gauge bosons should remain. Notice that we can write

$$Z_\mu = \cos\theta_W\, A^3_\mu - \sin\theta_W\, B_\mu\ , \tag{21.28}$$

where we have used the definitions

$$\sin\theta_W = \frac{g'}{\sqrt{g^2+g'^2}} \ , \qquad \cos\theta_W = \frac{g}{\sqrt{g^2+g'^2}} \ . \tag{21.29}$$

Exercise 21.3 Show that $\sin\theta_W$ and $\cos\theta_W$ has the expected behavior of trigonometric functions. That is, show

$$|\sin\theta_W| \leq 1 \ , \qquad |\cos\theta_W| \leq 1 \ , \qquad \sin^2\theta_W + \cos^2\theta_W = 1 \ . \tag{21.30}$$

The orthogonal combination of B_μ and A^3_μ is

$$A_\mu = \sin\theta_W \, A^3_\mu + \cos\theta_W \, B_\mu \ . \tag{21.31}$$

This combination is orthogonal in the sense that

$$(\sin\theta_W, \cos\theta_W) \begin{pmatrix} \cos\theta_W \\ -\sin\theta_W \end{pmatrix} = 0 \ . \tag{21.32}$$

The A_μ field remains massless and is understood to be the gauge field of $\mathrm{U}(1)_{\mathrm{E\&M}}$. A useful way to understand how these fields are related is

$$\begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & -\sin\theta_W \\ \sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} A^3_\mu \\ B_\mu \end{pmatrix} \ . \tag{21.33}$$

Observe that the matrix that relates the two sets of vectors is orthogonal. We see that these are just two bases for describing these degrees of freedom. We might call the A^3_μ, B_μ basis the gauge basis because in that basis the underlying gauge symmetry is clear. And we call the Z_μ, A_μ basis the mass basis because the basis elements correspond to the mass eigenstates. This shows that what we call the photon is a nontrivial mixture of a part of the $\mathrm{SU}(2)_L$ force and the $\mathrm{U}(1)_Y$ force.

Exercise 21.4 From the definitions made above, show that

$$A^1_\mu = \frac{1}{\sqrt{2}}\left(W^+_\mu + W^-_\mu\right) \ , \qquad A^2_\mu = \frac{i}{\sqrt{2}}\left(W^+_\mu - W^-_\mu\right) \ , \tag{21.34}$$

and

$$A^3 = \cos\theta_W Z_\mu + \sin\theta_W A_\mu \ , \qquad B_\mu = \cos\theta_W A_\mu - \sin\theta_W Z_\mu \ . \tag{21.35}$$

Exercise 21.5 Using the results of the previous exercise show that the covariant deriva-

tive acting on an arbitrary field X can be written

$$\begin{aligned} D_\mu X &= \left(\partial_\mu + igT^a A^a_\mu + ig' y_X B_\mu\right) X \\ &= \Bigg[\partial_\mu + i\frac{g}{\sqrt{2}}(T^1 + iT^2)W^+_\mu + i\frac{g}{\sqrt{2}}(T^1 - iT^2)W^-_\mu \\ &\quad + i\left(g\cos\theta_W T^3 - g'\sin\theta_W y_X\right) Z_\mu \\ &\quad + i\frac{gg'}{\sqrt{g^2+g'^2}}\left(T^3 + y_X\right) A_\mu\Bigg] X \ , \end{aligned} \tag{21.36}$$

where the T^a are the generators of $\mathrm{SU}(2)_L$ appropriate to the representation of X and y_X is the $\mathrm{U}(1)_Y$ charge of X.

Subsection 21.1

Charges and Couplings

We can compare the form of the covariant derivative given in the above exercise with what we found in QED

$$D_\mu = \partial_\mu + ieqA_\mu \ . \tag{21.37}$$

This allows us to identify the electric coupling as

$$e = \frac{gg'}{\sqrt{g^2+g'^2}} = g\sin\theta_W \ . \tag{21.38}$$

This result underscores again that electricity and magnetism arises from a combination of $\mathrm{SU}(2)_L$, which has the g gauge coupling, and $\mathrm{U}(1)_Y$, which has the g' gauge coupling. The electric charge is

$$q = T^3 + y \ . \tag{21.39}$$

That is, what we measure as electric charge is a combination of a field's hyper charge y_X and what representation of $\mathrm{SU}(2)_L$ it transforms under. For example, a field in the trivial representation of $\mathrm{SU}(2)_L$ has $T^3 = 0$. For such fields their electric charge is equal to their hypercharge, $q = y$.

Example

Let us consider the Higgs field

$$H = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \ . \tag{21.40}$$

This field has a hypercharge of $y_H = 1/2$ and is in the fundamental

representation of $\mathrm{SU}(2)_L$. This implies that

$$T^3 = \frac{1}{2}\sigma_3 = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} . \tag{21.41}$$

For the field ϕ^+ which is in the upper slot of the doublet we have

$$\frac{1}{2}\sigma^3 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 \\ 0 \end{pmatrix} , \tag{21.42}$$

so we say that ϕ^+ has a T^3 value of $1/2$. On the other hand, ϕ^0 is the lower slot of the doublet so

$$\frac{1}{2}\sigma^3 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = -\frac{1}{2}\begin{pmatrix} 0 \\ 1 \end{pmatrix} . \tag{21.43}$$

This shows that ϕ^0 has a T^3 value of $-1/2$.

Putting this together we find that the electric charge q of ϕ^+ is

$$T^3 + y_H = \frac{1}{2} + \frac{1}{2} = 1 , \tag{21.44}$$

while the electric charge of ϕ^0 is

$$T^3 + y_H = -\frac{1}{2} + \frac{1}{2} = 0 . \tag{21.45}$$

This shows that superscript labels of these fields are simply their electric charges.

Having understood the A_μ coupling in the covariant derivative (21.36), we now turn to the Z_μ couplings

$$\partial_\mu + i\left(g\cos\theta_W T^3 - g'\sin\theta_W y_X\right) Z_\mu . \tag{21.46}$$

We identified the electric charge $q = T^3 + y$, which is typically a more familiar quantity. This identification allows us to rewrite the Z coupling as

$$\begin{aligned} g\cos\theta_W T^3 - g'\sin\theta_W y_X =& g\cos\theta_W T^3 - g'\sin\theta_W \left(q - T^3\right) \\ =& T^3\left(g\cos\theta_W + g'\sin\theta_W\right) - qg'\sin\theta_W \\ =& T^3 \frac{g^2 + g'^2}{\sqrt{g^2+g'^2}} - qg'\sin\theta_W \\ =& \sqrt{g^2+g'^2}\left(T^3 - q\sin^2\theta_W\right) . \end{aligned} \tag{21.47}$$

In summary, we find the Z_μ couples to fermions through

$$\partial_\mu + i\frac{g}{\cos\theta_W}\left(T^3 - q\sin^2\theta_W\right)Z_\mu \ , \tag{21.48}$$

which depends on only the $\mathrm{SU}(2)_L$ gauge coupling g, the electric charge q, the $\mathrm{SU}(2)_L$ representation T^3, and $\sin\theta_W$. We see that the Z_μ couples to left-handed fermions through T^3 and q. However, it also couples to right-handed fermions if they have an electric charge, even through all the known right-handed particles have $T^3 = 0$. In other words, because the Z_μ is a mixture of A^3_μ and B_μ it couples to both left-handed particles (with nonzero T^3) and right-handed particles (with $T^3 = 0$).

Finally, we consider the couplings of the $W^\pm_\mu$ field. This coupling, from Eq. (21.36), is

$$\partial_\mu + i\frac{g}{\sqrt{2}}\left(T^1 + iT^2\right)W^+_\mu + i\frac{g}{\sqrt{2}}\left(T^1 - iT^2\right)W^-_\mu \ . \tag{21.49}$$

Unlike the Z, the $W^\pm$ is only known to couple to left-handed fermions because all known right-handed fermions have $T^1 = T^2 = 0$. Recalling the form of the raising and lowering operators (39.51) of the $\mathfrak{su}(2)$ Lie algebra we see that the $W^\pm_\mu$ interactions take the form

$$\partial_\mu + i\frac{g}{\sqrt{2}}T^+W^+_\mu + i\frac{g}{\sqrt{2}}T^-W^-_\mu \ . \tag{21.50}$$

Remember that these $T^\pm$ operators take states with T^3 of j to states with T^3 of $j \pm 1$. This also means that a state with electric charge $q = T^3 + y$ is taken to $q \pm 1$. In other words, the $W^\pm_\mu$ field has electric charge ± 1.

Example

Consider the action of the $W^\pm_\mu$ part of the gauge covariant derivative on the Higgs field

$$H = \begin{pmatrix}\phi^+\\ \phi^0\end{pmatrix} \ , \tag{21.51}$$

which is in the fundamental representation of $\mathrm{SU}(2)_L$. In this case

$$T^+ = \frac{1}{2}(\sigma_1 + i\sigma_2) = \begin{pmatrix}0&1\\0&0\end{pmatrix} \ , \quad T^- = \frac{1}{2}(\sigma_1 - i\sigma_2) = \begin{pmatrix}0&0\\1&0\end{pmatrix} \ . \tag{21.52}$$

This means that

$$\begin{aligned}\frac{g}{\sqrt{2}}\left(T^+W^+_\mu + T^-W^-_\mu\right)H &= \frac{g}{\sqrt{2}}\begin{pmatrix}0 & W^+_\mu\\ W^-_\mu & 0\end{pmatrix}\begin{pmatrix}\phi^+\\ \phi^0\end{pmatrix}\\ &= \frac{g}{\sqrt{2}}\begin{pmatrix}W^+_\mu\phi^0\\ W^-_\mu\phi^+\end{pmatrix} \ .\end{aligned} \tag{21.53}$$

To match electric charges of the derivative term in the full covariant derivative the upper term should have electric charge of one and the lower term and electric charge of zero. This only follows if the electric charge of $W_\mu^\pm$ is ± 1.

Section 22

Beta Decay

Now that we have determined the couplings of the electroweak gauge bosons we are ready to understand the process of radioactive beta decay

$$n \to p + e^- + \overline{\nu}_e \ . \tag{22.1}$$

Historically it was determined that this process always produces left-handed electrons and left-handed neutrinos. One might suppose that the leading order process is described by a Feynman diagram like Fig. 23. This, in turn, would seem to arise from a Lagrangian interaction like

$$\frac{4G_F}{\sqrt{2}} \overline{\psi}_n \gamma^\mu P_L \psi_p \, \overline{\psi}_{\nu_e} \gamma_\mu P_L \psi_e \tag{22.2}$$

This normalization of the coupling is related to the factor of 1/2 included in each P_L projector. Note also that the factors of γ^μ ensure that only left-handed fields are involved. That is, recall that

$$\overline{\psi} P_L \psi = \overline{\psi}_R \psi_L \ , \qquad \overline{\psi} \gamma^\mu P_L \psi = \overline{\psi}_L \gamma^\mu \psi_L \ . \tag{22.3}$$

In Sec. 14.1 it was pointed out that the interaction in (22.2) is nonrenormalizeable. Because each fermion field has dimension 3/2, the dimension of G_F must be -2. This nonrenormalizable theory can be used to make predictions, but it can only do so at energies well below the scale associated with G_F. Based on the γ^μs that appear in the interaction, we might suspect that a spin-1 field is involved. Because beta decay was also observed to be governed by a short-range force, it was correctly understood to result from the exchange of a massive gauge boson.

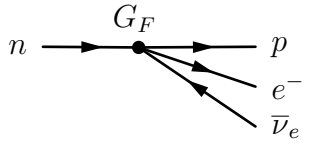


Figure 23. Schematic diagram of β-decay.

We now understand this process to be mediated by the W_μ^- boson. The valence quarks of the neutron are two down quarks and one up quark. The valence quarks of the proton are two up quarks and one down quark. If we gather the left-handed up- and down quarks into an $\mathrm{SU}(2)_L$ doublet

$$Q = \begin{pmatrix} u_L \\ d_L \end{pmatrix} \ , \tag{22.4}$$

then the interaction

$$\overline{Q} i\gamma^\mu D_\mu Q \ , \tag{22.5}$$

includes the $W^\pm_\mu$ terms

$$-\frac{g}{\sqrt{2}}\left(\overline{u}_L,\, \overline{d}_L\right)\gamma^\mu \begin{pmatrix} 0 & W^+_\mu \\ W^-_\mu & 0 \end{pmatrix}\begin{pmatrix} u_L \\ d_L \end{pmatrix} = -\frac{g}{\sqrt{2}}\left(\overline{u}_L W^+_\mu \gamma^\mu d_L + \overline{d}_L W^-_\mu \gamma^\mu u_L\right) \ . \tag{22.6}$$

This allows one of the neutron down quarks to become a W^-_μ and an up quark. Note that electric charge is conserved, $-1/3 = -1 + 2/3$.

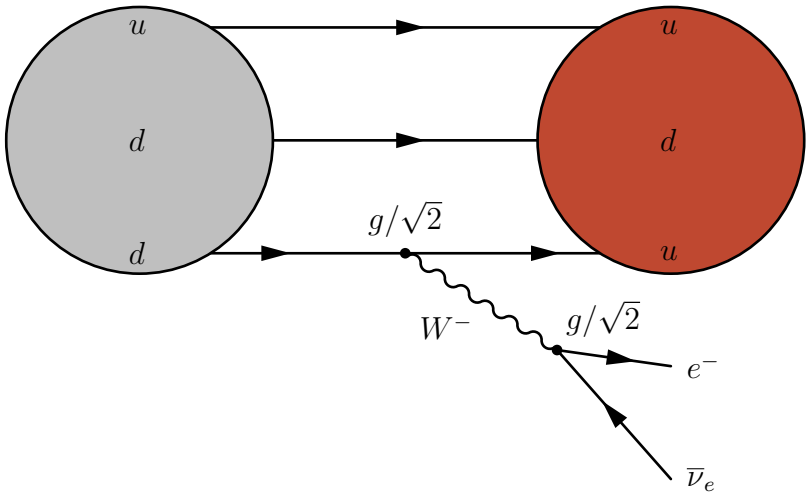


Figure 24. Schematic diagram of β-decay in terms of W boson and quarks. The neutron (udd) is shaded grey while the proton (uud) is shaded red.

However, the W^-_μ is a very heavy particle (about 80 GeV) and so is produced very off-shell. Consequently it must convey its energy and momentum into lighter, on-shell fields. The lightest fields composed of quarks are the pions, with masses near 140 MeV. But, the difference in the neutron and proton masses is just over 1 MeV, so an on-shell pion cannot be produced. The only particles light enough are the left-handed leptons.

These leptons have nonzero electric charge—the electron, the muon, and the tau—are each the lower component of an $\mathrm{SU}(2)_L$ doublet. The upper component in each doublet is a neutrino. So far, no right-handed companions to these neutrinos has been discovered. Of these three charged leptons only the electron is light enough to be produced in neutron decay.

Exercise 22.1 Using the methods employed above, show that the $W^\pm_\mu$ couples to an $\mathrm{SU}(2)_L$ doublet of leptons

$$L_e = \begin{pmatrix} \nu_e \\ e_L \end{pmatrix} \ , \tag{22.7}$$

as

$$-\frac{g}{\sqrt{2}}\left(\overline{\nu}_e W_\mu^+ \gamma^\mu e_L + \overline{e}_L W_\mu^- \gamma^\mu \nu_e\right) . \tag{22.8}$$

The interactions found in the above exercise allow the excess energy of changing a neutron into a proton to become an electron and an antineutrino, each with considerable kinetic energy. It is significant that among all the force carriers we have encountered the W boson is the first that connects particles of different type. That is, a photon couples to a particle and its antiparticle, as does the Z boson. The W couples an up quark to an antidown quark and the electron to an antineutrino.

Exercise 22.2 Using the fact that the electric charge of the electron is -1 and that $q = T^3 + y$ determine the hypercharge y of the left-handed lepton field L_e defined in (22.7). Then, determine the electric charge of ν_e.

The interactions of the W are also new in that they carry electric charge. This is what allows a neutron to become a proton, negative charge is carried out of the strong force bound state, leaving the positively charged proton behind. This is why the effects of the W are the ones that are associated with the "weak nuclear force" in popular science. You could also clam the Z contributes to this force, but its effects are somewhat similar to the photon, both are electrically neutral. The Z does contribute to many important phenomena, especially at high energy colliders where the Z can be close to on-shell while the photon is very off-shell.

The fact that the W fields are so off-shell during beta decay also allows us to understand the four-Fermi interaction in (22.2). Our high energy understanding of the interactions leads us to express the amplitude for leading order process as

$$\mathcal{M} \propto \frac{g}{\sqrt{2}}\overline{\psi}_{\nu_e L}\gamma^\mu \psi_e \frac{\eta_{\mu\nu}}{p^2 - m_W^2}\frac{g}{\sqrt{2}}\overline{\psi}_{dL}\gamma^\nu \psi_{uL} . \tag{22.9}$$

Here the two leptons are connected to the two quarks by the exchanged (the propagator of) the W boson. Because $p^2 \ll m_W^2$ we can expand the denominator of the propagator and find the approximate amplitude

$$\mathcal{M} \approx \frac{g^2}{2m_W^2}\overline{\psi}_{\nu_e L}\gamma^\mu \psi_e \overline{\psi}_{dL}\gamma_\mu \psi_{uL} . \tag{22.10}$$

This has the form of the four-Fermi interaction and makes the connection

$$G_F = \frac{g^2\sqrt{2}}{8m_W^2} . \tag{22.11}$$

This also illustrates an important general point. A nonrenormalizable theory can be used to describe nature, but not to arbitrarily high energies. At

higher energies a new particle appears (in this case the W) with a mass related to the scale of the nonrenormalizable coupling G_F.

The Standard Model

PART VI

Sec. 23. Fermion Masses
Sec. 24. Flavor & CKM
Sec. 25. Bound States & Flavor
Sec. 26. Symmetries

Section 23

Fermion Masses

In the preceding sections we described the $\mathrm{SU}(3)_c \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ gauge structure of the Standard Model (SM) of particle physics. In doing so, however, we have actually created a problem. When we discussed QED we considered the electron field

$$\psi_e = \begin{pmatrix} e_L \\ e_R \end{pmatrix} \ , \tag{23.1}$$

and mentioned that the mass term had the form

$$m_e \overline{\psi}_e \psi_e = m_e \left(\overline{e}_L e_R + \overline{e}_R e_L \right) \ . \tag{23.2}$$

But, when discussing beta decay we claimed that the $W^{\pm}_{\mu}$ couples to only the left-handed electron. But it must couple to a field in a nontrivial representation of $\mathrm{SU}(2)_L$, which turns out to be

$$L_e = \begin{pmatrix} \nu_e \\ e_L \end{pmatrix} \ . \tag{23.3}$$

The problem is that there seems to be no way write down a gauge invariant mass term for the electron, and the electron mass has been measured.

The way through this puzzle is to use the Higgs boson. Note that the operators

$$\lambda_\ell \overline{L}_{e\,j} H^j e_R + \lambda^*_\ell \overline{e}_R H^\dagger_j L^j_e \ , \tag{23.4}$$

are completely gauge invariant. The Higgs and L_e are in the **2** of $\mathrm{SU}(2)_L$, the two dimensions of which we have indexed like H^j. Under a gauge transformation these transform as

$$\begin{aligned} \overline{L}_{e\,j} H^j e_R + \overline{e}_R H^\dagger_j L^j_e \rightarrow & \overline{L}_{e\,k} U^{\dagger\,k}{}_j U^j{}_m H^m e_R + \overline{e}_R H^\dagger_k U^{\dagger\,k}{}_j U^j{}_m L^m_e \\ = & \overline{L}_{e\,j} H^j e_R + \overline{e}_R H^\dagger_j L^j_e \ . \end{aligned} \tag{23.5}$$

Under a $\mathrm{U}(1)_Y$ transformation each field X goes to $e^{-iy_X\theta}$. The Higgs has a hypercharge of $y_H = 1/2$. The right-handed electron has a hypercharge of $y_{eR} = -1$. The left-handed lepton field has a hypercharge of $y_{Le} = -1/2$

so that the electric charge of the left-handed electron is $-1/2 - 1/2 = -1$ and the electric charge of the neutrino is $1/2 - 1/2 = 0$. This means that under a $\mathrm{U}(1)_Y$ transformation we have

$$\begin{aligned}\overline{L}_{e\,j} H^j e_R + \overline{e}_R H^\dagger_j L^j_e \to & \overline{L}_{e\,j} H^j e_R e^{-i(y_{eR}+y_H-y_{Le})} + \overline{e}_R H^\dagger_j L^j_e e^{i(y_{eR}+y_H-y_{Le})} \\ = & \overline{L}_{e\,j} H^j e_R + \overline{e}_R H^\dagger_j L^j_e \ . \end{aligned} \tag{23.6}$$

So far, we have only shown that these operators in (23.4) are gauge invariant. Note that we must include both of them for the Lagrangian to be Hermitian. We have allowed λ_ℓ to be a complex number, which we can express as $r_\ell e^{i\theta_\ell}$. It is the case that

$$\overline{e}_R i\partial\!\!\!/ e_R + \overline{L}_{e\,j} i\partial\!\!\!/ L^j_e \ , \tag{23.7}$$

is invariant under the global rephasing

$$e_R \to e^{-i\theta_{eR}} e_R \ , \qquad L_e \to e^{-i\theta_{Le}} L_e \ . \tag{23.8}$$

But under this same rephasing choice

$$r_\ell e^{i\theta_\ell} \overline{L}_{e\,j} H^j e_R \to r_\ell e^{i(\theta_\ell+\theta_{Le}-\theta_{eR})} \overline{L}_{e\,j} H^j e_R \ . \tag{23.9}$$

So, no matter the value of θ_ℓ we can choose to rephase e_R and L_e such that

$$\theta_{eR} - \theta_{eL} = \theta_\ell \ . \tag{23.10}$$

For this choice the coefficient of $\overline{L}_{e\,j} H^j e_R$ is just the real number r_ℓ. We can always make this choice without affecting any other part of the theory, so we simply assume λ_ℓ is real.

Why are we spending all this time discussing these three field interaction terms when we claimed that we are interested in masses? When electroweak symmetry breaks then the Higgs field gets a vacuum expectation value

$$\langle H \rangle = \frac{v}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \ . \tag{23.11}$$

Neglecting (for now) the fluctuations around the VEV, this leads to

$$\frac{\lambda_\ell v}{\sqrt{2}} \left[(\overline{\nu}_e, \, \overline{e}_L) \begin{pmatrix} 0 \\ 1 \end{pmatrix} e_R + \overline{e}_R \, (0, \, 1) \begin{pmatrix} \nu_e \\ e_L \end{pmatrix} \right] = \frac{\lambda_\ell v}{\sqrt{2}} \left(\overline{e}_L e_R + \overline{e}_R e_L \right) \ . \tag{23.12}$$

This is exactly the form of a fermion mass term. We see that the Higgs, whose VEV already is tied to the mass of the W and Z bosons, can also be responsible for the lepton masses.

Exercise 23.1 For the left-handed quark field

$$Q_{ud} = \begin{pmatrix} u_L \\ d_L \end{pmatrix} \ , \tag{23.13}$$

determine the hypercharge of Q_{ud} such that the electric charges of u_L and d_L match those of their corresponding right-handed field, which are $2/3$ and $-1/3$, respectively. Then show that

$$\lambda_d \overline{Q}_{ud\,j} H^j d_R + \lambda_d^* \overline{d}_R H_j^\dagger Q_{ud}^j \ , \tag{23.14}$$

is gauge invariant and Hermitian. Argue that one can always choose λ_d to be real and find the mass term for down quarks when H is set to its VEV.

The Higgs can be used to provide masses, after electroweak symmetry breaking, to the leptons and down quarks. What about the up quark? We cannot write down

$$\overline{Q}_{ud\,j} H^j u_R \ , \tag{23.15}$$

because this is not invariant under $\mathrm{U}(1)_Y$ transformations. The left-handed quarks have hypercharge $y_Q = 1/6$ and the right-handed up quark has hypercharge $y_{uR} = 2/3$ so $y_{uR} - y_Q = 1/2$. This means we need a field with hypercharge of $-1/2$ to make a $\mathrm{U}(1)_Y$ invariant combination. The field $H^\dagger$ has this hypercharge, but under $\mathrm{SU}(2)_L$ transformations

$$\overline{Q}_{ud\,j} H_j^\dagger \to \overline{Q}_{ud\,k} U^{\dagger\,k}{}_j H_m^\dagger U^{\dagger\,m}{}_j \ , \tag{23.16}$$

which is not $\mathrm{U}(2)_L$ invariant. There is, however, one other option. Recall that for any two-by-two matrix U that

$$\varepsilon_{ij} U^i{}_m U^j{}_n = \mathrm{Det}(U) \varepsilon_{mn} \ . \tag{23.17}$$

The fundamental representation Us have a determinant of one. This means that

$$\varepsilon_{ij} H^i Q_{ud}^j \to \varepsilon_{ij} U^i{}_m H^m U^j{}_n Q_{ud}^n = \varepsilon_{mn} H^m Q_{ud}^n \ , \tag{23.18}$$

is a gauge invariant combination. This means that

$$\varepsilon^{ij} \overline{Q}_{ud\,i} H_j^\dagger u_R + \varepsilon_{ij} \overline{u}_R H^i Q_{ud}^j \ , \tag{23.19}$$

is completely gauge invariant. Often, the notation

$$\widetilde{H}^i = \varepsilon^{ij} H_j^\dagger = \begin{pmatrix} \phi^{0*} \\ -\phi^{+*} \end{pmatrix} \ , \tag{23.20}$$

is used and all the $\mathrm{SU}(2)_L$ indices are suppressed. The full fermion-Higgs

couplings are then

$$-\lambda_\ell \overline{L}_e H e_R - \lambda_d \overline{Q}_{ud} H d_R - \lambda_u \overline{Q}_{ud} \widetilde{H} u_R + \text{H.c} \ , \tag{23.21}$$

where H.c denotes the Hermitian conjugate.

Exercise 23.2 Show that by making global rephasings of the quark fields one can ensure that both λ_d and λ_u are real numbers.

When the Higgs is set to its VEV this leads to the mass terms

$$-\frac{\lambda_\ell v}{\sqrt{2}} \left(\overline{e}_L e_R + \overline{e}_R e_L\right) - \frac{\lambda_d v}{\sqrt{2}} \left(\overline{d}_L d_R + \overline{d}_R d_L\right) - \frac{\lambda_u v}{\sqrt{2}} \left(\overline{u}_L u_R + \overline{u}_R u_L\right) \ . \tag{23.22}$$

In this way the dynamics of the Higgs field leads to masses for both the W and Z bosons as well as the fermions of the Standard Model.

Section 24

Flavor and the CKM Matrix

Up to this point we have outlined the full gauge symmetry structure of the Standard Model, including how the spontaneous breaking of that symmetry produces masses for the W, Z, electron, up quark, and down quark. It can be sometimes useful to describe the fermion fields in terms of their representations under the gauge symmetry. For example, we might write

$$e_R \sim (\mathbf{1}, \mathbf{1}, -1) \ , \tag{24.1}$$

to convey that the right-handed electron field is in the singlet representation of $\text{SU}(3)_c$, the singlet representation of $\text{SU}(2)_L$, and has a hypercharge of -1. The other fermion fields are described by

$$L_e \sim \left(\mathbf{1}, \mathbf{2}, -\frac{1}{2}\right) \ , \quad Q_{ud} \sim \left(\mathbf{3}, \mathbf{2}, \frac{1}{6}\right) \ , \quad u_R \sim \left(\mathbf{3}, \mathbf{1}, \frac{2}{3}\right) \ , \quad d_R \sim \left(\mathbf{3}, \mathbf{1}, -\frac{1}{3}\right) \ . \tag{24.2}$$

This gauge structure among these fields actually exists in three copies. The muon and the tau along with their associated neutrinos have the same gauge structure as the electron

$$\mu_R \sim (\mathbf{1}, \mathbf{1}, -1) \ , \qquad L_\mu \sim \left(\mathbf{1}, \mathbf{2}, -\frac{1}{2}\right) \ , \tag{24.3}$$

$$\tau_R \sim (\mathbf{1}, \mathbf{1}, -1) \ , \qquad L_\tau \sim \left(\mathbf{1}, \mathbf{2}, -\frac{1}{2}\right) \ . \tag{24.4}$$

There are also two more copies of the up and down quark structure. The charm and top quarks are like the up quarks, while the strange and bottom

quarks are like the down quark:

$$Q_{cs} \sim \left(\mathbf{3},\, \mathbf{2},\, \frac{1}{6}\right) , \quad c_R \sim \left(\mathbf{3},\, \mathbf{1},\, \frac{2}{3}\right) , \quad s_R \sim \left(\mathbf{3},\, \mathbf{1},\, -\frac{1}{3}\right) , \tag{24.5}$$

$$Q_{tb} \sim \left(\mathbf{3},\, \mathbf{2},\, \frac{1}{6}\right) , \quad t_R \sim \left(\mathbf{3},\, \mathbf{1},\, \frac{2}{3}\right) , \quad b_R \sim \left(\mathbf{3},\, \mathbf{1},\, -\frac{1}{3}\right) . \tag{24.6}$$

There is some whimsical vocabulary associated with these copies of the same structure. The electron, muon, tau, and the three neutrinos are all called **Leptons**. The three types are called **Flavors**, as in electron flavored or tau flavored. Similarly, the quarks come in six flavors, up, charm, top, down, strange, and bottom. While there is a lot that is the same about these particles, they do have different masses.

	1st Generation	2nd Generation	3rd Generation
Name	electron	muon	tau
Mass	0.511	106	1777
Name	up quark	charm quark	top quark
Mass	2.16	1270	172000
Name	down quark	strange quark	bottom quark
Mass	4.67	93.4	4180

Table 2. Table of fermion masses in MeV. For more precise values and uncertainties see [20].

As shown in Table 2, a set of particles (one lepton, one quark with electric charge of $2/3$, on quark with electric charge of $-1/3$) grouped by mass is called a generation. There are three generations of particles in the Standard Model and the origin of their very different masses is not known.

Particles of similar type can be grouped together. The right-handed particles of the three generation are organized into flavor triplets

$$U = \begin{pmatrix} u_R \\ c_R \\ t_R \end{pmatrix} , \quad D = \begin{pmatrix} d_R \\ s_R \\ b_R \end{pmatrix} , \quad E = \begin{pmatrix} e_R \\ \mu_R \\ \tau_R \end{pmatrix} , \tag{24.7}$$

and the left-handed fields as

$$L = \begin{pmatrix} L_e \\ L_\mu \\ L_\tau \end{pmatrix} , \quad Q = \begin{pmatrix} Q_{ud} \\ Q_{cs} \\ Q_{tb} \end{pmatrix} . \tag{24.8}$$

With these definitions we can write down the full Standard Model La-

grangian

$$\begin{aligned}\mathcal{L} = &-\frac{1}{4}G^a_{\mu\nu}G^{a\mu\nu} - \frac{1}{4}W^a_{\mu\nu}W^{a\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} \\ &+ i\overline{Q}\not{D}Q + i\overline{L}\not{D}L + i\overline{D}\not{D}D + i\overline{U}\not{D}U + i\overline{E}\not{D}E \\ &+ (D_\mu H)^\dagger D^\mu H + \mu^2|H|^2 - \lambda|H|^4 \\ &+ (\lambda_\ell)^i{}_j\,\overline{L}_i H E^j + (\lambda_d)^i{}_j\,\overline{Q}_i H D^j + (\lambda_u)^i{}_j\,\overline{Q}_i\widetilde{H}U^j + \text{ H.c. }. \end{aligned} \tag{24.9}$$

The top line describes the gauge fields for each of the gauge symmetries. The second line is the kinetic energy terms for each of the fermions. Note that these are in terms of the flavor triplets. The third line describes the Higgs field's interactions with the gauge fields and itself. The last line gives the interactions of the Higgs field and the fermion fields.

This final line is also responsible for all producing all elementary fermion masses. We also emphasize that the Yukawa couplings, λ_ℓ etc, are matrices in the flavor spaces of the different fields. The different entries in these matrices, multiplied by the Higgs vacuum expectation value, give rise to the variety of different fermion masses.

What kind of matrices are $(\lambda_\ell)^i{}_j$, $(\lambda_d)^i{}_j$, and $(\lambda_u)^i{}_j$? In general, they are arbitrary complex matrices. However, we can choose bases in which they are diagonal.

Example

Let us consider the leptons. When the Higgs has a nonzero VEV the leptons have the mass terms

$$\left(\overline{L}_1\,,\,\overline{L}_2\,,\,\overline{L}_3\right) M_\ell \begin{pmatrix} E^1 \\ E^2 \\ E^3 \end{pmatrix} + \left(\overline{E}_1\,,\,\overline{E}_2\,,\,\overline{E}_3\right) M_\ell^\dagger \begin{pmatrix} L^1 \\ L^2 \\ L^3 \end{pmatrix}\,, \tag{24.10}$$

where the mass matrix is related to the lepton Yukawa matrix by

$$\frac{v}{\sqrt{2}}\,(\lambda_\ell)^i{}_j \equiv (M_\ell)^i{}_j\ . \tag{24.11}$$

Note that we have labeled the fields by numbers deliberately.

The only other parts of the Lagrangian that depend on the lepton fields are the kinetic terms

$$\overline{L}i\not{D}L + \overline{E}i\not{D}E\ . \tag{24.12}$$

It is simple to see that this part of the Lagrangian has a $\mathrm{U}(3)_E\times\mathrm{U}(3)_L$ global symmetry. That is, if we consider the transformations

$$E \to UE\,,\quad L \to VL\,, \tag{24.13}$$

where U and V are unitary matrices then

$$\begin{aligned}\overline{L}i\not{D}L + \overline{E}i\not{D}E \rightarrow & \overline{L}V^\dagger i\not{D}VL + \overline{E}U^\dagger i\not{D}UE \\ = & \overline{L}i\not{D}L + \overline{E}i\not{D}E \ . \end{aligned} \tag{24.14}$$

The last equality follows as the covariant derivatives are simply identity operators in the flavor spaces of these fields. Under this same transformation the lepton mass terms become

$$\overline{L}M_\ell E + \overline{E}M_\ell^\dagger L \rightarrow \overline{L}V^\dagger M_\ell U E + \overline{E}U^\dagger M_\ell^\dagger V L \ . \tag{24.15}$$

As we shortly show, any complex matrix can be diagonalized by using two unitary matrices

$$D = U^\dagger M V \ . \tag{24.16}$$

So, no matter what the lepton Yukawa matrix is we can make a transformation in the flavor space of the leptons such that the mass matrix is diagonal. It is this diagonal basis that we associate with the measured lepton masses.

The above process of diagonalization is sometimes called biunitary diagonalization. We sketch the process in what follows.

Exercise 24.1 Suppose that M is an arbitrary complex matrix. Show that both $MM^\dagger$ and $M^\dagger M$ are Hermitian matrices.

Given some arbitrary complex matrix M we note that $MM^\dagger$ is Hermitian. However, all Hermitian matrices can be diagonalized by some unitary matrix, which we denote U. That is

$$U^\dagger M M^\dagger U = D \ , \tag{24.17}$$

where D is a diagonal matrix with real, non-negative entries. Because each entry is a non-negative real number we can take their square-root to make the diagonal matrix d where

$$dd = D \ . \tag{24.18}$$

Now, it could be that one of the diagonal entires is a zero, in which case d does not have an inverse. But, we could separate out the part of the space with the zero eigenvalue (or eigenvalues if there are more than one). On the remaining space d does have an inverse. For simplicity, we simply assume that none of the eigenvalues are zero. Under this assumption d^{-1} is the diagonal matrix whose entries are the inverse of the corresponding

entries of d. We can then define the matrix V by

$$V = M^\dagger U d^{-1} \ . \tag{24.19}$$

Exercise 24.2 Using the above definition of V find $V^\dagger$ and show that $V^\dagger V = \mathbb{I}$. That is, show that V is a unitary matrix. Then, show that

$$d = U^\dagger M V \ . \tag{24.20}$$

This shows that an arbitrary complex matrix M can be diagonalized by unitary matrices U and V.

It is worth emphasizing that particles propagate though spacetime in the mass basis, or the basis in which their mass matrix is diagonal. However, they might have diagonal interactions with other fields in a different basis. The misalignment between mass and interaction basis can lead to interesting experimental signals. For instance, in the Standard Model we have shown that we can diagonalize the lepton mass matrix without breaking the flavor symmetry of the lepton kinetic terms, such as $\overline{E} i \not{D} E$. This means that the interactions the arise from the kinetic terms, like all the interactions with gauge bosons, preserve the mass basis eigenstates. In other words, gauge interactions take electrons to electrons, muons to muons, and taus to taus.

This result may seem uninteresting, but we should contrast it with the quark case. The quark Yukawa matrices

$$(\lambda_d)^i{}_j \, \overline{Q}_i H D^j + (\lambda_u)^i{}_j \, \overline{Q}_i \widetilde{H} U^j + \text{ H.c. } . \tag{24.21}$$

Using the notation

$$Q = \begin{pmatrix} U_L \\ D_L \end{pmatrix} \ , \tag{24.22}$$

where U_L is three dimensional vector of the left-handed up-type quark fields and D_L is three dimensional vector of the left-handed down-type quark fields. When the Higgs is expanded about it VEV we find the mass matrices

$$\overline{D}_L M_d D + \overline{U}_L M_u U + \text{H.c. } . \tag{24.23}$$

There is also a $\mathrm{U}(3)_Q \times \mathrm{U}(3)_U \times \mathrm{U}(3)_D$ flavor symmetry of the quark kinetic terms.

Exercise 24.3 Show that the quark kinetic terms

$$\overline{Q} i \not{D} Q + \overline{U} i \not{D} U + \overline{D} i \not{D} D \ , \tag{24.24}$$

are invariant under

$$Q \to U_Q Q \ , \quad U \to U_U U \ , \quad D \to U_D D \ , \tag{24.25}$$

where U_Q, U_U, and U_D are unitary matrices acting on the flavor space of the quark fields.

The question is, can we diagonalize both of the quark mass terms with these transformations? We find

$$\overline{D}_L M_d D + \overline{U}_L M_u U \to \overline{D}_L U_Q^\dagger M_d U_D D + \overline{U}_L U_Q^\dagger M_u U_U U \ , \tag{24.26}$$

where both D_L and U_L are rotated by the same matrix because they are both part of the left-handed quark field Q. We see immediately that there is a problem. We could choose U_Q and U_D to diagonalize M_d, but then we only have the freedom of U_U to try and diagonalize M_u. In general, this is not enough freedom to diagonalize both matrices. In order to be guaranteed to diagonalize both we must take

$$Q = \begin{pmatrix} U_L \\ D_L \end{pmatrix} \to \begin{pmatrix} U_u U_L \\ U_d D_L \end{pmatrix} \neq U_Q \begin{pmatrix} U_L \\ D_L \end{pmatrix} \ . \tag{24.27}$$

In other words, to diagonalize both quark matrices we must break the flavor symmetry of the kinetic terms.

What consequence does this have? It can only affect the left-handed quark fields as this is the flavor symmetry we are breaking. This kinetic terms has the schematic form

$$\overline{Q} i \not{D} Q = \tag{24.28}$$
$$\left(\overline{U}_L \, , \, \overline{D}_L\right) \begin{pmatrix} i\not{\partial} - g_s \not{G} - e q_u \not{A} - g_z z_u \not{Z} & -g_W \not{W}^+ \\ -g_W \not{W}^- & i\not{\partial} - g_s \not{G} - e q_d \not{A} - g_z z_d \not{Z} \end{pmatrix} \begin{pmatrix} U_L \\ D_L \end{pmatrix} \ ,$$

where $g_z z_{u,d}$ stands in for the specific coupling of the Z boson to the different fields. Under the unitary transformations

$$U_L \to U_u U_L \ , \quad D_L \to U_d D_L \ , \tag{24.29}$$

we find

$$\overline{Q} i \not{D} Q = \tag{24.30}$$
$$\left(\overline{U}_L \, , \, \overline{D}_L\right) \begin{pmatrix} i\not{\partial} - g_s \not{G} - e q_u \not{A} - g_z z_u \not{Z} & -g_W \not{W}^+ U_u^\dagger U_d \\ -g_W \not{W}^- U_d^\dagger U_u & i\not{\partial} - g_s \not{G} - e q_d \not{A} - g_z z_d \not{Z} \end{pmatrix} \begin{pmatrix} U_L \\ D_L \end{pmatrix} \ .$$

This shows that that only effect of this flavor violation appears in the

W_μ interactions. This occurs through the matrix

$$V_{\rm CKM} = U_u^\dagger U_d \ , \tag{24.31}$$

where CKM stands for Cabibbo, Kobayashi, and Maskawa. Nicola Cabibbo developed a similar matrix for two quark generations [45], which was generalized to three by Makoto Kobayashi and Toshihide Maskawa [46]. Note that

$$V_{\rm CKM}^\dagger = U_d^\dagger U_u \ , \tag{24.32}$$

so

$$V_{\rm CKM}^\dagger V_{\rm CKM} = U_d^\dagger U_u U_u^\dagger U_d = \mathbb{I} = U_u^\dagger U_d U_d^\dagger U_u = V_{\rm CKM} V_{\rm CKM}^\dagger \ , \tag{24.33}$$

so the CKM matrix is predicted to be unitary.

The CKM matrix allows the W interactions to mix the quark generations. Remember that these generations are defined in terms of the quark masses, so the notion of generation is a mass basis concept. But in this basis the W interactions are not diagonal, so they mix the generations. We see this through

$$\begin{aligned} &\frac{g}{\sqrt{2}} \overline{U}_L \not{W}^+ V_{\rm CKM} D_L \\ =&\frac{g}{\sqrt{2}} \left(\overline{u}_L \ , \ \overline{c}_L \ , \ \overline{t}_L\right) \not{W}^+ \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d_L \\ s_L \\ b_L \end{pmatrix} \ . \end{aligned} \tag{24.34}$$

The off diagonal terms of the matrix allow for generation mixing. The magnitudes of all the CKM matrix components have been obtained experimentally [20]

$$|V|_{\rm CKM} = \begin{pmatrix} 0.97373 \pm 0.00031 & 0.2243 \pm 0.0008 & 0.00382 \pm 0.0002 \\ 0.221 \pm 0.004 & 0.975 \pm 0.006 & 0.0408 \pm 0.0014 \\ 0.0086 \pm 0.0002 & 0.0415 \pm 0.0009 & 1.014 \pm 0.029 \end{pmatrix} \ . \tag{24.35}$$

Note that the diagonal terms are close to one, the off-diagonal terms connecting the 1st and 2nd generations have values closer to 1/10 while the mixing of the 2nd and 3rd generations is nearer 1/100. The mixing of the 1st and 3rd generations mixing is even smaller, a bit more than 1/1000. This seems to indicate some kind of structure in the quark flavors. As of yet, we have no confirmed explanation of the origin of these measured values.

Subsection 24.1

Weak CP-Violation

The fact that there are three generations leads to a very interesting consequence. Specifically, it allows for CP violation in the W interactions with quarks. To see why this happens for three generations let us first suppose there was only one generation. In this case we would have

$$\frac{g}{\sqrt{2}}\overline{U}_L \not{W}^+ V_{\rm CKM} D_L \ , \tag{24.36}$$

where $V_{\rm CKM}$ is simply a unitary matrix of dimension one, or in other words a complex phase $e^{-i\theta_{\rm CKM}}$. Then, by making the transformation

$$\overline{U}_L \to \overline{U}_L \ , \qquad D_L \to e^{i\theta_{\rm CKM}} D_L \ , \tag{24.37}$$

the interaction becomes $e^0 = 1$. This simple interaction is real and consequently does not lead to CP violation. (See Sec. 15.2 for a review of real couplings and CP violation.)

Next, let us consider the case of two generations. Historically, this case was thought to be the correct model of nature, by some at least, until the particles of the third generation were discovered. In this case the interaction looks like

$$\frac{g}{\sqrt{2}}\overline{U}_L \not{W}^+ V D_L = \frac{g}{\sqrt{2}} (\overline{u}_L \, , \, \overline{c}_L) \not{W}^+ \begin{pmatrix} V_{ud} & V_{us} \\ V_{cd} & V_{cs} \end{pmatrix} \begin{pmatrix} d_L \\ s_L \end{pmatrix} \ . \tag{24.38}$$

If $V_{\rm CKM}$ were an arbitrary matrix then it would be described by eight real parameters

$$V_{ud} = r_{ud} e^{-i\theta_{ud}} \ , \quad V_{us} = r_{us} e^{-i\theta_{us}} \ , \quad V_{cd} = r_{cd} e^{-i\theta_{cd}} \ , \quad V_{cs} = r_{cs} e^{-i\theta_{cs}} \ . \tag{24.39}$$

The unitarity of $V_{\rm CKM}$, however, implies that

$$V V^\dagger = V^\dagger V = \mathbb{I}_2 \ . \tag{24.40}$$

Exercise 24.4

Show that Eq. (24.40) implies

$$r_{ud}^2 + r_{us}^2 = 1 \ , \quad r_{cd}^2 + r_{cs}^2 = 1 \ , \quad r_{ud} r_{cd} e^{i(\theta_{cd} - \theta_{ud})} = - \, r_{us} r_{cs} e^{i(\theta_{cs} - \theta_{us})} \ , \tag{24.41}$$

$$r_{ud}^2 + r_{cd}^2 = 1 \ , \quad r_{us}^2 + r_{cs}^2 = 1 \ , \quad r_{ud} r_{us} e^{i(\theta_{us} - \theta_{ud})} = - \, r_{cd} r_{cs} e^{i(\theta_{cs} - \theta_{cd})} \ . \tag{24.42}$$

How many independent constraints does this put on the phases of the matrix components?

We see from the exercise that the magnitudes matrix components can

be specified by one angle

$$r_{ud} = r_{cs} = \cos\psi \ , \qquad r_{us} = -r_{cd} = \pm\sin\psi \ , \tag{24.43}$$

while the four phases must satisfy

$$\theta_{cd} + \theta_{us} = \theta_{ud} + \theta_{cs} \ . \tag{24.44}$$

Note that even though there were two off-diagonal components this only leads to one phase constraint. This one constraint on four phases seems to imply that there are complex couplings and hence CP violation in this two generation model. However, we can make phase rotations of the quark fields

$$u_L \to e^{-i\theta_u} u_L \ , \quad d_L \to e^{-i\theta_d} d_L \ , \quad c_L \to e^{-i\theta_c} c_L \ , \quad s_L \to e^{-i\theta_s} s_L \ . \tag{24.45}$$

Exercise 24.5

Show that the quark rephasing in Eq. (24.45) is equivalent to

$$\begin{aligned} \theta_{ud} &\to \theta_{ud} + \theta_d - \theta_u \ , & \theta_{us} &\to \theta_{us} + \theta_s - \theta_u \ , \\ \theta_{cd} &\to \theta_{cd} + \theta_d - \theta_c \ , & \theta_{cs} &\to \theta_{cs} + \theta_s - \theta_c \ . \end{aligned} \tag{24.46}$$

Show that Eq (24.44) is preserved, or in other words $V_{\rm CKM}$ remains unitary.

Notice that the phases of the matrix are modified by differences of the quark rephasing angles. This can be understood in following way. If all the quark fields were rephased by the same angle then the interaction is completely unaffected as there are as many quark and anti-quark fields. This means that we can only remove three phases from the matrix, but the constraint in Eq. (24.44) shows that there are only three phases to remove. The end result is that by rephasing the quark fields one can make the two generation $V_{\rm CKM}$ completely real. Consequently, there is not CP violation in this case.

What about the three-generation case? For two generations there were four phases in the matrix and one constraint (from one off-diagonal unitarity relationship) leaving three independent phases. By rephasing four quark fields we can remove three phases and hence have no physical phase and no CP violation. In the case of three generations we have a three-by-three matrix which would in general have nine phases. However, unitarity produces three constraints on these phases (from three off-diagonal relationships) leaving six phases in the matrix. We have six quark fields that we can rephase, but this only allows us to remove five of the phases. Therefore, there is one physical phase in the three-generation CKM matrix. This phase does produce CP violation in the interactions of the W boson.

Exercise 24.6 Show that the unitarity requirement leads to three independent constraints on the phases of the entries of the complex three-by-three matrix in Eq. (24.34). Next, show that by rephasing the quark fields that at most five of these phases can be removed.

The value of this CP violating phase has been measured in experiment. This value corresponds to the specific parameterization of the CKM matrix in terms of three mixing angles θ_{12}, θ_{23}, and θ_{13} and the CP violating phase δ:

$$V_{\rm CKM} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} , \qquad (24.47)$$

where $c_{ij} = \cos\theta_{ij}$ and $s_{ij} = \sin\theta_{ij}$. The measured values [20] of these parameters confirm the hierarchy of generation mixing:

$$\begin{aligned} \sin\theta_{12} &= 0.22500 \pm 0.00067 \\ \sin\theta_{23} &= 0.04182^{+0.00085}_{-0.00074} \\ \sin\theta_{13} &= 0.00369 \pm 0.00011 \ . \end{aligned} \qquad (24.48)$$

The mixing between the first and second generations is largest, then the second and third. The mixing between the first and third is the smallest. The CP violating phase, in radians, is

$$\delta = 1.144 \pm 0.027 \ , \qquad (24.49)$$

which is inconsistent with $\delta = 0$. This confirms the presence of CP violation from the interactions of quarks with the W.

Historically, CP violation was first observed in bound states. The Kaons (spin-0 mesons that include one strange quark and either an up or down quark) were observed to violate CP long before the origin of this violation was understood. In addition to being an aspect of the Standard Model to be understood, the nature of the Universe provides a reason to care about CP violation, specifically the fact that we observe much more matter (baryons) than antimatter (antibaryons).

The origin of this matter-antimatter asymmetry is unknown and is typically referred to as **baryogenesis**. Not long after CP violation was first observed in Kaons, Andrei Sakharov [47] showed that in order to produce baryogenesis from an initial state that has equal amount of matter and antimatter three conditions must be satisfied:

1. Baryon number B violation. Baryon number is defined carefully in Sec. 26. Essentially, it is like the conserved particle number of a complex scalar field or a Dirac fermion, but applied to the entire collection of baryons, not just a single particle type (like a proton). In

order to proceed from a state with equal numbers of baryons n_B and antibaryons $n_{\overline{B}}$, that is at state with $B = n_B - n_{\overline{B}} = 0$, to the nonzero baryon number we observe today there must be baryon number violation.

2. Violation of C and CP. Suppose there is a process that takes some initial state X with no baryon number to a state with $n_B \neq 0$. If C is s symmetry then the initial state $\overline{X}$ has exactly the same cross section to produce $n_{\overline{B}} = n_B$. So, even if baryon number is violated in each process the symmetric initial state produces a $B = 0$ final state unless C is violated. That CP violation is *also* necessary can be seen in a few ways. First, if the initial state has equal numbers of particles and antiparticles then it is CP symmetric. A final state with nonzero baryon number is not CP symmetric, so the process of going from initial to final states must violate CP. One can also note that, because CPT is always a symmetry, that violation of CP is equivalent to violation of T. If time reversal were a symmetry then processes that produce a net baryon number would be off set by equally efficient processes that eliminate a net baryon number.

3. Out of Equilibrium Conditions. This condition relies on statistical mechanics, which is beyond the scope of this text. In short, however, if the initial state is in thermal equilibrium then there is no chemical potential for the baryons. This ensures that even if conditions 1 and 2 are satisfied that no net baryon number is produced.

This may lead one to wonder if the Standard Model includes all the necessary ingredients for baryogenesis. The answer is somewhat interesting. Each type of the above conditions are met. As discussed Sec. 26 baryon number is a symmetry of the Standard Model Lagrangian, but is violated by quantum effects. The structure of the $SU(2)_L$ part of the Standard Model violates both C and CP. Finally, if what we describe as the Universe initially had a temperature above the Higgs VEV, then as it cools the Universe would go through a phase transition, which can provide out of equilibrium conditions.

For better or worse, even though all the ingredients seem to be there, the Standard Model does not produce enough of them. The measured CP violation in the Standard Model is not enough to produce the observed matter/antimatter asymmetry [48] and the electroweak phase transition is not strongly first order [49], meaning that any produced baryon number could be "washed out." Therefore, whatever describes the physics beyond the Standard Model may very well include additional sources of CP violation and the thermal history of the early Universe may well include strong phase transitions.

Section 25

Bound States & Flavor

The fact that there are several quark flavors was first inferred from the nature of the many quark bound states produced at particle colliders. While this was a fascinating and confusing time in particle physics we do not follow the historical process. Instead, we proceed along a nearly opposite logical direction. We use the established nature of the quarks to understand the nature of the bound states. The powerful tool for organizing and understanding these bound states is an approximate quark flavor symmetry.

Isospin is the name given to one of these symmetries. The idea is that up quark and the down quark look pretty much the same to the strong force. To begin with, they are in the same representation of the Lorentz group and the same representation of $SU(3)_c$. Despite these similarities, the assertion that they are the same might bother you seeing as the two quarks have different masses and different electric charges. However, the difference in their masses is small compared to the characteristic mass scale of the strong scale $\sqrt{\sigma} \approx 485$ MeV. This means that to the strong force, their masses may as well be the same. While their two electric charges are different (they even have different signs) the electric force is much, much weaker than the strong force within bound states. Therefore, the fact their charges differ is a small correction to the strong force seeing them as the same.

In short, the isospin symmetry is that we can imagine that the up and down quarks belong to a vector is isospin space

$$N^i = \begin{pmatrix} u \\ d \end{pmatrix} \ . \tag{25.1}$$

These are complex spinor fields, so each is governed by the Dirac Lagrangian. These can be combined into

$$\overline{N}_i \left(i\partial\!\!\!/ - m_N \right) N^i \ , \tag{25.2}$$

where m_N is some approximate mass for the doublet of fields. This Lagrangian is invariant under transformations

$$N^i \to U^i_{\ j} N^j \ , \tag{25.3}$$

where $U^i_{\ j}$ is a unitary matrix with unit determinant. In other words, the symmetry is described by the Lie group SU(2). This is the same structure as spin, which is why this symmetry is called isospin.

Let us start with mesons, two of these quarks bound together. One

is in the fundamental representation **2** and the other is in the conjugate representation $\overline{\mathbf{2}}$. We showed (39.95) that these two representations are equivalent, so we can consider a meson as belonging to the

$$\mathbf{2} \otimes \mathbf{2} = \mathbf{3} \oplus \mathbf{1} \ , \tag{25.4}$$

representation. However, in this combination the conjugate state is given by

$$\overline{N}^i = \varepsilon^{ij} N_j = \begin{pmatrix} \overline{d} \\ -\overline{u} \end{pmatrix} \ . \tag{25.5}$$

There are standard rules for writing the tensor product states in terms of the original basis states. You can look them up in Clebsch-Gordan tables. In the case at hand we are combining the states

$$\begin{pmatrix} \overline{d} \\ -\overline{u} \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} u \\ d \end{pmatrix} \ . \tag{25.6}$$

The states of the **3** are

$$\overline{d}u \ , \qquad \frac{1}{\sqrt{2}} \left(\overline{d}d - \overline{u}u \right) \ , \qquad \overline{u}d \ , \tag{25.7}$$

while the the **1** state is

$$\frac{1}{\sqrt{2}} \left(\overline{d}d + \overline{u}u \right) \ . \tag{25.8}$$

A similar story plays out with the SU(2) representations of spin for the quarks. We denote the combination of two spin-1/2 particles as

$$\begin{pmatrix} \uparrow \\ \downarrow \end{pmatrix} \otimes \begin{pmatrix} \uparrow \\ \downarrow \end{pmatrix} \tag{25.9}$$

where $\uparrow$ denotes positive spin angular momentum along the z-axis (spin-up) and $\downarrow$ denotes spin-down. The result is three states that make up a spin-1 representation

$$\uparrow\uparrow \ , \qquad \frac{1}{\sqrt{2}} \left(\uparrow\downarrow + \downarrow\uparrow \right) \ , \qquad \downarrow\downarrow \ , \tag{25.10}$$

and one state in the spin-0 representation

$$\frac{1}{\sqrt{2}} \left(\uparrow\downarrow - \downarrow\uparrow \right) \ . \tag{25.11}$$

Putting everything together, isospin symmetry predicts that mesons composed of up and down quarks should come in groups of three, where the particles have the same mass and spin, or a single particle. Because isospin is only an approximate symmetry we expect that the masses are

not exactly the same, but should be close. Part of the symmetry breaking is that the quarks have different electric charges, the u has charge 2/3 and the d has charge -1/3. The antiquarks have opposite charge. This allows us to make predictions of the electric charges of isospin states. The three triplet states have electric charges 1, 0, and -1, while the isospin singlet has charge 0.

Particle	Mass	Electric Charge	Spin	Isospin
π^+	139.57	1	0	$(1,1)$
π^0	134.98	0	0	$(1,0)$
π^-	139.57	-1	0	$(1,\text{-}1)$
ω	782.65	0	1	$(0,0)$
ρ^+	775.11	1	1	$(1,1)$
ρ^0	775.26	0	1	$(1,0)$
ρ^-	775.11	-1	1	$(1,\text{-}1)$

Table 3. Table of mesons composed of up and down quarks. Masses given in MeV. Isospin values give the representation and the value within the representation. More precise values (with uncertainties) and additional properties are found in [20].

The mesons made of up quark and down quarks are listed in Table 3. We see that there are two collections of isospin triplets, the pions (which are spin-0) and the rhos (which are spin-1). These mesons in each triplet all have similar masses and differ in electric charge by one unit steps. The omega meson is not part of a triplet, and it has zero electric charge, as predicted.

We now turn to the baryons composed of up and down quarks. In this case we are considering the combination of three quarks

$$\mathbf{2} \otimes \mathbf{2} \otimes \mathbf{2} = \mathbf{2} \otimes (\mathbf{3} \oplus \mathbf{1}) = \mathbf{4} \oplus \mathbf{2} \oplus \mathbf{2} \ . \tag{25.12}$$

This shows that we expect baryons to come in groups of four or groups of two. In each group the bound states should have equal spins and nearly equal masses. Because each quark has spin-1/2, the total allowed spins are either spin-1/2 (the **2**) or spin-3/2 (the **4**).

In order to determine the electric charges of these bound states we must consider what combinations of quarks arise in the different representations of isospin. In doing so we neglect the normalization factors (like $\sqrt{2}$) that showed up in the meson analysis. We simply attempt to keep track of the number of up and down quarks. In the case of spin, the four states have the schematic form

$$\uparrow\uparrow\uparrow, \quad \uparrow\uparrow\downarrow \ , \quad \uparrow\downarrow\downarrow \ , \quad \downarrow\downarrow\downarrow \ . \tag{25.13}$$

Similarly, the **4** of isospin has the schematic form of

$$uuu\ , \quad uud\ , \quad udd\ , \quad ddd\ . \tag{25.14}$$

This shows that the four baryons in this multiplet have electric charges of 2, 1, 0, and -1. The two isospin **2** representations distinct, leading to different configurations of the up and down quarks and hence to different magnetic dipole moments and other quantities. However, for our purposes it is sufficient to understand that the two states have the schematic form

$$uud\ , \quad udd\ . \tag{25.15}$$

Consequently, baryons in the **2** of isospin have electric charges 1 and 0.

Particle	Mass	Electric Charge	Spin	Isospin
p	0.9383	1	1/2	$(1/2, 1/2)$
n	0.9396	0	1/2	$(1/2, -1/2)$
Δ^{++}	1.232	2	3/2	$(3/2, 3/2)$
Δ^{+}	1.232	1	3/2	$(3/2, 1/2)$
Δ^{0}	1.232	0	3/2	$(3/2, -1/2)$
Δ^{-}	1.232	-1	3/2	$(3/2, -3/2)$

Table 4. Table of baryons composed of up and down quarks. Masses given in GeV. Isospin values give the representation and the value within the representation. More precise values (with uncertainties) and additional properties are found in [20].

Table 4 lists the baryons made of up and down quarks. We see that the proton p and neutron n comprise an isospin doublet and have masses that are nearly identical. There are also four Δ baryon states. It is difficult to measure the masses of these extremely short-lived particles, so to the precision they are known these spin-3/2 states all have equal mass, in complete agreement with the predictions of isospin symmetry.

Subsection 25.1

SU(3) Flavor

Isospin is not a perfect symmetry because the up and down quarks have different charges and different masses. The mass difference is a smaller effect in part because both quark masses are small compared to $\sqrt{\sigma} \approx 485$ MeV. The strange quark s has a mass of about 93 MeV. This may seem significantly larger than the up and down quarks, but the mass difference still somewhat smaller than $\sqrt{\sigma}$. In addition, the strange quark has the same electric charge as the down quark. Therefore, we might hope to extend the SU(2) isospin symmetry to an SU(3) flavor symmetry in which

the fundamental objects are quark triplets

$$N^i = \begin{pmatrix} u \\ d \\ s \end{pmatrix} \ . \tag{25.16}$$

We then expect the mesons and baryons to be in at least approximate representations of SU(3).

It is worth emphasizing that this SU(3) symmetry is completely distinct from the color gauge symmetry, which is sometimes denoted $SU(3)_c$. The gauge symmetry is understood to be the related to the number of gluons and the reason why three quarks make up a baryon. The flavor symmetry arises because three of the quarks happen to have masses that are within $\sqrt{\sigma}$ of each other. This flavor symmetry has no effect on the structure of the QCD Lagrangian, but it does provide a useful way to organize the large number of quark bound states.

We begin with mesons, combinations of a quark in the **3** of flavor and an antiquark in the $\overline{\mathbf{3}}$. Because

$$\mathbf{3} \otimes \overline{\mathbf{3}} = \mathbf{8} \oplus \mathbf{1} \tag{25.17}$$

we expect an octet of 8 mesons with similar mass or a singlet state. The singlet should have the form

$$\overline{N}_i N^i \to \overline{u}u + \overline{d}d + \overline{s}s \ , \tag{25.18}$$

which implies that it has zero electric charge.

The elements of the octet have the form

$$N^i \overline{N}_j - \frac{1}{3}\delta^i_j \overline{N}_i N^i \ . \tag{25.19}$$

We can write the first term as a three-by-three matrix

$$N^i \overline{N}_j \to \begin{pmatrix} u\overline{u} & u\overline{d} & u\overline{s} \\ d\overline{u} & d\overline{d} & d\overline{s} \\ s\overline{u} & s\overline{d} & s\overline{s} \end{pmatrix} \ . \tag{25.20}$$

This includes two states with electric charge of 1, two with electric charge -1, and four with zero charge.

While it seems that states with a down quark replaced by a strange quark are the same, this is only true in regard to electric charge. Remember that the strange quark has a much larger mass than the down quark. Historically, physicists striving to understand the experimental data regarding mesons and baryons noticed a new property (which was called strangeness after these strange new particles) that was conserved in all process. Now we understand strangeness to simply count the number of strange antiquarks.

So, one s has strangeness -1 and one $\overline{s}$ has strangeness 1.

Particle	Mass	Electric Charge	Isospin	Strange
η'	957.78	0	$(0,0)$	0
η	547.86	0	$(0,0)$	0
K^+	493.68	1	$(1/2,1/2)$	1
K^-	493.68	-1	$(1/2,-1/2)$	-1
K^0	497.61	0	$(1/2,-1/2)$	1
$\overline{K}^0$	497.61	0	$(1/2,1/2)$	-1

Table 5. Table of spin-0 mesons that include strange quarks. Masses given in MeV. Isospin values give the representation and the value within the representation. More precise values (with uncertainties) and additional properties are found in [20].

In Table 5 we list the spin-0 mesons that include strange quark contributions. Of these the η' mass is the largest by a considerable margin. Consequently, we associate this state with the **1** of SU(3) flavor. The other five states have similar masses. These are then combined with the three pion states to fill out the complete octect.

One might point out that the pions are considerable lighter than the other spin-0 mesons. There is a separate symmetry origin of this. Within the three flavor symmetry, one can imagine treating the left-handed and right-handed quarks separately. That is, there is an approximate $\mathrm{SU}(3)_L \times \mathrm{SU}(3)_R$ symmetry, where left-handed fields transform by way of matrices U_L and right-handed fields transform with U_R. It is believed that during the phase transition from free quarks and gluons to bound states that the combination of quark fields $\overline{q}q$ develops a vacuum expectation value, denoted $\langle \overline{q}q \rangle$.

Exercise 25.1 Argue that a nonzero quark condensate $\langle \overline{q}q \rangle \neq 0$ is not in general invariant under separate transformations of the left-handed quarks by U_L and the right-handed quark by U_R. However, show that the subgroup of transformations $U_L = U_R$ does not change the quark condensate. These transformations are referred to as the vector subgroup $\mathrm{SU}(3)_V$ because they treat the two chiralities in the same way. In essence, you are to show that quark condensation implements the spontaneous symmetry breaking $\mathrm{SU}(3)_L \times \mathrm{SU}(3)_R \to \mathrm{SU}(3)_V$. How many pseudo-Nambu-Goldstone bosons would be produced by this symmetry breaking pattern?

As the above exercise illustrates, these spin-0 states can be understood as pseudo-Nambu-Goldstone bosons of the spontaneously broken $\mathrm{SU}(3)_L \times \mathrm{SU}(3)_R$ symmetry. This explains why they are all somewhat lighter than the other meson states. The pions are even lighter. We can understand this as a consequence of $\mathrm{SU}(2)_L \times \mathrm{SU}(2)_R$ being a closer to exact

symmetry than $SU(3)_L \times SU(3)_R$. Because the symmetry has less explicit breaking (from the unequal quark masses) the pseudo-Nambu-Goldstone masses are closer to zero.

A popular way of exhibiting these states is shown in Fig. 25. Here the horizontal axis is isospin and vertical axis is strangeness. Lines of constant electric charge are also shown. This illustration is very suggestive of the root system of the adjoint representation (the **8**) shown in Fig. 35. This way of grouping the meson states makes the similarities among these particles more apparent.

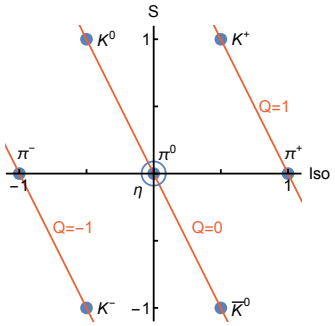


Figure 25. Plot of spin-0 meson octet. Note that two states overlap at the origin.

Particle	Mass	Electric Charge	Isospin	Strange
ϕ	1019.46	0	$(0,0)$	0
K^{*+}	891.66	1	$(1/2,1/2)$	1
K^{*-}	891.66	-1	$(1/2,-1/2)$	-1
K^{*0}	895.81	0	$(1/2,-1/2)$	1
$\overline{K}^{*0}$	895.81	0	$(1/2,1/2)$	-1

Table 6. Table of spin-1 mesons that include strange quarks. Masses given in MeV. Isospin values give the representation and the value within the representation. More precise values (with uncertainties) and additional properties are found in [20].

A similar structure governs the spin-1 mesons, shown in Table 6. In this case the ϕ particle and the ω from Table 3 seem to be combinations of the **1** and the the center point of the octet. For this reason, both are sometimes included in the spin-1 diagram (Fig. 26) and it is referred to as a nonet, even though this is not a representation of SU(3) flavor. Though the ϕ does not seem to be a true singlet of the flavor symmetry, it does have the most different mass of the collection, so removing this particle might be the most motivated octet choice.

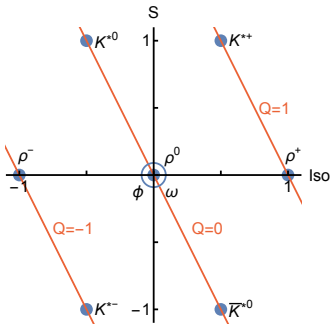


Figure 26. Plot of spin-1 meson nonet. Note that three states overlap at the origin.

Particle	Mass	Electric Charge	Isospin	Strange
Λ^0	1.115	0	$(0,0)$	-1
Σ^+	1.189	1	$(1,1)$	-1
Σ^0	1.192	0	$(1,0)$	-1
Σ^-	1.197	-1	$(1,-1)$	-1
Ξ^0	1.314	0	$(1/2,1/2)$	-2
Ξ^-	1.321	-1	$(1/2,-1/2)$	-2

Table 7. Table of spin-1/2 baryons that include strange quarks. Masses given in GeV. Isospin values give the representation and the value within the representation. More precise values (with uncertainties) and additional properties are found in [20].

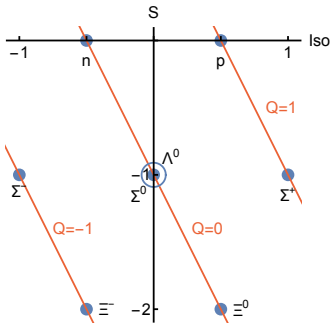


Figure 27. Plot of spin-1/2 baryon octet. Note that two states overlap at point with isospin=0 and strangeness=-1.

Finally, we consider the baryons. These are combinations of three

quarks so we consider the tensor product

$$\mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3} = \mathbf{10} \oplus \mathbf{8} \oplus \mathbf{8} \oplus \mathbf{1} \ . \tag{25.21}$$

This implies that we expect baryons that are grouped into eights or tens or by themselves. In Table 7 we list the properties of the spin-1/2 baryons that include strange quarks. These states (along with the proton and neutron) can also be organized in an octet, as shown in Fig. 27.

At this point the novelty of these groupings might begin to fade. However, there is one more collection of baryons to consider, and some interesting history is associated with it. The spin-3/2 baryons that include strange quarks are listed in Table 8. These particles, along with the Δ bound states in Table 4 fill out all ten states of the so-called baryon decuplet.

Particle	Mass	Electric Charge	Isospin	Strange
Σ^{*+}	1.383	1	$(1, 1)$	-1
Σ^{*0}	1.384	0	$(1, 0)$	-1
Σ^{*-}	1.387	-1	$(1, -1)$	-1
Ξ^{*0}	1.532	0	$(1/2, 1/2)$	-2
Ξ^{*-}	1.535	-1	$(1/2, -1/2)$	-2
Ω^-	1.672	-1	$(0, 0)$	-3

Table 8. Table of spin-3/2 baryons that include strange quarks. Masses given in GeV. Isospin values give the representation and the value within the representation. More precise values (with uncertainties) and additional properties are found in [20].

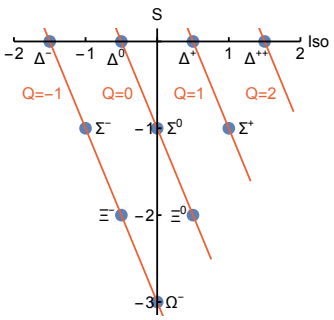


Figure 28. Plot of spin-3/2 baryon decuplet.

Back in 1961 the existence of quarks had not been established, but a great many mesons and baryons had been discovered at colliders. In particular, all the spin-3/2 baryons listed above had been discovered, except the Ω^-. In that year Susumu Okubo and Murray Gell-Mann independently predicted the existence of another spin-3/2 baryon with electric charge -1 and strangeness -3 with a mass near 1.6 GeV. Their motivation was the observation that all the particles fit in the decuplet shown in Fig. 28, except for one missing piece. When the Ω^- was discovered in 1964 this was seen as a great triumph of the quark hypothesis and led to further experimental testing and eventual confirmation.

Section 26

Symmetries of the Standard Model

We have now constructed the full Lagrangian of the Standard Model, which is worth recording again

$$\begin{aligned}\mathcal{L} = &-\frac{1}{4}G^a_{\mu\nu}G^{a\mu\nu} - \frac{1}{4}W^a_{\mu\nu}W^{a\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} \\ &+ i\overline{Q}\not{D}Q + i\overline{L}\not{D}L + i\overline{D}\not{D}D + i\overline{U}\not{D}U + i\overline{E}\not{D}E \\ &+ (D_\mu H)^\dagger D^\mu H + \mu^2|H|^2 - \lambda|H|^4 \\ &+ (\lambda_\ell)^i{}_j\,\overline{L}_i H E^j + (\lambda_d)^i{}_j\,\overline{Q}_i H D^j + (\lambda_u)^i{}_j\,\overline{Q}_i \widetilde{H} U^j + \text{ H.c. }. \end{aligned} \tag{26.1}$$

Symmetries of various kinds have enabled us to determine this structure. First and foremost is the structure of spacetime encoded by the Lorentz group. These symmetries led us to use representations of the Lorentz group to describe the flow of momentum and energy. Each representation gives rise to different type of field—spin-0, spin-1/2, and spin-1—and these are the basic ingredients of the Standard Model.

Gauge symmetries also provide great insight into the structure of the Standard Model. The product group $\mathrm{SU}(3)_c{\times}\mathrm{SU}(2)_L{\times}\mathrm{U}(1)_Y$ governs how all of the elementary spin-1 fields interact with each other and other particles. These symmetries also provide a significant constraint on the types of interactions that can occur amongst the fermions and the Higgs.

In the rest of this section we discuss other symmetries and approximate symmetries that can be used to understand the dynamics of the Standard Model and can also be useful in guiding our understanding of what may lie beyond the Standard Model. It may seem surprising that approximate symmetries have a place in this discussion. But if even a part of a model exhibits a symmetry then the interactions that have to do with that part of the model can be best understood by considering that symmetry. We have already seen some examples of this. Isospin and the SU(3) flavor symmetry are not exact, but they still provide a useful framework for organizing the quark bound states.

Another example is CP. Most of the Standard Model is invariant under the discrete CP symmetry. As shown in section 24.1, however, the interactions of the W boson can violate CP when all three quark generations are involved in a process. Therefore, any Standard Model process that violates CP must involve not only the W but all three quark generations, which is a quite restrictive requirement.

Another approximate symmetry of the Standard Model is the flavor symmetry described in the previous sections. Each type of fermionic field in a given generation is extended to a U(3) multiplet. This leads to an approximate $\mathrm{U}(3)_E{\times}\mathrm{U}(3)_L{\times}\mathrm{U}(3)_U{\times}\mathrm{U}(3)_D{\times}\mathrm{U}(3)_Q$ flavor symmetry. This symmetry is broken only by the Yukawa interactions of the fermions with the Higgs.

One might ask if any subgroup of this large flavor symmetry is preserved by the Yukawa terms. The answer is, "Yes." Remember that when we determined how many phases could be removed from the CKM matrix that a global rephasing of *all* the quark fields had no effect on the matrix. This means that such a rephasing is a continuous global symmetry. The conserved charge is called **Baryon Number** because its effects were first observed in baryon interactions. We can describe the symmetry breaking as $\mathrm{U}(3)_U \times \mathrm{U}(3)_D \times \mathrm{U}(3)_Q \to \mathrm{U}(1)_B$.

We assign each quark field a baryon number charge of $1/3$ and all other fields have a charge of 0. This also implies that all anti-quark fields have baryon number of $-1/3$. Therefore, mesons (which are composed of one valence quark field and one valence anti-quark field) have a baryon number of 0. In contrast, baryons (which have three valence quark fields) have a baryon number of 1. Anti-baryons have a baryon number of -1. A simple and familiar example of baryon number conservation is neutron decay

$$n \to p + e^- + \overline{\nu}_e \ . \tag{26.2}$$

The initial state (the neutron) has baryon number 1. The final state has one proton (with baryon number 1) and two leptons, each with baryon number 0.

Within the Standard Model there is a larger residual lepton flavor symmetry. After the flavor symmetry is used to diagonalize the lepton masses

$$(\overline{e}_L \, , \, \overline{\mu}_L \, , \, \overline{\tau}_L) \begin{pmatrix} m_e & 0 & 0 \\ 0 & m_\mu & 0 \\ 0 & 0 & m_\tau \end{pmatrix} \begin{pmatrix} e_R \\ \mu_R \\ \tau_R \end{pmatrix} \ , \tag{26.3}$$

we can individually rephase any single one of the mass eigenstate lepton flavors. This leads to a conserved electron number, muon number, and tau number. We might describe this as $\mathrm{U}(3)_E \times \mathrm{U}(3)_L \to \mathrm{U}(1)_e \times \mathrm{U}(1)_\mu \times \mathrm{U}(1)_\tau$. Under these flavor symmetries both a lepton and its associated neutrino have charge 1, while all others have charge 0. The antiparticle and antineutrino have charge -1. This is why in neutron decay (26.2), which has zero electron number in the initial state, we find an electron (with electron number 1) and an electron flavored antineutrino (with electron number -1). We see that neutron decay is largely determined by Lorentz invariance (which ensures a particle can only decay into a collection of particles with total mass less than the decaying particle mass), conservation of electric charge, conservation of baryon number, and conservation of electron number. Another example is muon decay

$$\mu^- \to e^- + \nu_\mu + \overline{\nu}_e \ . \tag{26.4}$$

The initial state is muon number of 1. The final state has one muon neutrino (with muon number of 1) one electron (with electron number of

1) and an electron antineutrino (with electron number of -1). Note that conservation of electric charge, muon number, and electron number are also on display.

As useful as baryon number and the lepton numbers are, it turns out they are actually not quite conserved. The only flaw in the arguments we have made up to this point is that we have not taken into account quantum effects. While the classical field theory is often a good guide to the leading order quantum theory there are cases where the current associated with a continuous symmetry J^μ of the Lagrangian is *not* conserved

$$\partial_\mu J^\mu \neq 0 \ , \tag{26.5}$$

when quantum effects are included.

These types of symmetries are called **Anomalous Symmetries** and baryon number along with each lepton flavor number are examples. Taking into account quantum corrections one finds

$$\partial_\mu J^\mu_B = \frac{3g^2}{32\pi^2}\varepsilon^{\mu\nu\alpha\beta}W^a_{\mu\nu}W^a_{\alpha\beta} \tag{26.6}$$

$$\partial_\mu J^\mu_{L_i} = \frac{g^2}{32\pi^2}\varepsilon^{\mu\nu\alpha\beta}W^a_{\mu\nu}W^a_{\alpha\beta} \ , \tag{26.7}$$

where the $W^a_{\mu\nu}$ are the nonabelian field-strength tensors of the $\mathrm{SU}(2)_L$ gauge group. The right-hand side is only effectively nonzero at energies high enough to produce the heavy gauge bosons, this is why these symmetries are useful for low energy processes like neutron decay and muon decay. It is also interesting that the right-hand sides of both of these equations have the same form, differing only by the number in front. This means that if we consider the baryon number current minus the sum of all the lepton number currents we find

$$\begin{aligned}\partial_\mu J^\mu_{B-L} =& \partial_\mu \left(J^\mu_B - J^\mu_{L_e} - J^\mu_{L_\mu} - J^\mu_{L_\tau} \right) \\ =& \frac{(3-1-1-1)g^2}{32\pi^2}\varepsilon^{\mu\nu\alpha\beta}W^a_{\mu\nu}W^a_{\alpha\beta} = 0 \ .\end{aligned} \tag{26.8}$$

That is, the $B-L$ symmetry has a conserved current even when quantum effects are included.

This analysis matters because (as mentioned during our discussion of CP violation) the Universe seems to have more baryons than antibaryons. This asymmetry is good because it allows things like stars and planets to form. But we do not know the origin of this baryogenesis. Our analysis of the $B-L$ symmetry allows for processes at high energy can change the total baryon number (say from zero to something nonzero) by changing the total lepton number by the same amount. Not all models of baryogenesis make use of this process, but many do.

Let us turn from the fermion flavor symmetries to the Higgs field. We spent a lot of time discussing the spontaneous breaking of the gauge symmetries of the Standard Model $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$. However, there is a different pattern of *global* symmetry breaking in the Higgs potential itself. Writing the Higgs field in terms of four real fields

$$H = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \begin{pmatrix} \phi_3 + i\phi_4 \\ \phi_1 + i\phi_2 \end{pmatrix} \ . \tag{26.9}$$

The Higgs potential can then be written as

$$\begin{aligned} V =& -\mu^2 \left(\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2\right) + \lambda \left(\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2\right)^2 \\ =& -\mu^2 \Phi \cdot \Phi + \lambda \left(\Phi \cdot \Phi\right)^2 \ , \end{aligned} \tag{26.10}$$

where

$$\Phi = \begin{pmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \phi_4 \end{pmatrix} \ . \tag{26.11}$$

This shows that the potential is completely equivalent to that of a real scalar field in the fundamental representation of a global $\mathrm{SO}(4) \simeq \mathrm{SU}(2) \times \mathrm{SU}(2)$ symmetry getting a VEV, $\langle \phi_1 \rangle \neq 0$, in only the first component. The remaining global symmetry is then $\mathrm{SO}(3) \simeq \mathrm{SU}(2)$. This symmetry breaking pattern $\mathrm{SO}(4) \to \mathrm{SO}(3)$ would produce three Nambu-Goldstone bosons. These would-be Nambu-Goldstone bosons are those degrees of freedom that end up as the longitudinal modes of the massive W and Z bosons. The interactions of these vector modes are governed by the residual, or **Custodial**, SU(2) symmetry. This leads to specific predictions about W and Z interactions at low energy, which have been verified to great accuracy.

The final approximate symmetry we consider is scale symmetry of the action

$$S = \int d^4x \, \mathcal{L} \ . \tag{26.12}$$

This symmetry considers a rescaling of each coordinate $x^\mu \to \xi x^\mu$ by a positive quantity ξ. This means that

$$d^4x \to \xi^4 d^4x \ . \tag{26.13}$$

This also means that

$$\partial_\mu = \frac{\partial}{\partial x^\mu} \to \frac{\partial}{\xi \partial x^\mu} = \xi^{-1} \partial_\mu \ . \tag{26.14}$$

The particle fields themselves also change under rescaling. Recall from Sec. 14.1 that each field has dimensions of some power of energy. Spin-0 and spin-1 fields have dimension 1 while spin-1/2 fields have dimension

3/2. The energy (or mass) dimension of the fields is equivalent in natural units to inverse length. Therefore, we expect that under a rescaling of the coordinates that the fields should be rescaled by the inverse of their energy dimension. That is, for spin-0 ϕ, spin-1/2 ψ, and spin-1 A_μ that

$$\phi \to \xi^{-1}\phi \ , \qquad \psi \to \xi^{-3/2}\psi \ , \qquad A_\mu \to \xi^{-1}A_\mu \ . \tag{26.15}$$

How then do various terms in the Standard Model Lagrangian transform?

Exercise 26.1

Show that the massless free field (no interactions) Lagrangians for spin-0, spin-1/2, and spin-1 fields all transform as

$$\mathcal{L}_{\text{free}} \to \xi^{-4}\mathcal{L}_{\text{free}} \ , \tag{26.16}$$

under scale transformations. This implies, using Eq. (26.13), that the action for free field theories is invariant under scale transformations.

Moving beyond the massless free particle parts of the Lagrangian, we can consider how the interaction terms transform. Most of these terms include two fermion fields and one boson, either spin-1 or spin-0. Each of these terms transform as (for instance)

$$B_\mu \overline{L}\gamma^\mu L \to \xi^{-1}B_\mu \xi^{-3/2}\overline{L}\gamma^\mu \xi^{-3/2}L = \xi^{-4}B_\mu \overline{L}\gamma^\mu L \ . \tag{26.17}$$

Because the measure of the action integral transforms with ξ^4 (26.13), we see that all these action terms are invariant under scale transformations. The only other terms to consider is the Higgs potential. These terms transform as

$$-\mu^2|H|^2 + \lambda|H|^4 \to -\frac{\mu^2}{\xi^2}|H|^2 + \frac{\lambda}{\xi^4}|H|^4 \ . \tag{26.18}$$

The λ term leads to a scale invariant interaction, but (including the transformation of the action measure) the μ^2 term does transform under rescaling.

What we have found is that there is exactly one term in the Standard Model that transforms under scale transformations. It is the Higgs mass parameter and under scaling $\mu \to \xi\mu$. Another way of interpreting this is to say that there is only one explicit energy scale in the Standard Model, and it is μ. (There are some implicit scales from the quantum mechanical running of the couplings, but μ is the only classical scale.) From measurements of the electroweak and Higgs bosons we know that $\mu \sim 90$ GeV. However, we do not know why this specific scale appears. In fact, there is one other physical scale from gravity, which we take as beyond the Standard Model, see Sec. 27. This is the Planck scale, $M_{\text{Pl}} \sim 10^{19}$ GeV. We do not really know where this scale comes from either. The huge hierarchy between the

two scales is also, so far, an unsolved mystery.

PART VII

Beyond The Standard Model

Sec. 27. Gravity
Sec. 28. Neutrino Masses
Sec. 29. Strong CP
Sec. 30. Unification
Sec. 31. EW Hierarchy

There are, of course, an unending number of ways that the Standard Model could be extended, even within the confines of needing to agree with existing experimental data. This section makes no claim of even coming close to being comprehensive in this regard. Rather, we mention a few extensions of the Standard Model that students often first ask about.

There are a few things that are conspicuously absent from this section. These include the nature of the cosmological dark matter and baryogengesis as well as possible models of early universe inflation or dark energy. These are all very interesting, but generally require a more detailed discussion of cosmology than is within the scope of this text.

Section 27

Gravity

We have, so far, described the dynamics responsible for the electromagnetic as well as the weak and strong nuclear forces. Of course, in day-to-day experience living near a planet the gravitational force is the most important. This is largely because the gravitational interactions are long range and always attractive. At large scales the nuclear forces are unimportant, though electrodynamics does play a role. However, most matter is electrically neutral and so gravity is the force that remains.

Like the other forces we have encountered gravity can be described by the classical field. The quantum field theory considers the quantum fluctuations about a classical vacuum. This can be done in a relatively straightforward way for gravity, similar to what we have seen for other fields. The fluctuations lead to particles, gravitons, that communicate the force of gravity. However, there are also reasons to believe that this process does not describe the ultimate theory of gravity, which would be needed to describe quantum mechanical gravitational effects when gravity is very strong. Still, for the effects of gravity on earth, where the effects are quite weak compared to other forces, this effective theory of gravity is sufficient.

The classical field theory that describes gravity is called General Relativity which is introduced in many texts [50, 51]. In this theory the primary object is the metric $g_{\mu\nu}$. This is a symmetric rank-2 tensor that describes the geometrical properties of spacetime. One specific example is the flat space metric (in Cartesian coordinates) that we have used throughout this

text

$$\eta_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} . \tag{27.1}$$

This is the classical background (or vacuum expectation value) for the graviton field. We can then describe fluctuations about this background by

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} , \tag{27.2}$$

where $h_{\mu\nu}$ is the dynamical graviton field. Notice that $\eta_{\mu\nu}$ is dimensionless and so $g_{\mu\nu}$ and $h_{\mu\nu}$ are dimensionless. Similarly, because $g_{\mu\nu}$ and $\eta_{\mu\nu}$ are symmetric it follows that $h_{\mu\nu}$ is also.

One might be tempted to write the inverse metric as

$$g^{\mu\nu} \stackrel{?}{=} \eta^{\mu\nu} + h^{\mu\nu} , \tag{27.3}$$

but this cannot be correct. Suppose we schematically write the inverse metric in powers of h

$$g^{\mu\nu} = \eta^{\mu\nu} + c_1 h^{\mu\nu} + c_2 h^{\mu}{}_{\alpha} h^{\nu\alpha} + c_3 h^{\alpha}{}_{\alpha} h^{\mu\nu} + \mathcal{O}(h^3) . \tag{27.4}$$

This is the most general possibility that is symmetric in μ and ν. Note that in this expansion we have raised and lowered the indices of $h_{\mu\nu}$ using the Minkowski metric $\eta_{\mu\nu}$. We can then use the fact that

$$g^{\mu\alpha} g_{\alpha\nu} = \delta^{\mu}_{\nu} , \tag{27.5}$$

to determine the values of the unknown coefficients order by order in h.

Exercise 27.1 Use the method described above to show that

$$g^{\mu\nu} = \eta^{\mu\nu} - h^{\mu\nu} + h^{\mu}{}_{\alpha} h^{\nu\alpha} + \mathcal{O}(h^3) . \tag{27.6}$$

Is this results symmetric in μ and ν?

What equations of motion do $g_{\mu\nu}$ and $h_{\mu\nu}$ satisfy, or equivalently what is the Lagrangian of general relativity? This is quite easy to write down

$$S_{\rm GR} = \frac{1}{2} M_{\rm Pl}^2 \int d^4x \sqrt{-g} R , \tag{27.7}$$

but it takes a bit of work to define all the quantities. The Planck mass is a huge energy compared to all known particle masses and interaction

energies. It is related to Newton's gravitational constant G_N by

$$M_{\rm Pl} = \frac{1}{\sqrt{8\pi G_N}} \approx 2.435 \times 10^{18} \text{ GeV} . \tag{27.8}$$

The quantity g is the determinant of the metric, which for the choice of metric signature we use is always negative. For instance, the determinant of $\eta_{\mu\nu}$ is -1. As above, we can make a schematic expansion with unknown coefficients

$$g = -1 + c_1 h^\alpha{}_\alpha + c_2 h^\alpha{}_\alpha h^\alpha{}_\alpha + c_3 h^{\alpha\beta} h_{\alpha\beta} + \mathcal{O}(h^3) . \tag{27.9}$$

Then, by using the identity

$$\partial_\mu g = g g^{\alpha\beta} \partial_\mu g_{\alpha\beta} , \tag{27.10}$$

we can determine the c_i coefficients.

Exercise 27.2 Use the above identity to show that

$$g = -1 - h^\alpha{}_\alpha - \frac{1}{2} h^\alpha{}_\alpha h^\beta{}_\beta + \frac{1}{2} h^{\alpha\beta} h_{\alpha\beta} + \mathcal{O}(h^3) . \tag{27.11}$$

The meaning of R takes more work and a lot more formalism. We do not take the space to describe all that could be covered in defining curvature. However, as we go through the process of defining R we point to several ways in which general relativity is similar to a nonabelian gauge theory.

In gauge theories we found it useful to define a gauge covariant derivative $D_\mu = \partial_\mu + igT^a A^a_\mu$. In differential geometry (which is the mathematical framework used to describe calculus in curved spaces) we also define a covariant derivative ∇_μ. For gauge theories the covariant derivative changes depending on what representation T^a of the Lie algebra corresponds to the field it is acting on. In general relativity the covariant derivative changes according to what representation of the Lorentz group a particular field is. This is because the tangent space at any point (which plays a similar role to the Lie algebra) is given by flat space, even if the spacetime is curved. For example, for a spin-0 field the covariant derivative is just the usual derivative $\nabla_\mu \phi = \partial_\mu \phi$.

For a spin-1 field V^α

$$\nabla_\beta V^\alpha = \partial_\beta V^\alpha + \Gamma^\alpha_{\beta\gamma} V^\gamma , \tag{27.12}$$

where

$$\Gamma^\alpha_{\beta\gamma} = \frac{1}{2} g^{\alpha\delta} \left(\partial_\beta g_{\gamma\delta} + \partial_\gamma g_{\beta\delta} - \partial_\delta g_{\beta\gamma} \right) , \tag{27.13}$$

is the Christoffel symbol of the second kind, or the **metric connection**.

Note that for fluctuations about flat space that

$$\Gamma^{\alpha}_{\beta\gamma} = \frac{1}{2}\left(\eta^{\alpha\delta} - h^{\alpha\delta}\right)\left(\partial_\beta h_{\gamma\delta} + \partial_\gamma h_{\beta\delta} - \partial_\delta h_{\beta\gamma}\right) + \mathcal{O}(h^3) \ . \tag{27.14}$$

In gauge theories a commutator of covariant derivatives is related to the field-strength tensor, $[D_\mu, D_\nu]\,\phi = igF_{\mu\nu}\phi$. In general relativity there is a similar relation

$$[\nabla_\mu, \nabla_\nu]\,V^\alpha = R^\alpha{}_{\beta\mu\nu}V^\beta \ , \tag{27.15}$$

where

$$R^\alpha{}_{\beta\mu\nu} = \partial_\mu\Gamma^\alpha_{\nu\beta} - \partial_\nu\Gamma^\alpha_{\mu\beta} + \Gamma^\alpha_{\mu\delta}\Gamma^\delta_{\nu\beta} - \Gamma^\alpha_{\nu\delta}\Gamma^\delta_{\mu\beta} \ , \tag{27.16}$$

is the Riemann tensor. This object is related to the curvature of the spacetime. A different object, the Ricci tensor, is given by

$$R_{\alpha\beta} = R^\gamma{}_{\alpha\gamma\beta} = \partial_\gamma\Gamma^\gamma_{\alpha\beta} - \partial_\beta\Gamma^\gamma_{\gamma\alpha} + \Gamma^\gamma_{\gamma\delta}\Gamma^\delta_{\alpha\beta} - \Gamma^\gamma_{\beta\delta}\Gamma^\delta_{\alpha\gamma} \ . \tag{27.17}$$

The Ricci scalar, which appears in the action, is defined to be

$$R = g^{\alpha\beta}R_{\alpha\beta} \ , \tag{27.18}$$

which is the quantity that appears in the action (27.7).

One essential feature of the nonabelian gauge theories we have considered is that their Lagrangians are invariant under gauge transformations. One might naturally ask, considering all the connections we have drawn with gravity, if $h_{\mu\nu}$ manifests something like gauge transformations and if R is invariant under them? In general relativity one consider the manifold (the spacetime) to be the physical object. However, to make calculations we must typically use some coordinate system, which we might denote x^μ, to label points in spacetime. Or course, the physics had better not depend on what coordinates we choose. Therefore, any coordinate transformation

$$x^\mu \to x'^\mu = x^\mu + \xi^\mu(x) \ , \tag{27.19}$$

should leave all physical quantities unchanged. This leads to

$$h_{\mu\nu} \to h'_{\mu\nu} = h_{\mu\nu} + \partial_\mu\xi_\nu + \partial_\nu\xi_\mu \ , \tag{27.20}$$

which has some similarity to the transformations $A_\mu \to A'_\mu = A_\mu + \partial_\mu\alpha(x)$ of gauge fields. After some algebra, one can show that $\sqrt{-g}R$ is invariant under these transformations.

Exercise 27.3 Show that the trace $h^\alpha{}_\alpha$ transforms as

$$h^\alpha{}_\alpha \to h'^\alpha{}_\alpha = h^\alpha{}_\alpha + 2\partial_\alpha\xi^\alpha \ . \tag{27.21}$$

For massless gauge fields, as discussed in Sec. 11, the gauge transfor-

mations of the potential were used to eliminate two degrees of freedom from the four vector A_μ. One finds that the massless vector field has two degrees of freedom which are transverse (or orthogonal) to the momentum flowing through the field. In Sec. 10 we also discuss this field as in the (1/2, 1/2) representation of the Lorentz group, which implies that it describes particles with spin representation

$$\frac{\mathbf{1}}{\mathbf{2}} \otimes \frac{\mathbf{1}}{\mathbf{2}} = \mathbf{1} \oplus \mathbf{0} \ . \tag{27.22}$$

Here the three degrees of freedom in $\mathbf{1}$ and the single degree in $\mathbf{0}$ match the four components of a four-vector field. After applying the gauge transformation analysis only the transverse part of the spin-1 degrees of freedom remain.

Exercise 27.4 Argue that $h_{\mu\nu}$ has 10 independent degrees of freedom. How many degrees of freedom does

$$h_{\mu\nu} - \frac{1}{4}\eta_{\mu\nu}h^{\alpha}{}_{\alpha} \ , \tag{27.23}$$

have?

The massless field described by $h_{\mu\nu}$ can be analyzed in a similar way, though with more involved algebra. As a symmetric tensor, $h_{\mu\nu}$ is described by 10 independent components. By using the gauge freedom eight of these can be removed. Here the number eight is twice the number of components of ξ^μ, similar to how the two components removed from A_μ is twice the degrees of freedom in the scalar field $\alpha(x)$. Therefore, $h_{\mu\nu}$ only describes two physical degrees of freedom, or in other words it has two polarizations. These turn out to be transverse to the three-momentum of the field.

What is the spin of this field? It has two spacetime indices, no spinor indices, so the (1, 1/2) and (1/2, 1) representations are not the right fit. What about (1, 1)? The spin described by such a field is

$$\mathbf{1} \otimes \mathbf{1} = \mathbf{2} \oplus \mathbf{1} \oplus \mathbf{0} \ . \tag{27.24}$$

In this case the spin-2 representation describes five degrees of freedom. When added to the four of $\mathbf{1} \oplus \mathbf{0}$ we only have nine, one short of the 10 components in $h_{\mu\nu}$. It turns out that the (1, 1) representation of the Lorentz group describes symmetric, traceless tensors, like the one shown in (27.23). However, the trace of the tensor is clearly a scalar field, so it is in the $\mathbf{0}$ representation. Therefore, the total spins of $h_{\mu\nu}$ are

$$\mathbf{1} \otimes \mathbf{1} \oplus \mathbf{0} = \mathbf{2} \oplus \mathbf{1} \oplus \mathbf{0} \oplus \mathbf{0} \ . \tag{27.25}$$

By using the gauge freedom of this field all of the degrees of freedom are eliminated except for two of $\mathbf{2}$, which are the transverse, traceless part.

Therefore, the graviton is a spin-2 particle.

The fact that we have arrived at a spin-2 particle is highly significant. Weinberg showed [52] that a massless spin-2 particle couples with the same strength to every other field. Viewing this from the perspective of gravity, this shows that a graviton must interact with everything in the same way, or that all matter and energy interact with gravity universally, which is what we observe. This same analysis can be extended to higher spin fields. The result is that no massless bosonic particle with spin higher than two can consistently couple to other fields.

Now that we are thinking of $h_{\mu\nu}$ as describing the graviton field, we would like to understand it better. It's interactions are obtained from the action defined Eq. (27.7). To properly determine all the gravitational dynamics we should expand both R and $\sqrt{-g}$ around flat space. This is algebraically tedious and not so very illuminating, so we proceed somewhat schematically. For instance, we find that three types of terms arise in

$$R \sim \partial h \partial h + h \partial h \partial h + h^2 \partial h \partial h \ , \tag{27.26}$$

which are seen to be two derivative terms that lead to particle propagation through spacetime along with self-interaction terms involving three or four gravitons. This is similar to the gauge theories outlined above, we find a kinetic term composed of several couplings involving two derivatives and also self-interaction terms. Therefore, the Lagrangian includes terms like

$$\mathcal{L} \supset M_{\rm Pl}^2 \frac{1}{2} \partial h \partial h \ . \tag{27.27}$$

This is similar to the kinetic terms (two derivative terms) that we found for scalar, spinor, and vector fields. But, in all of those cases there was no dimensionful coefficient. To produce a theory with the usual coefficient of one-half, we must absorb a factor of $M_{\rm Pl}$ into the definition of $h_{\mu\nu}$. That is, we consider fluctuations around flat space of the form

$$g_{\mu\nu} = \eta_{\mu\nu} + \frac{1}{M_{\rm Pl}} h_{\mu\nu} \ , \tag{27.28}$$

where $h_{\mu\nu}$ is a field with mass dimension one, similar to other bosonic fields. While this choice has removed the factor of $M_{\rm Pl}$ from the two derivative terms, it has also ensured that graviton self interactions have the schematic form

$$\frac{1}{M_{\rm Pl}} h \partial h \partial h + \frac{1}{M_{\rm Pl}^2} h^2 \partial h \partial h \ . \tag{27.29}$$

In other words, the self-interactions are suppressed by factors of the Planck scale. For a process of energy E the derivatives (in momentum space) provide a factor of E meaning the Lagrangian terms go like $E/M_{\rm Pl}$ and $E^2/M_{\rm Pl}^2$. The current highest energy colliders run at about 10^4 GeV so the

leading order graviton interactions at this energy are suppressed by about 10^{-15}.

This is the sense in which quantum effects due to gravity can be systematically included in any theory. However, the effects are very small making such predictions difficult to test. This is also part of what makes a fully quantum mechanical theory of gravity elusive. While several (quite distinct) directions for such a theory have been put forward, we have not yet been able to use experimental results to determine how well any of them models nature.

Throughout this section we have not seen any issue or conflict with combining general relativity and quantum field theory. We can, however, see something suspicious in this quantum mechanical description of gravity. The couplings of the graviton to itself in (27.29) are all nonrenormalizeable, as discussed in Sec. 14.1. In fact, this is true of the graviton's interactions with other particles as well. The coupling to a fermionic field ψ would arise from

$$\sqrt{-g}\,\overline{\psi}\,(i\gamma^\mu\partial_\mu - m)\,\psi\ . \tag{27.30}$$

The leading interaction is then

$$\frac{1}{M_{\rm Pl}}h^\alpha{}_\alpha\overline{\psi}\,(i\gamma^\mu\partial_\mu - m)\,\psi\ , \tag{27.31}$$

which is also nonrenormalizable.

As we saw with the four-Fermi description of beta decay in Sec. 22, nonrenormalizable theories can be used to specific, testable predictions at energies well below the scale associated with the couplings with negative mass dimension. At high energies we expect new particles to arise whose mass is related to the same mass scale. If this thinking is applied to gravity, we might then expect that general relativity is a very good theory at energies well below the Planck mass, but needs to be corrected at higher energies. We might also expect that at energies near the Planck mass that new, very heavy, particles appear. Of course, we are nowhere near having controlled experimental access to these energies in order to probe what the high energy version of gravity might be. Several possibilities have been put forward, but as yet there is no experimental data that prefers one above the rest.

Section 28

Neutrino Masses

Experimentally, we have strong evidence that neutrinos have mass. However, there are several different ways that their mass might be included in a model of particle interactions and as yet there is no experimental evidence

that distinguishes one model from another. Until we know more, the typical choice is to leave the neutrino masses out of the official description of the Standard Model. Within the Standard Model neutrinos only appear as a component of left-handed lepton doublets, such as

$$L = \begin{pmatrix} \nu_e \\ e_L \end{pmatrix} \ . \tag{28.1}$$

Unlike every other known fermion, no right-handed partner to the neutrinos have been detected. Without such a particle we cannot write down the types of masses the other fields develop.

Despite all this we have very convincing evidence that neutrinos *do* have mass! This evidence comes from what are called **Neutrino Oscillations**. These oscillations have to do with a misalignment between different bases for the neutrinos. This is similar to the appearance of the CKM matrix when the quark interactions with the W boson are expressed in the basis with diagonal masses.

Example

We can understand the principles that lead to neutrino oscillations with only two neutrino flavors. Considering that two-by-two matrices are much easier to manipulate we consider this case in some detail. We begin by defining the two neutrino flavors in the "interaction basis." This is the basis in which interactions with the W boson are diagonal

$$\nu_f = \begin{pmatrix} \nu_e \\ \nu_\mu \end{pmatrix} \ . \tag{28.2}$$

We also suppose that there is some basis in which the neutrinos have a diagonal mass matrix with masses m_1 and m_2. We label this basis

$$\nu_m = \begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix} \ . \tag{28.3}$$

These two basis are related by some linear transformation $\nu_f = O\nu_m$, where O is a two-by-two matrix. All the linear transformations that preserve the normalization of the kinetic terms, that is

$$\overline{\nu}_f i \not{D} \nu_f \to \overline{\nu}_m O^\dagger i \not{D} O \nu_m = \overline{\nu}_m i \not{D} \nu_m \ , \tag{28.4}$$

are unitary $O^\dagger = O^{-1}$. However, as we saw in our discussion of a two-by-two CKM matrix in Sec. 24.1 by changing our basis by global phases O can be made into a real, orthogonal matrix. These have the form

$$O = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \ . \tag{28.5}$$

Or that the lepton flavors of neutrinos can be expressed as a linear combination of the two mass eigenstates and vice versa

$$\nu_e = \cos\theta\nu_1 + \sin\theta\nu_2 \ , \qquad \nu_\mu = -\sin\theta\nu_1 + \cos\theta\nu_2 \ , \tag{28.6}$$

$$\nu_1 = \cos\theta\nu_e - \sin\theta\nu_\mu \ , \qquad \nu_2 = \sin\theta\nu_e + \cos\theta\nu_\mu \ . \tag{28.7}$$

Why does this matter? When we measure something about a neutrino we are effectively using the interaction basis ν_f. But when neutrinos travel from one point in spacetime to another they propagate in the mass basis. Let us consider a simple experiment. An electron flavored neutrino is produced at the point $x = (0,0,0)$ at time $t = 0$. An apparatus which can detect electron flavored neutrinos is separated from the production point by a length L. We also denote the time of detection as $t = T$. If the neutrinos were massless then they would move at the speed of light ($c = 1$) and so $L/T = 1$. For a particle whose mass is much smaller than the energy of the decay process that produces it

$$\frac{L}{T} = v = \frac{p}{E} = \frac{\sqrt{E^2 - m^2}}{E} = \sqrt{1 - \frac{m^2}{E^2}} \approx 1 - \frac{m^2}{2E^2} \ . \tag{28.8}$$

Given a quantum state at $x = 0$ and $t = 0$ $|\nu_1(0,0)\rangle$, with no potential, we can translate it in space using Eq. (33.64) and in time by Eq. (33.67). Or in short, we can translate in spacetime by

$$|\nu_1(T,L)\rangle = e^{-ip_{1\mu}x^\mu}|\nu_1(0,0)\rangle \ , \tag{28.9}$$

where (choosing the distance to be in the z-direction) $x^\mu = (T,0,0,L)$. Here $p_{1\mu}$ is a four-momentum of the neutrino state, and in particular $p_1^2 = m_1^2$. This can be seen as consistent with the claim made earlier that propagation through spacetime is specific to the mass basis. It is only in this basis that one can connect the four-momentum of a state to a specific mass.

We are ready to consider our simple experiment. At the initial time and location we have

$$|\nu_e(0,0)\rangle = \cos\theta|\nu_1(0,0)\rangle + \sin\theta|\nu_2(0,0)\rangle \ . \tag{28.10}$$

At the detector and at time T we have

$$\begin{aligned} |\nu_e(T,L)\rangle =& e^{-ip_{1\mu}x^\mu}\cos\theta|\nu_1(0,0)\rangle + e^{-ip_{2\mu}x^\mu}\sin\theta|\nu_2(0,0)\rangle \\ =& e^{-ip_{1\mu}x^\mu}\left[\cos\theta|\nu_1(0,0)\rangle + e^{i(p_{1\mu}-p_{2\mu})x^\mu}\sin\theta|\nu_2(0,0)\rangle\right] \ , \end{aligned} \tag{28.11}$$

where in the second line we have written the state in a way that emphasizes that there is a relative phase between the two.

Let us focus on this phase to leading order in m/E. That is, let us suppose that both neutrino masses are much smaller than the energy E of the process. It is important to recognize that this energy does not correspond to either of E_1 or E_2 which are the energies of a particle in a mass eigenstate. Remember that the neutrino was not produced in a mass eigenstate. However, if the neutrino were massless then all of these energies would be equal. The differences between these energies go like m_i^2/E_i^2. This means that

$$E - E_1 \sim \frac{m^2}{E} \ , \quad E - E_2 \sim \frac{m^2}{E} \ , \tag{28.12}$$

where m is the general scale of the neutrino masses. This implies that to leading order in m/E

$$\frac{m_1^2}{E_1} = \frac{m_1^2}{E} \ , \quad \frac{m_2^2}{E_2} = \frac{m_2^2}{E} \ , \tag{28.13}$$

and

$$E_1 - E_2 \sim \frac{m^2}{E} \ . \tag{28.14}$$

Similarly, $T/L - 1 \sim m^2/E^2$. With these results in mind we find the phase is

$$\begin{aligned}
\varphi =& (p_{1\mu} - p_{2\mu})x^\mu = T(E_1 - E_2) - L(\sqrt{E_1^2 - m_1^2} - \sqrt{E_2^2 - m_2^2}) \\
=& L\left[\frac{T}{L}(E_1 - E_2) - E_1 + \frac{m_1^2}{2E_1} + E_2 - \frac{m_2^2}{2E_2} + \dots\right] \\
=& L\left[\left(\frac{T}{L} - 1\right)(E_1 - E_2) + \frac{m_1^2}{2E_1} - \frac{m_2^2}{2E_2} + \dots\right] \\
=& L\frac{m_1^2 - m_2^2}{2E} + \mathcal{O}\left(\frac{m^3}{E^3}\right) \ .
\end{aligned} \tag{28.15}$$

The process of measuring the neutrino to be electron flavored is captured by

$$|\langle \nu_e | \nu_e(T, L)\rangle|^2 \ . \tag{28.16}$$

We begin by writing (dropping the unimportant global phase)

$$\begin{aligned}
|\nu_e(T, L)\rangle &= \cos\theta|\nu_1\rangle + e^{i\varphi}\sin\theta|\nu_2\rangle \\
&= \cos\theta\left(\cos\theta|\nu_e\rangle - \sin\theta|\nu_\mu\rangle\right) + e^{i\varphi}\sin\theta\left(\sin\theta|\nu_e\rangle + \cos\theta|\nu_\mu\rangle\right) \\
&= \left(\cos^2\theta + e^{i\varphi}\sin^2\theta\right)|\nu_e\rangle + \sin\theta\cos\theta\left(e^{i\varphi} - 1\right)|\nu_\mu\rangle \ .
\end{aligned} \tag{28.17}$$

This leads to

$$\begin{aligned}|\langle\nu_e|\nu_e(T,L)\rangle|^2 &= \left(\cos^2\theta + e^{i\varphi}\sin^2\theta\right)\left(\cos^2\theta + e^{-i\varphi}\sin^2\theta\right)\\ &= \cos^4\theta + \cos^2\theta\sin^2\theta\left(e^{i\varphi}+e^{-i\varphi}\right) + \sin^4\theta\\ &= 1 - 2\cos^2\theta\sin^2\theta + 2\cos^2\theta\sin^2\theta\cos\varphi\\ &= 1 - \frac{1}{2}\sin^2(2\theta)\,(1-\cos\varphi)\\ &= 1 - \sin^2\frac{\varphi}{2}\sin^2(2\theta)\ . \end{aligned} \tag{28.18}$$

This shows that the probability of finding the neutrino to be electron flavored depends on φ, which depends on L, E, and $|m_1^2 - m_2^2|$.

Exercise 28.1 Show by direct calculation that the probability of measuring the neutrino to be muon flavored is

$$\sin^2\frac{\varphi}{2}\sin^2(2\theta)\ . \tag{28.19}$$

A similar result holds in the three flavor case with different oscillation phases between the electron and muon flavors compared to the oscillation phase for electron and tau neutrinos. The different phases depend on the difference of squared masses, similar to the two-flavor case. The experimental measurements of these oscillations confirm not just that at least two of the neutrino masses are nonzero, but also the difference of the masses. The largest of these splittings is about 2.5×10^{-3} eV^2, which implies that at least one of the neutrino masses must be at least 0.05 eV. The current direct experimental upper bound on the mass of the electron neutrino [53] is about 0.45 eV. One can also put cosmological bounds on the sum of the neutrino masses under the assumption of the standard cosmological model (ΛCDM) and that there is not a cosmological relic of other light particles. This bound [54] is about 0.13 eV. Both of these results indicate that the expansion in m/E made above is a good one.

Of course, the fact of nonzero neutrino masses implies that the kinetic term of the left-handed lepton fields must include a flavor mixing matrix. This parallels exactly the CKM matrix that appears in the left-handed quark kinetic term (24.30). For the lepton the matrix is called the PMNS matrix after Bruno Pontecorvo [55], Ziro Maki, Masami Nakagawa, and Soichi Sakata [56]. As with the CKM matrix (24.35), the magnitudes of the components of the PMNS matrix have been determined experimentally [57]

$$|V|_{\text{PMNS}} = \begin{pmatrix} 0.8215 \pm 0.0205 & 0.5495 \pm 0.0305 & 0.1485 \pm 0.0065 \\ 0.3765 \pm 0.1245 & 0.588 \pm 0.092 & 0.704 \pm 0.052 \\ 0.397 \pm 0.121 & 0.579 \pm 0.094 & 0.690 \pm 0.053 \end{pmatrix}\ , \tag{28.20}$$

where the 3σ uncertainties are shown along with the central value.

While the CKM matrix is close to diagonal with some mixing between generations, we see that the PMNS matrix generally has significant mixing among the generations. We do not currently understand the origin of either matrix's structure. However, the fact that these two matrix seem quite different in character is somewhat suggestive that each matrix might be explained in different ways.

Now that we understand that some of the neutrinos must have a mass we consider how this might occur. The simplest and most straightforward idea is to give masses to the neutrinos in the same way that we give mass to the other fermions of the Standard Model. This requires that in addition to the Higgs-fermion interactions listed in Eq. (24.9) we include

$$i\overline{N}\not{D}N + \left[(\lambda_N)^i{}_j\,\overline{L}_i\widetilde{H}N^j + \text{ H.c.}\right] \ , \tag{28.21}$$

where the N^j are a multiplet of right-handed neutrino fields. Taking this multiplet to be a triplet matches with the three-generation structure of the rest of the Standard Model. The bound mentioned above indicate that the neutrino masses are smaller than one eV. This indicates that if this Higgs interaction is the sole origin of neutrino mass then the Yukawa couplings must be very small, no larger than about 10^{-12}. One might wonder where such a small, nonzero, number comes from. In addition, such a small coupling makes detecting this right-handed neutrino in any way very difficult.

Exercise 28.2

Show that we must use the $\widetilde{H}$ version of the Higgs field to ensure that a neutrino mass term is produced. Using the quantum numbers of the Higgs and left-handed lepton fields, show that the gauge structure of the right-handed neutrino field is

$$N \sim (\mathbf{1},\, \mathbf{1},\, 0) \ . \tag{28.22}$$

As shown in the exercise above, the structure of this interaction implies that the right-handed neutrinos have no interactions with any of the Standard Model gauge fields. This means that we cannot use any of the gauge interactions to directly produce such a field. The only way this field interacts with the particles of the Standard Model is through the tiny coupling to the Higgs, making it very difficult to test this possibility.

However, this special property of being neutral under all gauge interactions allows for another very interesting possibility. This neutral nature actually allows the right-handed neutrino to have a different kind of mass term. We explain this in the following subsection.

Subsection 28.1

Majorana Fermions

The first field we discussed in this text was the real scalar. This simple field satisfies $\phi^\dagger = \phi$, which means that it cannot be rephased. A consequence of this is that it has no conserved particle number. In essence, the field ϕ is its own antiparticle. While a complex scalar field has a mass term

$$m^2\phi^\dagger\phi \ , \tag{28.23}$$

which involves both the particle and antiparticle fields, a real scalar mass term only involves one field

$$\frac{m^2}{2}\phi^2 \ . \tag{28.24}$$

It turns out that there is a type of spin-1/2 field that has similar properties. This is called a Majorana fermion after Ettore Majorana who discovered the equation that describes them [58]. What kind of Lagrangian term might this look like? We could try a spinor that is real

$$\psi^* = \psi \ , \tag{28.25}$$

but this is too restrictive. In essence, because the $(\frac{1}{2}, 0)$ and $(0, \frac{1}{2})$ representations of the Lorentz group are not real representations, we cannot make this requirement. Instead, we require that the charge conjugation of our field is equal to the original field

$$\psi \xrightarrow{C} -i\gamma^2\psi^* = \psi \ . \tag{28.26}$$

Exercise 28.3 Show that the requirement in Eq. (28.26) leads to

$$\psi_R = i\sigma_2\psi_L^* \ . \tag{28.27}$$

In order to be more efficient in writing, we introduce the notation

$$\psi^c \equiv -i\gamma^2\psi^* \ . \tag{28.28}$$

Then the Majorana condition is simply

$$\psi^c = \psi \ . \tag{28.29}$$

Exercise 28.4 Show that for any fermion fields ψ_i and ψ_j

$$\overline{\psi^c}_i\psi_j^c = \overline{\psi}_j\psi_i \ . \tag{28.30}$$

Suppose we have a Dirac field ψ_D. We define the combination

$$\psi_{M_+} = \frac{1}{\sqrt{2}}\left(\psi_D - i\gamma^2\psi_D^*\right) \ . \tag{28.31}$$

It is straightforward to check that by acting with charge conjugation we find

$$\begin{aligned}\psi_{M_+} \xrightarrow{C} & \frac{1}{\sqrt{2}}\left[-i\gamma^2\psi_D^* - i\gamma^2\left(-i\gamma^2\psi_D^*\right)^*\right] \\ &= \frac{1}{\sqrt{2}}\left[-i\gamma^2\psi_D^* + i(-i)\gamma^2(-\gamma^2)\psi_D\right] \\ &= \frac{1}{\sqrt{2}}\left[-i\gamma^2\psi_D^* + \psi_D\right] = \psi_{M_+} \ . \end{aligned} \tag{28.32}$$

In other words, ψ_{M_+} is a Majorana field.

Exercise 28.5 Show that for a Dirac fermion ψ_D that

$$\psi_{M_-} = \frac{-i}{\sqrt{2}}\left(\psi_D + i\gamma^2\psi_D^*\right) \ , \tag{28.33}$$

is also a Majorana field.

This means that we can write any Dirac field as

$$\psi_D = \frac{1}{\sqrt{2}}\left(\psi_{M_+} + i\psi_{M_-}\right) \ , \tag{28.34}$$

in terms of two Majorana fields. A Dirac mass term can then be written as

$$\begin{aligned} m\overline{\psi}_D\psi_D =& \frac{m}{2}\left(\overline{\psi}_{M_+} - i\overline{\psi}_{M_-}\right)\left(\psi_{M_+} + i\psi_{M_-}\right) \\ =& \frac{m}{2}\left(\overline{\psi}_{M_+}\psi_{M_+} + i\overline{\psi}_{M_+}\psi_{M_-} - i\overline{\psi}_{M_-}\psi_{M_+} + \overline{\psi}_{M_-}\psi_{M_-}\right) \\ =& \frac{m}{2}\left(\overline{\psi}_{M_+}\psi_{M_+} + \overline{\psi}_{M_-}\psi_{M_-}\right) \ , \end{aligned} \tag{28.35}$$

where we have used Eq. (28.30) to obtain the last line. This shows that a Dirac fermion can be thought of as two Majorana fermions with equal mass, similar to a complex scalar being expressed as two real fields with equal mass. We also find the a factor of one-half in front of the mass terms, just as we would for a real scalar mass.

We know that the Dirac mass(as discussed in Sec. 9.1) is Hermitian and Lorentz invariant. It turns out that for Majorana fields there is another type of mass that can appear in the Lagrangian. In general, we have something like

$$\psi_A M^A{}_B \psi^B = \psi^T M \psi \ , \tag{28.36}$$

where M is a constant matrix in the spinor space.

Exercise 28.6 Show that for each the Pauli matrices σ_i

$$\sigma_2 \sigma_i^T \sigma_2 = -\sigma_i \ . \tag{28.37}$$

Then, using Eq. (38.7) show that

$$e^{a_i \sigma_i^T} \sigma_2 = \sigma_2 e^{-a_i \sigma_i} \ . \tag{28.38}$$

This mass term must be Lorentz invariant. From Eq. (9.55) we find that

$$\begin{aligned} \psi_L^T \sigma_2 \psi_L \rightarrow & \psi_L e^{\frac{1}{2}(i\theta_i + \beta_i)\sigma_i^T} \sigma_2 e^{\frac{1}{2}(i\theta_i + \beta_i)\sigma_i} \psi_L \\ = & \psi_L \sigma_2 e^{-\frac{1}{2}(i\theta_i + \beta_i)\sigma_i} e^{\frac{1}{2}(i\theta_i + \beta_i)\sigma_i} \psi_L \\ = & \psi_L^T \sigma_2 \psi_L \ . \end{aligned} \tag{28.39}$$

Exercise 28.7 Show that for each the Pauli matrices σ_i

$$\psi_R^T \sigma_2 \psi_R \rightarrow \psi_R^T \sigma_2 \psi_R \ . \tag{28.40}$$

These results indicates that the Majorana mass can be proportional to

$$\gamma^0 \gamma^2 \ , \tag{28.41}$$

because we have just shown that

$$\psi^T \gamma^0 \gamma^2 \psi = -\psi_L^T \sigma_2 \psi_L + \psi_R^T \sigma_2 \psi_R \ , \tag{28.42}$$

is Lorentz invariant.

Exercise 28.8 Show that for a Majorana fermion the operator

$$i \psi^T \gamma^0 \gamma^2 \psi \ , \tag{28.43}$$

is invariant under charge conjugation and is Hermitian.

Now, clearly a mass term of the type

$$i M \psi^T \gamma^0 \gamma^2 \psi \ , \tag{28.44}$$

cannot apply to a field that carries a U(1) gauge charge. This means that no fermion of the Standard Model has a Majorana mass. However, if we include a right-handed neutrino field, which is neutral under all the Standard Model gauge groups, then such a mass would be allowed. Including a factor of one-half so that we obtain the correct normalization, the Majorana mass is then

$$\frac{1}{2} M \overline{\psi^c} \psi \ . \tag{28.45}$$

Subsection 28.2

Seesaw Mechanism

It is shown above that one possible model of neutrino masses is simply include right-handed neutrino fields N_i and use the Higgs VEV to produce a mass for the neutrinos. This required, however, very small Yukawa couplings. Suppose we allow for this type of coupling *and* a Majorana mass for the right handed neutrino fields

$$m_D \overline{\nu}_L \nu_R + \frac{1}{2} M \overline{\nu^c}_R \nu_R + \text{H.c.} \ , \tag{28.46}$$

where the Dirac mass originates from the Higgs VEV like the SM fermions

$$m_D = \frac{\lambda_\nu v}{\sqrt{2}} \ . \tag{28.47}$$

From Eq. (28.30) we know that

$$\overline{\nu}_L \nu_R = \overline{\nu^c}_R \nu^c_L \ , \tag{28.48}$$

so we can write

$$m_D \overline{\nu}_L \nu_R = \frac{1}{2} m_D \left(\overline{\nu}_L \nu_R + \overline{\nu}_L \nu_R \right) = \frac{1}{2} m_D \left(\overline{\nu}_L \nu_R + \overline{\nu^c}_R \nu^c_L \right) \ . \tag{28.49}$$

Why have we written a perfectly simple mass term in such a complicated way? This way of writing things allows us to write the combined mass (28.46) terms as

$$\frac{1}{2} \left(\overline{\nu}_L, \ \overline{\nu^c}_R \right) \mathcal{M}_\nu \begin{pmatrix} \nu^c_L \\ \nu_R \end{pmatrix} = \frac{1}{2} \left(\overline{\nu}_L, \ \overline{\nu^c}_R \right) \begin{pmatrix} 0 & m_D \\ m_D & M \end{pmatrix} \begin{pmatrix} \nu^c_L \\ \nu_R \end{pmatrix} \ . \tag{28.50}$$

Exercise 28.9 What are the eigenvalues of $\mathcal{M}_\nu$? Expand the results in $m_D/M \ll 1$. Does anything about the result surprise you?

Exercise 28.10 Consider the matrix that defines the squared masses of the fermions

$$\mathcal{M}^\dagger_\nu \mathcal{M}_\nu \ . \tag{28.51}$$

What are the eigenvalues of this matrix and how do they compare with the results of the previous exercise.

The above excises show that the combination of Dirac and Majorana masses, with $M \gg m_D$ lead to two mass eigenstates. One has a mass of

about M. The other has a mass of about

$$m_D \frac{m_D}{M} = \frac{\lambda_\nu^2 v^2}{2M} \ . \tag{28.52}$$

If this lighter state is associated with the nonzero neutrino masses then having $M \gg v$ implies that λ_ν need not be so small. This is the essence of the seesaw mechanism [59–62]. A large Majorana mass for the right-handed neutrino can explain why the neutrino states we have measured have such a small mass.

Exercise 28.11 Taking v =246 GeV, and assuming the above seesaw mechanism produces a state with about eV scale mass, what value of the Majorana mass M produces a neutrino Yukawa λ_ν of similar size to the electron Yukawa $\lambda_e \sim 10^{-5}$? What value of M produces a Yukawa about the size of the tau Yukawa $\lambda_\tau \sim 10^{-2}$?

Section 29

Strong CP Problem

In the last section we considered a physical measurement, nonzero neutrino mass, and some of the ways the standard model might be extended in order to account for it. This section focuses on something of an opposite issue. There is a term that we expect to be part of the standard model which does not seem to be there. This section is only a short introduction to this puzzle and one possible solution. More information can be found in [19, 63]

In Sec. 18 we wrote down a Lagrangian for QCD. In doing so, we neglected to write a term that is allowed by many symmetries. This term is

$$\mathcal{L}_\theta = \theta \frac{\alpha_s}{32\pi} \varepsilon^{\alpha\beta\gamma\delta} G^a_{\alpha\beta} G^a_{\gamma\delta} \ , \tag{29.1}$$

where $\alpha_s = g_s^2/(4\pi)$ and the gluon field strength is $G^a_{\alpha\beta}$. This term is often written in terms of the dual-field strength

$$\widetilde{G}^a_{\mu\nu} = \frac{1}{2} \varepsilon_{\mu\nu\alpha\beta} G^{a\alpha\beta} \ . \tag{29.2}$$

This Lagrangian term is Lorentz invariant and gauge invariant, so why did we not include it before?

Remember that the usual gluon kinetic term

$$-\frac{1}{4} G^a_{\mu\nu} G^{a\mu\nu} \ , \tag{29.3}$$

is invariant under each of C, P, and T. It is invariant under C because the gluons are in the adjoint representation of $\mathrm{SU}(3)_c$, which is a real represen-

tation. This means that C acts trivially. This term is also invariant under P and T. Recall that

$$G^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu - g_s f^{abc} A^b_\mu A^c_\nu \ , \tag{29.4}$$

and that the gluon fields A^a_μ transform like a derivative under both P and T, see Sec. 15. This means that each term in (29.3) includes an even number (at least two) instances of fields and derivatives that all transform in the same say. This means that no matter the transformation type, either one or minus one we have

$$1^2 = (-1)^2 = 1 \ , \tag{29.5}$$

so the total term is invariant.

This is not the case, however, for the θ term in (29.1). The Levi-Civita connects different types of indices together leading to terms that are odd under P and T. For example, consider a term like

$$g_s^2 f^{abc} f^{ade} \varepsilon^{\alpha\beta\gamma\delta} A^b_\alpha A^c_\beta A^d_\gamma A^e_\delta \ . \tag{29.6}$$

The epsilon tensor ensures that only one instance of time and each spatial dimension appear in each term of this sum (remember that all repeated indices are summed over). For instance, one term is

$$g_s^2 f^{abc} f^{ade} \varepsilon^{0123} A^b_0 A^c_1 A^d_2 A^e_3 \ . \tag{29.7}$$

Under parity $A_0 \to A_0$ while $A_{1,2,3} \to -A_{1,2,3}$ so in total this term changes sign. Time reversal produces $A_0 \to -A_0$ and $A_{1,2,3} \to A_{1,2,3}$ which also leads to a sign change. Similar analyses apply to each term in (29.1), showing that it is odd (changes sign) under both P and T. Usually, we simply say that this term in the Lagrangian is odd under CP.

This means that if we have a CP invariant theory then we should not include this term in the Lagrangian. Or equivalently, if we take $\theta = 0$ then no other effects generate a nonzero θ. However, we already saw in Sec. 24 that the CKM matrix allows for CP violation, which has been measured to be nonzero. So, considering that the SM as a whole does not preserve CP it seems we should include a nonzero θ term.

In fact, we can take a step back to our Abelian hypercharge vector field B_μ with field strength $B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$. Should we consider the term

$$\theta_Y \widetilde{B}^{\mu\nu} B_{\mu\nu} \ , \tag{29.8}$$

also? The usual response is "No," because even if we include this term with

nonzero θ_Y, it is a total derivative

$$\begin{aligned}\widetilde{B}^{\mu\nu}B_{\mu\nu} =&\frac{1}{2}\varepsilon^{\alpha\beta\gamma\delta}B_{\alpha\beta}B_{\gamma\delta} = \frac{1}{2}\varepsilon^{\alpha\beta\gamma\delta}\left(\partial_\alpha B_\beta - \partial_\beta B_\alpha\right)\left(\partial_\gamma B_\delta - \partial_\delta B_\gamma\right)\\ =&2\varepsilon^{\alpha\beta\gamma\delta}\left(\partial_\alpha B_\beta\right)\left(\partial_\gamma B_\delta\right) = 2\varepsilon^{\alpha\beta\gamma\delta}\left[\partial_\alpha\left(B_\beta\partial_\delta B_\gamma\right) - B_\beta\partial_\alpha\partial_\gamma B_\delta\right]\\ =&\partial_\alpha\left(2\varepsilon^{\alpha\beta\gamma\delta}B_\beta\partial_\delta B_\gamma\right)\ . \end{aligned} \tag{29.9}$$

Within the action, a total derivative produces a boundary term, an integral over the field at infinite distance. We usually assume that the fields all go to zero at infinity, so we can safely drop boundary terms. It turns out there is an important exception. In the case of nontrivial topology this specific boundary term can be nontrivial. For abelian gauge theories this can only occur when there are particles with both electric type charges (all the fields we have considered in this text) and magnetic type charges. In the SM we have no particles with magnetic hypercharge, so the boundary term is trivial and the θ_Y term can be safely set to zero.

Exercise 29.1 Write down explanations for each equality in (29.9). Work through any index manipulations to make sure you understand.

What about in the nonabelian case? The form of the gluon field strength (29.4) is more complicated, so it may be that the θ term is not a total derivative in any case. However, it turns out that

$$\widetilde{G}^a_{\mu\nu}G^{a\mu\nu} = \partial_\alpha\varepsilon^{\alpha\beta\gamma\delta}\left(A^a_\beta G^a_{\gamma\delta} + \frac{g_s}{3}f^{abc}A^a_\beta A^b_\gamma A^c_\delta\right)\ . \tag{29.10}$$

Note that this reduces to what we found above in the abelian limit.

Exercise 29.2 Starting with the right-hand side, show that Eq. (29.10) is true.

Now, it turns out to be true that the volume integral

$$\int d^4x\partial_\alpha\varepsilon^{\alpha\beta\gamma\delta}\left(A^a_\beta G^a_{\gamma\delta} + \frac{g_s}{3}f^{abc}A^a_\beta A^b_\gamma A^c_\delta\right) = \frac{8\pi}{\alpha_s}\left(n_+ - n_-\right)\ , \tag{29.11}$$

where n_- is the integer (called the winding number) that characterizes the topology of the gauge fields at $t = -\infty$ and n_+ is the winding number of the gauge fields at $t = +\infty$. In other words, the possible gauge field configurations of QCD can be classified, topologically, by an integer called the winding number. Dynamics that change winding number are associated with a nonzero θ term. In particular, the vacuum of QCD is contains contributions from *all* possible winding numbers, so it seems we should include this term in the Lagrangian and determine the value of θ experimentally.

One experimental measurement that is sensitive to θ is the electric dipole moment of the neutron, d_e. That is, if $\theta = 0$ then the neutron has

no electric dipole moment, $d_e = 0$. Experimentally, the bounds on the dipole moment are such that

$$|\theta| < 10^{-10} \ . \tag{29.12}$$

The strong CP problem is simply that we have no explanation of why θ should be this small. One could simply say that $\theta = 10^{-12}$ and that's just the way things are, but the appearance of such a small number in the Lagrangian is a bit unsettling. As with the possibility of small neutrino Yukawa couplings, significant effort has gone into alternative explanations.

Subsection 29.1

Quark Mass Connection

The analysis above is correct, but misses an important aspect of the strong CP problem. In Sec. 26 we mentioned that some classical symmetries of a Lagrangian do not persist as symmetries of the fully quantum field theory. Such symmetries are labeled anomalous. In that section we found that the current associated with an anomalous symmetry is not conserved, see Eqs. (26.6) and (26.7).

Further back, in Sec. 9.1 we found that chiral rephasing

$$\psi = e^{i\theta_c\gamma_5}\psi' \ , \tag{29.13}$$

of *massless* fermions was a symmetry of the Dirac Lagrangian. This is not a symmetry of the SM, because the quarks all get masses. But even in the massless limit this symmetry is anomalous. In the massless limit, the current J^μ associated with this symmetry satisfies

$$\partial_\mu J^\mu = \theta_c \frac{g_s^2}{16\pi^2}\widetilde{G}^a_{\mu\nu}G^{a\mu\nu} + \theta_c\frac{g^2}{16\pi^2}\widetilde{W}^a_{\mu\nu}W^{a\mu\nu} + \theta_c\frac{g'^2}{16\pi^2}\widetilde{B}_{\mu\nu}B^{\mu\nu} \ , \tag{29.14}$$

if the fermion field being rephased is charged under all of $\mathrm{SU}(3)_c\times\mathrm{SU}(2)_L\times\mathrm{U}(1)_Y$. These contributions also appear in the massive limit. The consequence is that if we include theta terms for all the SM gauge fields, then by making chiral rephasing we can change the value of the thetas. In particular, it can be shown that the $\mathrm{SU}(2)_L$ and $\mathrm{U}(1)_Y$ theta terms can be completely removed by making chiral rephasings.

In the case of the quark we are not so free to rephase. In Sec. 24 we discussed the quark mass matrices that result from electroweak symmetry breaking

$$\overline{D}_L M_d D + \overline{U}_L M_u U + \text{H.c} \ . \tag{29.15}$$

To reach the mass basis we made unitary transformations

$$U_L \to U_u U_L \ , \quad D_L \to U_d D_L \ , \quad U_R \to U_U U_R \ , \quad D_R \to U_D D_R \tag{29.16}$$

for the components of the of left-handed and right-handed quark fields. We can equivalently write all the quark mass terms as

$$\overline{\psi}_{qL}\mathcal{M}_q\psi_{qR} + \text{H.c} \ , \tag{29.17}$$

where ψ_{qL} and ψ_{qR} are six-dimensional vectors of fermions containing all left-handed and right-handed quark flavors. This six-by-six mass matrix $\mathcal{M}_q$ can be diagonalized

$$M_q = V_L^\dagger \mathcal{M}_q V_R \ , \tag{29.18}$$

by using two six dimensional unitary matrices V_L and V_R.

These unitary matrices can be expressed as SU(6) matrices multiplied by an overall phase

$$V_L = e^{i\phi_L} S_L \ , \quad V_R = e^{i\phi_R} S_R \ . \tag{29.19}$$

In general, we expect these two matrices to be different as well as the two phases. However, rephasing the left- and right-handed fermions differently corresponds to a chiral rephasing

$$\psi \to e^{i(\phi_R - \phi_L)\gamma_5/2}\psi' \ . \tag{29.20}$$

Exercise 29.3 Show that the transformation

$$\psi = e^{i\theta_c\gamma_5}\psi' \ , \tag{29.21}$$

leads to

$$\psi_L = e^{-i\theta_c}\psi'_L \ , \quad \psi_R = e^{i\theta_c}\psi'_R \ . \tag{29.22}$$

Then show that the transformations

$$\psi_L = e^{i\phi_L}\psi'_L \ , \quad \psi_R = e^{i\phi_R}\psi'_R \ , \tag{29.23}$$

amount to a total rephasing

$$\psi = e^{i(\phi_R + \phi_L)/2}\psi' \ , \tag{29.24}$$

in addition to a chiral rephasing

$$\psi \to e^{i(\phi_R - \phi_L)\gamma_5/2}\psi' \ . \tag{29.25}$$

The effect of the chiral rephasing that accompanies the mass diagonalization is to produce a term in the Lagrangian

$$-2 \times 6 \frac{\phi_R - \phi_L}{2} \frac{g_s^2}{16\pi^2} \tilde{G}^a_{\mu\nu} G^{a\mu\nu} \ , \tag{29.26}$$

where the factor of 6 comes from the number of quark flavors. This means

that the total term in the Lagrangian after the mass diagonalization is

$$[\theta + 6\,(\phi_L - \phi_R)]\,\frac{g_s^2}{16\pi^2}\widetilde{G}^a_{\mu\nu}G^{a\mu\nu}\ . \tag{29.27}$$

However, this is usually, not how this term is expressed. Recall that the diagonalized matrix M_q is completely real, so

$$\arg\left(\mathrm{Det}(M_q)\right) = 0\ , \tag{29.28}$$

where $\arg(re^{i\phi}) = \phi$. This means that

$$\begin{aligned} 0 =& \arg\left(\mathrm{Det}(M_q)\right) = \arg\left(\mathrm{Det}(V_L^\dagger \mathcal{M}_q V_R)\right) \\ =& \arg\left(e^{-i6\phi_L}\mathrm{Det}(S_L^\dagger)\right) + \arg\left(\mathrm{Det}(\mathcal{M}_q)\right) + \arg\left(e^{i6\phi_R}\mathrm{Det}(S_R)\right) \\ =& -6\phi_L + \arg\left(\mathrm{Det}(\mathcal{M}_q)\right) + 6\phi_R\ . \end{aligned} \tag{29.29}$$

In other words,

$$6(\phi_L - \phi_R) = \arg\left(\mathrm{Det}(\mathcal{M}_q)\right)\ , \tag{29.30}$$

where $\mathcal{M}_q$ is the non-diagonal quark mass matrix. So, after making the definition

$$\overline{\theta} = \theta + \arg\left(\mathrm{Det}(\mathcal{M}_q)\right)\ , \tag{29.31}$$

we find that the term in the Lagrangian (in the mass basis) is

$$\overline{\theta}\frac{g_s^2}{16\pi^2}\widetilde{G}^a_{\mu\nu}G^{a\mu\nu}\ . \tag{29.32}$$

It is this quantity $\overline{\theta}$ that is constrained by experiment to be very small, not θ itself.

Subsection 29.2

The QCD Axion

There are many ways that the strong CP problem might be resolved. One that has garnered a significant amount of attention is the axion [64, 65]. The idea is to introduce a new U(1) global symmetry (often called a Peccei-Quinn [66] symmetry) that is spontaneously broken at some scale f_a and is explicitly broken by the chiral anomaly discussed in the above section. The axion field a arises as the pNGB of the broken symmetry. It is convenient to describe the axion as the phase of a pseudoscalar field

$$\Phi(x) = r(x)e^{ia(x)/f_a}\ , \tag{29.33}$$

similar to other nonlinear descriptions of pNGBs. Note that the theory if invariant under shifts

$$a \to a + 2\pi f_a \ , \tag{29.34}$$

so it is only meaningful to discuss field values up to $2\pi f_a$.

Because a is s pNGB, we expect its couplings all to involve derivatives, like $\partial_\mu a$, so that the theory is invariant under arbitrary shifts

$$a \to a + c \ . \tag{29.35}$$

However, because the U(1) symmetry is explicitly broken by the chiral anomaly the Lagrangian contains one coupling without a derivative

$$\frac{a}{f_a}\frac{g_s^2}{16\pi^2}\widetilde{G}^a_{\mu\nu}G^{a\mu\nu} \ . \tag{29.36}$$

This means that the Lagrangian contains the total coupling

$$\left(\overline{\theta} + \frac{a}{f_a}\right)\frac{g_s^2}{16\pi^2}\widetilde{G}^a_{\mu\nu}G^{a\mu\nu} \ . \tag{29.37}$$

At low energies, this leads to the effective potential

$$V_{\rm a} = -m_\pi^2 f_\pi^2\sqrt{1 - \frac{4m_u m_d}{(m_u+m_d)^2}\sin^2\left(\frac{\overline{\theta}+a/f_a}{2}\right)} \ , \tag{29.38}$$

where m_u and m_d are the masse of the up- and down quarks, respectively. The parameter m_π is the (neutral) pion mass and $f_\pi \approx 92$ MeV is called the pion decay constant.

Exercise 29.4 Show that V_a has minima and maxima and determine what the value of a is at the minima.

The exercise above shows that the potential has an extremum whenever

$$\sin\left(\overline{\theta} + \frac{a}{f_a}\right) = 0 \ . \tag{29.39}$$

The minima turn out to be whenever

$$a = 2N\pi f_a - \overline{\theta} f_a \ , \tag{29.40}$$

where N is an integer. The other extrema are maxima. Because we are only interested in in field values less than $2\pi f_a$ the minimum of the potential, and the value of a's vacuum expectation value, is

$$\langle a\rangle = -\overline{\theta} f_a \ . \tag{29.41}$$

If we expand the term in (29.36) about the axion VEV

$$a = \langle a \rangle + \delta a \ , \tag{29.42}$$

we find

$$\left(\overline{\theta} + \frac{-f_a \overline{\theta}}{f_a} + \frac{\delta a}{f_a} \right) \frac{g_s^2}{16\pi^2} \widetilde{G}^a_{\mu\nu} G^{a\mu\nu} \ . \tag{29.43}$$

This means that the effective constant coefficient of this terms is zero, which explains why there has been no experimental measurement of a neutron dipole moment, even if $\overline{\theta}$ is not small.

While the neutron does not provide evidence of the θ term when there is an axion field there are experimental signatures. The most obvious is the pseudoscalar field δa which interacts with the gluon fields. Indeed, other interactions with SM fields are also possible and provide opportunities to discover the axion field. What are the aspects of this field? Well, it is easy to determine of the mass of δa by

$$m_a^2 = \left. \frac{d^2 V_a}{da^2} \right|_{a=\langle a \rangle} = \frac{m_\pi^2 f_\pi^2 m_u m_d}{f_a^2 (m_u + m_d)^2} \ . \tag{29.44}$$

All of the scales involved are of the MeV or hundred MeV size, except for f_a which would need to be determined experimentally. In most cases, experimental analyses require f_a to be very large implying the axion mass is very small. Still, there is a vigorous experimental program looking for evidence of axions and axion-like-particles. In addition, a background axion field could be responsible for all or part of the cosmological dark matter.

Section 30

Grand Unified Theories

In Sec. 21 we learned that larger gauge symmetries can be spontaneously broken down to smaller gauge symmetries. One might then look at the gauge structure of the standard model $\mathrm{SU}(3)_c \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ and wonder if it might come from another, perhaps simpler, gauge structure at higher energies. Pati and Salam [67] made the first steps in this direction by considering a model with gauge structure $\mathrm{SU}(4)_c \times \mathrm{SU}(2)_L \times \mathrm{SU}(2)_R$. This might be termed a partial unification.

Shortly thereafter, Georgi and Glashow [68] proposed a grand (in the sense of complete) unification model. They pointed out that an SU(5) gauge theory could undergo spontaneous symmetry breaking to produce $\mathrm{SU}(3) \times \mathrm{SU}(2) \times \mathrm{U}(1)$. The authors made a very careful discussion of how this might relate to the standard model. In this section we consider a few of these issues. However, the full story of unification is vast and detailed,

so this section should only be understood as a short introduction to some of the salient ideas. More information can be found in [69, 70]

The symmetry breaking pattern SU(5)→SU(3)×SU(2)×U(1) is a very interesting mathematical fact, but it is not sufficient to indicate that the standard model might actually arise from such a process. Suppose that at some high scale there is a fermion f that is charged under this SU(5) gauge theory. The Lagrangian for the theory would include

$$i\overline{f}\not{D}f \ , \tag{30.1}$$

where the covariant derivative is given by

$$(D_\mu)^i{}_j = \mathbb{I}^i{}_j\partial_\mu + ig_G A^a_\mu (T^a)^i{}_j \ , \tag{30.2}$$

where g_G is the grand unified coupling, the A^a_μ are the SU(5) gauge bosons, and the $(T^a)^i{}_j$ are the generators of the SU(5) group.

Let us focus on the gauge coupling. There is only one. That is part of what the unification idea means. Instead of one coupling for hypercharge, one for $SU(2)_L$, and one for QCD there is just one coupling. That is a very interesting suggestion, but at the energies we can probe the three couplings are all very different, so it seems that they cannot be unified, right?

The important idea to remember is that couplings run with energy scale, they are not fixed. We can then ask the question, is there an energy scale at which the couplings all become the same value? Figure 29 shows how the three gauge couplings in the standard model change with energy scale. These were calculated at the two-loop order and plotted in [71]. The combination $\alpha_i = g_i^2/(4\pi)$ is used because that is the combination that often naturally appears in calculations. The quantity α_i^{-1} is shown because that is the quantity that naturally appears in the equations that govern how the coupling changes scales.

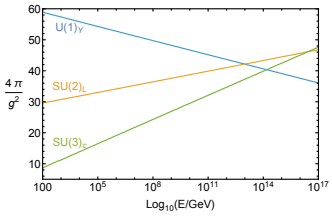


Figure 29. Two-loop running of the standard model gauge couplings. Data taken from Fig. 6.8 of [71].

The figure confirms that the hypercharge gets larger at higher energies, while the other two couplings get smaller. The couplings do all get near to each other at high energies, but they do not meet at a single point. Their intersection triangle spans the energy range from 10^{13} GeV to 10^{16} GeV. In the context of the plot this is only three orders of magnitude, but it is hard to see it as true unification.

However, we should consider what assumptions have gone into this plot. Perhaps the most important for this discussion is that we have assumed that there are no particles beyond those already discovered. Experimentally, we have explored very little of the energy scales on the plot. For certain kinds of particles we have probed up to a few TeV (10^3 GeV) for others only a few hundred GeV. In short, it is completely possible that there are as yet undiscovered particles in nature that can affect how some or all of these couplings run. For instance, in [71] it is shown that in some supersymmetric extensions of the standard model the couplings really do seem to unify at

about 10^{16} GeV. Of course, new particles might push toward or away from unification. Perhaps all we can say is that unification of the gauge couplings seems possible.

In addition to a single, unified coupling, if the high energy symmetry structure of nature is a single group gauge theory then we would expect the standard model particles to fit nicely into representations (or perhaps even a single representation) of the fundamental gauge group. In order to determine whether or not this is possible we need to be a little more definite about the structure of SU(5).

The generators of SU(5) can be separated into four types. Remember that the Lie group SU(N) has $N^2 - 1$ generators, so SU(5) has 24. In the fundamental representation we can express these generators as 5-by-5, traceless, Hermitian matrices. Most of the entries are zeros. To help guide the reader I have put lines into the matrices to separate it into blocks. For instance, the first eight generators have the form

$$T_G^{1,\ldots,8} = \left(\begin{array}{ccc|cc} & & & 0 & 0 \\ & \frac{1}{2}\lambda^{1,\ldots,8} & & 0 & 0 \\ & & & 0 & 0 \\ \hline 0 & 0 & 0 & & \\ 0 & 0 & 0 & \multicolumn{2}{c}{0} \\ & & & & \end{array}\right) , \tag{30.3}$$

where the zero in the bottom right represents a 2-by-two matrix of zeros. The $\frac{1}{2}\lambda^a$ are the 3-by-3 generators of SU(3) given in (39.125). The next 12 generators have very similar for. They are only nonzero in the off-diagonal blocks. Each pair look similar to the first two Pauli matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} , \qquad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} . \tag{30.4}$$

For instance, we have

$$T_G^9 = \frac{1}{2}\left(\begin{array}{ccc|cc} & & & 1 & 0 \\ & 0 & & 0 & 0 \\ & & & 0 & 0 \\ \hline 1 & 0 & 0 & & \\ 0 & 0 & 0 & \multicolumn{2}{c}{0} \\ & & & & \end{array}\right) , \quad T_G^{10} = \frac{1}{2}\left(\begin{array}{ccc|cc} & & & -i & 0 \\ & 0 & & 0 & 0 \\ & & & 0 & 0 \\ \hline i & 0 & 0 & & \\ 0 & 0 & 0 & \multicolumn{2}{c}{0} \\ & & & & \end{array}\right) , \tag{30.5}$$

and

$$T_G^{11} = \frac{1}{2}\left(\begin{array}{ccc|cc} & & & 0 & 1 \\ & 0 & & 0 & 0 \\ & & & 0 & 0 \\ \hline 0 & 0 & 0 & & \\ 1 & 0 & 0 & & 0 \end{array}\right) \quad , \quad T_G^{12} = \frac{1}{2}\left(\begin{array}{ccc|cc} & & & 0 & -i \\ & 0 & & 0 & 0 \\ & & & 0 & 0 \\ \hline 0 & 0 & 0 & & \\ i & 0 & 0 & & 0 \end{array}\right) \quad , \tag{30.6}$$

which only have nonzero values in the first row and first column. Generators T_G^{13} through T_G^{16} are similar, but with nonzero values in the second row and second column while T_G^{17} through T_G^{20} have nonzero values in the third row and third column.

Generators T_G^{21} through T_G^{23} are only nonzero in the bottom right, 2-by-2 block. The matrices that make up the block are exactly $\frac{1}{2}\sigma_i$, the generators of SU(2)

$$T_G^{2i} = \left(\begin{array}{ccc|cc} & & & 0 & 0 \\ & 0 & & 0 & 0 \\ & & & 0 & 0 \\ \hline 0 & 0 & 0 & & \\ 0 & 0 & 0 & & \frac{1}{2}\sigma_i \end{array}\right) \quad . \tag{30.7}$$

The last generator is diagonal

$$T_G^{24} = \sqrt{\frac{3}{5}}\left(\begin{array}{ccc|cc} & & & 0 & 0 \\ & -\frac{1}{3}\mathbb{I} & & 0 & 0 \\ & & & 0 & 0 \\ \hline 0 & 0 & 0 & & \\ 0 & 0 & 0 & & \frac{1}{2}\mathbb{I} \end{array}\right) \quad , \tag{30.8}$$

where $\mathbb{I}$ denotes the appropriate identity matrix.

Exercise 30.1 Show that all the generators satisfy

$$\mathrm{Tr}\left(T_G^A T_G^B\right) = \frac{1}{2}\delta^{AB} \quad . \tag{30.9}$$

The fact that the SU(3) and SU(2) generators show up in blocks makes it plausible that SU(5) can be broken into SU(3)×SU(2). However, it is not guaranteed. Suppose that there is some scalar field Φ that is in the fundamental representation **5** of SU(5) and has some potential that leads

to a vacuum expectation value (VEV) of the form

$$\langle \Phi \rangle = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ v \end{pmatrix} \ . \tag{30.10}$$

Every generator for which

$$T_G^A \langle \Phi \rangle = 0 \ , \tag{30.11}$$

is not spontaneously broken. All the unbroken generators correspond to the remaining symmetry group. Clearly, all of $T_G^{1,\ldots,8}$ are not broken. Neither are T_G^9, T_G^{10}, T_G^{13}, T_G^{14}, T_G^{17}, and T_G^{18}. Finally, the linear combination

$$T_G^{23} + \sqrt{\frac{5}{3}} T_G^{24} \ , \tag{30.12}$$

also satisfies (30.11). So, there are 15 unbroken generators, 8 of which clearly produce an SU(3) group. We have not proven it, but these are exactly the 15 generators of SU(4). In short, this is not the kind of symmetry breaking pattern that might produce the standard model gauge groups.

Any scalar field in the fundamental representation that gets a VEV leads to the same symmetry breaking pattern. So does any scalar field in the conjugate representation $\overline{\mathbf{5}}$. This leads us toconsider other (larger) representations. The tensor product of two fundamental representations can be decomposed into the symmetric tensor **15** and antisymmetric tensor **10** representations

$$\mathbf{5} \otimes \mathbf{5} = \mathbf{15} \oplus \mathbf{10} \ , \tag{30.13}$$

similar to the process considered in Eq. (39.243) for SU(3).

Suppose a field χ^{ij} is in one of these tensorial representations. Here the i and j indices label the five dimensions of the fundamental representation. This implies that under a transformation in which a field in the fundamental representation Φ^i transforms as

$$\Phi^i \to U^i{}_j \Phi^j \ , \tag{30.14}$$

that

$$\chi^{ij} \to U^i{}_\ell U^j{}_k \chi^{\ell k} \ . \tag{30.15}$$

Recall that these transformations $U^i{}_j$ have the form

$$U^i{}_j = \delta^i_j - i\theta^a \left(T_G^a\right)^i{}_j + \ldots \ . \tag{30.16}$$

For the fundamental field Φ^i the covariant derivative included the term

$$ig_G A^a \left(T_G^a\right)^i{}_j \ , \tag{30.17}$$

which is clearly related to the first nontrivial correction to the identity of $U^i{}_j$. The same holds true for higher tensor representations. To leading order in θ^a we find

$$U^i{}_k U^j{}_\ell = \delta^i_k \delta^j_\ell - i\theta^a \left[\delta^i_k \left(T_G^a\right)^j{}_\ell + \delta^j_\ell \left(T_G^a\right)^i{}_k\right] + \dots \ , \tag{30.18}$$

which implies that the covariant derivative for χ^{ij} is

$$\left(D_\mu \chi\right)^{ij} = \left[\delta^i_k \delta^j_\ell \partial_\mu + ig_G A^a \left(\delta^i_k \left(T_G^a\right)^j{}_\ell + \delta^j_\ell \left(T_G^a\right)^i{}_k\right)\right] \chi^{k\ell} \ . \tag{30.19}$$

Let us take some time to understand the second term in the covariant derivative. We can write the matrix part as

$$\left(\delta^i_k \left(T_G^a\right)^j{}_\ell + \delta^j_\ell \left(T_G^a\right)^i{}_k\right) \chi^{k\ell} = \chi^{i\ell} \left(T_G^a\right)^j{}_\ell + \left(T_G^a\right)^i{}_k \chi^{kj} \ . \tag{30.20}$$

The second term is already in the form of matrix multiplication. The first, however, is not. We can bring it into the usual matrix multiplication form by transposing one of the matrices. This leads to the the generalization of (30.11) for a scalar χ in a tensorial representation is

$$\langle\chi\rangle \left(T_G^a\right)^T + T_G^a \langle\chi\rangle = 0 \ , \tag{30.21}$$

where we emphasize the generator is transposed in the first term. Any symmetry whose generator satisfies this condition is not broken by the VEV of χ^{ij}.

If χ^{ij} is in the symmetric representation then there are two distinct VEV configurations. These are the trace of $\langle\chi\rangle$ and any traceless part of $\langle\chi\rangle$. The trace part can be expressed as

$$\langle\chi\rangle_{ST} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & v \end{pmatrix} \ . \tag{30.22}$$

Exercise 30.2

Show that a VEV of the form $\langle\chi\rangle_{ST}$ above does not break the symmetry associated with T_G^1 through T_G^{10}, nor T_G^{13}, T_G^{14}, T_G^{17} and T_G^{18}. Finally, show that

$$T_G^{23} + \sqrt{\frac{5}{3}} T_G^{24} \ , \tag{30.23}$$

is also a preserved symmetry.

The above exercise shows that a trace VEV produces the same SU(5)→SU(4) breaking pattern observed with the fundamental field VEV. What about a symmetric, traceless VEV? This has the form

$$\langle\chi\rangle_{S0} = \begin{pmatrix} 0&0&0&0&0\\ 0&0&0&0&0\\ 0&0&0&0&0\\ 0&0&0&0&v\\ 0&0&0&v&0 \end{pmatrix} . \tag{30.24}$$

Exercise 30.3 Show that a VEV of the form $\langle\chi\rangle_{S0}$ above does not break the symmetry associated with T^1_G through T^8_G, nor T^{23}_G. Argue that no linear combination (with real coefficients) of the other generators corresponds to an unbroken symmetry.

The exercise above shows that a symmetric, traceless VEV breaks more of the symmetry, leaving only SU(3)×U(1). This might be able to unify parts of the standard model gauge group, but clearly not all of it. Seeing as the symmetric tensor representation is not what we want, we turn to the antisymmetric tensor representation. This has the form

$$\langle\chi\rangle_{A} = \begin{pmatrix} 0&0&0&0&0\\ 0&0&0&0&0\\ 0&0&0&0&0\\ 0&0&0&0&v\\ 0&0&0&-v&0 \end{pmatrix} . \tag{30.25}$$

Exercise 30.4 Show that a VEV of the form $\langle\chi\rangle_{A}$ above does not break the symmetry associated with T^1_G through T^8_G, nor T^{21}_G, T^{22}_G, and T^{23}_G. Argue that no linear combination (with real coefficients) of the other generators corresponds to an unbroken symmetry.

This exercise shows that the antisymmetric VEV produces the symmetry breaking pattern SU(5)→SU(3)×SU(2). Again, this looks like it might account for part of the standard model gauge structure, but does not provide the full product group. What other representations might we try? Especially representations whose dimension is not too large.

Instead of taking a tensor product of two fundamental representations, we can consider the product of a fundamental and its conjugate. This produces

$$\mathbf{5} \otimes \overline{\mathbf{5}} = \mathbf{24} \oplus \mathbf{1} \ . \tag{30.26}$$

This **24** nothing other than the adjoint representation of SU(5). A scalar

field $\Sigma^i{}_j$ in the adjoint representation transforms according to

$$\Sigma^i{}_j \to U^i{}_k \Sigma^k{}_\ell U^{\dagger \ell}{}_j = \Sigma^i{}_j - i\theta^a \left[(T^a_G)^i{}_k \, \Sigma^k{}_j - \Sigma^i{}_k \, (T^a_G)^k{}_j \right] + \dots \ . \tag{30.27}$$

As we shown above for the tensor representations, this implies that condition that a VEV $\langle \Sigma \rangle$ does not break a symmetry associated with T^a_G is that

$$T^a_G \langle \Sigma \rangle - \langle \Sigma \rangle T^a_G = [T^a_G, \langle \Sigma \rangle] = 0 \ . \tag{30.28}$$

This condition clarifies how we should proceed. We recall that any field (and hence the VEV) in the adjoint representation can be expanded in the basis of the generators

$$\Sigma^i{}_j = \sum_{a=1}^{24} \Sigma_a \, (T^a_G)^i{}_j \ . \tag{30.29}$$

Any generator commutes with itself, so a residual U(1) symmetry is guaranteed to remain no matter what VEV we choose. By choose a VEV that commutes with the upper-left 3-by-3 block we produce a residual SU(3). A VEV that commutes with the lower-right 2-by-2 block produces a residual SU(2). By glancing at the generators we see that a VEV that is along the T^{24}_G direction produces exactly these characteristics.

Exercise 30.5 Show that a VEV of the form vT^{24}_G does not break the symmetry associated with T^1_G through T^8_G, nor T^{21}_G through T^{23}_G, nor T^{24}_G. Argue that no linear combination (with real coefficients) of the other generators corresponds to an unbroken symmetry.

The upshot is that a potential for $\Sigma^i{}_j$ which produces a VEV that can be parameterized as

$$\langle \Sigma \rangle^i{}_j = v \left(T^{24}_G \right)^i{}_j \ , \tag{30.30}$$

produces the symmetry breaking pattern SU(5)→SU(3)×SU(2)×U(1). This is what Georgi and Glashow determined in a more systematic way. We can now use this to determine how representations of SU(5) are decomposed into the SU(3), SU(2), and U(1) subgroups. For instance, the fundamental is decomposed as

$$\mathbf{5} = \left(\mathbf{3}, \, \mathbf{1}, \, -\frac{1}{\sqrt{15}} \right) \oplus \left(\mathbf{1}, \, \mathbf{2}, \, \frac{1}{2}\sqrt{\frac{3}{5}} \right) \ . \tag{30.31}$$

This means that a field in the fundamental of SU(5)

$$f = \begin{pmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \\ f_5 \end{pmatrix} \tag{30.32}$$

can be thought of as two fields

$$g = \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix} \ , \qquad h = \begin{pmatrix} f_4 \\ f_5 \end{pmatrix} \ , \tag{30.33}$$

where g is in the fundamental of SU(3) and the singlet representation of SU(2) while h is in the fundamental of SU(2) and singlet representation of SU(3). The U(1) charges are obtained by acting on f with T^{24}_G. One finds that g is multiplied by a factor of $-1/\sqrt{15}$ while h is multiplied by $\sqrt{3/5}/2$. It is important to remember that the numerical value U(1) charges are not physical, only their ratios. We can always absorb some factor into the definition of the coupling. So, the physically meaningful statement we can make is that

$$\frac{Y_g}{Y_h} = -\frac{2}{3} \ , \tag{30.34}$$

where I am associating the U(1) charge Y_i with our notation for hypercharge.

Exercise 30.6 Argue that we can write the decomposition of the $\overline{\mathbf{5}}$ representation as

$$\overline{\mathbf{5}} = \left(\overline{\mathbf{3}},\, \mathbf{1},\, \frac{1}{\sqrt{15}}\right) \oplus \left(\mathbf{1},\, \mathbf{2},\, -\frac{1}{2}\sqrt{\frac{3}{5}}\right) \ . \tag{30.35}$$

You may find it useful to consider the discussion of the relationship between $\mathbf{2}$ and $\overline{\mathbf{2}}$ discussed near Eq. (39.95).

We have reached the point where we can ask if the standard model fermions can be understood in terms of representations of SU(5). If we suspect that they are, it makes the most sense to write all the fields in terms of the same representation of the Lorentz group. In Sec. 24 we defined fields that were left-handed and those that were right-handed. However, we can express all the right-handed fields in terms of left-handed ones by way of the charge conjugation operator C. That is, if the right-handed up quarks are in the representation

$$U \sim \left(\mathbf{3},\, \mathbf{1},\, \frac{2}{3}\right) \ , \tag{30.36}$$

then the left-handed charge conjugated field is in the representation

$$CU \equiv U^c \sim \left(\overline{\mathbf{3}},\, \mathbf{1},\, -\frac{2}{3}\right) \ , \tag{30.37}$$

where, of course, we have used the fact that $\overline{\mathbf{1}} = \mathbf{1}$. Writing all the standard model fermions in terms of left-handed fields we have

$$Q \sim \left(\mathbf{3},\, \mathbf{2},\, \frac{1}{6}\right) \ , \quad L \sim \left(\mathbf{1},\, \mathbf{2},\, -\frac{1}{2}\right) \ , \tag{30.38}$$

and

$$U^c \sim \left(\overline{\mathbf{3}},\, \mathbf{1},\, -\frac{2}{3}\right) \ , \quad D^c \sim \left(\overline{\mathbf{3}},\, \mathbf{1},\, \frac{1}{3}\right) \ , \quad E^c \sim (\mathbf{1},\, \mathbf{1},\, 1) \ . \tag{30.39}$$

Exercise 30.7 Argue that D^c and L can belong to a single $\overline{\mathbf{5}}$ representation. Can any of the other fermion fields belong to a single $\mathbf{5}$ or $\overline{\mathbf{5}}$ representation?

The result of this exercise is encouraging! We find that two of the standard model fermion types can be unified into a single representation of SU(5). Significantly, this representation is a combination of quark and lepton fields, which is quite interesting and important. But, we also need to find a representation for the other fermions.

Because we have determined how a fundamental representation is decomposed we can determine the decomposition of larger representations. In Eq. (30.40) we recorded how two fundamental representations can be combined to make a symmetric tensor and an antisymmetric tensor. If we put in our decomposition from above we find

$$\mathbf{5} \otimes \mathbf{5} = \left[\left(\mathbf{3},\, \mathbf{1},\, -\frac{1}{3}\right) \oplus \left(\mathbf{1},\, \mathbf{2},\, \frac{1}{2}\right)\right] \otimes \left[\left(\mathbf{3},\, \mathbf{1},\, -\frac{1}{3}\right) \oplus \left(\mathbf{1},\, \mathbf{2},\, \frac{1}{2}\right)\right] \ . \tag{30.40}$$

In this relation we have used the parameterization of the U(1) charges that agree with D^c and L filling out a $\overline{\mathbf{5}}$. We can evaluate all the terms in (30.40) by simply distributing the tensor product over the direct sum. For instance, the tensor product of the first term in each direct sum is

$$\begin{aligned} \left(\mathbf{3},\, \mathbf{1},\, -\frac{1}{3}\right) \otimes \left(\mathbf{3},\, \mathbf{1},\, -\frac{1}{3}\right) &= \left(\mathbf{3} \otimes \mathbf{3},\, \mathbf{1} \otimes \mathbf{1},\, -\frac{1}{3} - \frac{1}{3}\right) \\ &= \left(\mathbf{6} \oplus \overline{\mathbf{3}},\, \mathbf{1},\, -\frac{2}{3}\right) \\ &= \left(\mathbf{6},\, \mathbf{1},\, -\frac{2}{3}\right) \oplus \left(\overline{\mathbf{3}},\, \mathbf{1},\, -\frac{2}{3}\right) \ . \end{aligned} \tag{30.41}$$

The first term in this result is the symmetric part of $\mathbf{3} \otimes \mathbf{3}$ while the second is the antisymmetric part. The correspondingly belong to the symmetric

(**15**) and antisymmetric (**10**) parts of $\mathbf{5} \otimes \mathbf{5}$.

Exercise 30.8 Show that

$$\left(\mathbf{1}, \mathbf{2}, \frac{1}{2}\right) \otimes \left(\mathbf{1}, \mathbf{2}, \frac{1}{2}\right) = (\mathbf{1}, \mathbf{3}, 1) \oplus (\mathbf{1}, \mathbf{1}, 1) \ , \tag{30.42}$$

where the first term belongs the **15** of SU(5) and the second belongs to the **10**.

The cross terms are slightly more subtle. It is true that

$$\left(\mathbf{3}, \mathbf{1}, -\frac{1}{3}\right) \otimes \left(\mathbf{1}, \mathbf{2}, \frac{1}{2}\right) = \left(\mathbf{3}, \mathbf{2}, \frac{1}{6}\right) \ , \tag{30.43}$$

and

$$\left(\mathbf{1}, \mathbf{2}, \frac{1}{2}\right) \otimes \left(\mathbf{3}, \mathbf{1}, -\frac{1}{3}\right) = \left(\mathbf{3}, \mathbf{2}, \frac{1}{6}\right) \ , \tag{30.44}$$

but these are not clearly symmetric or antisymmetric. It can help to consider the generated matrices in each case. We can express (30.43) as

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ 0 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 0 \\ 0 \\ b_1 \\ b_2 \end{pmatrix} = \left(\begin{array}{ccc|cc} & 0 & & a_1b_1 & a_1b_2 \\ & & & a_2b_1 & a_2b_2 \\ & & & a_3b_1 & a_3b_2 \\ \hline 0 & 0 & 0 & \multicolumn{2}{c}{0} \\ 0 & 0 & 0 & & \end{array}\right) \ , \tag{30.45}$$

while (30.44) is

$$\begin{pmatrix} 0 \\ 0 \\ 0 \\ c_1 \\ c_2 \end{pmatrix} \otimes \begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ 0 \\ 0 \end{pmatrix} = \left(\begin{array}{ccc|cc} & 0 & & 0 & 0 \\ & & & 0 & 0 \\ & & & 0 & 0 \\ \hline c_1d_1 & c_1d_2 & d_1d_3 & \multicolumn{2}{c}{0} \\ c_2d_1 & c_2d_2 & d_2d_3 & & \end{array}\right) \ . \tag{30.46}$$

The sum of these two fills out a matrix with 12 independent entries

$$M = \left(\begin{array}{ccc|cc} & 0 & & m_{14} & m_{15} \\ & & & m_{24} & m_{25} \\ & & & m_{34} & m_{35} \\ \hline m_{41} & m_{42} & m_{43} & \multicolumn{2}{c}{0} \\ m_{51} & m_{52} & m_{53} & & \end{array}\right) \ , \tag{30.47}$$

which can be written as the sum of a symmetric matrix M_S and an anti-

symmetric matrix M_A

$$M_S = \frac{1}{2}\left(\begin{array}{ccc|cc} & & & m_{14}+m_{41} & m_{15}+m_{51} \\ & 0 & & m_{24}+m_{42} & m_{25}+m_{52} \\ & & & m_{34}+m_{43} & m_{35}+m_{53} \\ \hline m_{14}+m_{41} & m_{24}+m_{42} & m_{34}+m_{43} & \multicolumn{2}{c}{\multirow{2}{*}{0}} \\ m_{15}+m_{51} & m_{25}+m_{52} & m_{35}+m_{53} & & \end{array}\right) , \tag{30.48}$$

$$M_A = \frac{1}{2}\left(\begin{array}{ccc|cc} & & & m_{14}-m_{41} & m_{15}-m_{51} \\ & 0 & & m_{24}-m_{42} & m_{25}-m_{52} \\ & & & m_{34}-m_{43} & m_{35}-m_{53} \\ \hline -m_{14}+m_{41} & -m_{24}+m_{42} & -m_{34}+m_{43} & \multicolumn{2}{c}{\multirow{2}{*}{0}} \\ -m_{15}+m_{51} & -m_{25}+m_{52} & -m_{35}+m_{53} & & \end{array}\right) . \tag{30.49}$$

Each of these matrices has six independent components and each is in the

$$\left(\mathbf{3},\, \mathbf{2},\, \frac{1}{6}\right) , \tag{30.50}$$

representation. So, the sum of (30.43) and (30.44) produces both a symmetric representation and an antisymmetric representation of this type.

In total then, we have shown that the symmetric representation is decomposed as

$$\mathbf{15} = \left(\mathbf{6},\, \mathbf{1},\, -\frac{2}{3}\right) \oplus \left(\mathbf{3},\, \mathbf{2},\, \frac{1}{6}\right) \oplus (\mathbf{1},\, \mathbf{3},\, 1) , \tag{30.51}$$

while the antisymmetric representation is

$$\mathbf{10} = \left(\overline{\mathbf{3}},\, \mathbf{1},\, -\frac{2}{3}\right) \oplus \left(\mathbf{3},\, \mathbf{2},\, \frac{1}{6}\right) \oplus (\mathbf{1},\, \mathbf{1},\, 1) . \tag{30.52}$$

The standard model does not include any fields in the **6** of $\mathrm{SU}(3)_c$, so we cannot use the **15** to account for the known particles. The **10**, however, is just right to account for U^c, Q, and E^c, which again groups quarks and leptons together. Perhaps even more than in the above case, that these fermions fit into this representation is quite remarkable! All the fermions of the standard model can be accounted for in terms of two representations (the $\overline{\mathbf{5}}$ and the **10**) of SU(5).

Though I do not discuss the details, SO(10) is often discussed in regard to grand unified theories. This group can break as $\mathrm{SO}(10) \to \mathrm{SU}(5) \times \mathrm{U}(1)$, which can then produce the standard model through the SU(5). Under this

breaking the **16** of SO(10) is decomposed as

$$\mathbf{16}_{\mathrm{SO}(10)} = \mathbf{10}_{\mathrm{SU}(5)} \oplus \overline{\mathbf{5}}_{\mathrm{SU}(5)} \oplus \mathbf{1}_{\mathrm{SU}(5)} \ . \tag{30.53}$$

From the analysis above we see that all the standard model fermions fit into this representation of SO(10). One must also include a new fermion field that is neutral under all the standard model gauge groups, which is exactly what might be expected of a right-handed neutrino.

In addition to the approximate unification of the gauge couplings, this structure of the standard model fermions is suggestive of unification at the very least. Remarkably, experiment can probe this possibility. Each of the SU(5) representations above contain both quarks and leptons. At energies above the symmetry breaking scale, these particles are treated in the same way. Below the symmetry breaking scale the particles appear very different, but they are still connected by massive vector bosons X_μ related to the broken symmetry generators. This is similar to the way the left-handed quarks interact through the $W^\pm_\mu$ bosons.

In short, in grand unified theories of this type, leptons and quarks can change into one another. For instance, baryon number can be violated such that the proton can decay into a pion and a positron.

$$p \to \pi^0 e^+ \ . \tag{30.54}$$

Very schematically, the decay width is given by

$$\Gamma_p \sim \frac{g_G^4 m_p^5}{m_X^4} \ , \tag{30.55}$$

where $m_p \sim 1$ GeV is the proton mass and m_X is the mass of the vector bosons associated with the broken symmetries. From Fig. 29 we can estimate

$$g_G \sim 0.55 \ , \quad m_X \sim 5 \times 10^{15} \text{ GeV} \ . \tag{30.56}$$

Exercise 30.9 Show that under the above assumptions that the proton decay width is about

$$\Gamma_p \sim 10^{-60} \text{ GeV} \ , \tag{30.57}$$

and that this corresponds to a lifetime of about

$$\tau_p \sim 2 \times 10^{28} \text{ years.} \tag{30.58}$$

The estimated lifetime above is very long, much longer than the 1.38×10^{10} year age of the Universe. However, there are a lot of protons. The experimental bounds [72] currently limit the lifetime from this process to

$$\tau_p > 2.4 \times 10^{34} \text{ years,} \tag{30.59}$$

which is about six orders of magnitude to large. Our estimate here was very rough and using a specific model can increase the proton lifetime. The general point is that even physics associated with very high energy scales can sometimes be probed. In this case, the prediction of proton decay is so radical (and we can collect a great many protons to observe) that some very high scale effects can be excluded.

Section 31

Electroweak Hierarchy Problem

The Grand Unified Theories discussed in the above section introduced a very high energy scale, about 10^{15} GeV, into beyond the standard model explorations. In order to make precise predictions to compare with experiment (such as the proton lifetime) physicists began to make quantum field theory calculations that included these high scales (see for instance [73, 74]). They found that when quantum field theory corrections were included that scalar fields "generally" developed masses on the order of the high mass scale. More specifically, when one calculates the value of the Higgs VEV in a grand unified theory, one finds the VEV is near 10^{15} GeV, unless the parameters are very finely adjusted.

This characteristic of grand unified theories was recognized to apply more generally [75]. Unless there is some structure that prevents it, scalar fields generally develop masses on the order of the highest scale they couple to. In other words, scalar fields are generally sensitive to high physics scales. We have not developed the quantum field theory required to show this sensitivity. However, we can use a very general result to develop some understanding.

Coleman and Weinberg [76] showed that the quantum corrections to the potential $U(\phi_1, \phi_2, \ldots)$ of a scalar fields ϕ_i can be calculated in a very general way. We can then simply use their calculation and apply it to a variety of circumstances. They found that the contributions from a set of scalars is

$$V_S(\phi_i) = \frac{\Lambda^2}{16\pi^2}\mathrm{Tr}\,\mathcal{M}_S^2 + \frac{1}{32\pi^2}\mathrm{Tr}\left[\mathcal{M}_S^4\left(\ln\frac{\mathcal{M}_S^2}{\Lambda^2} - \frac{1}{2}\right)\right] \ . \tag{31.1}$$

In this result Λ is a parameter with dimensions of energy (or mass) that is known as the cutoff scale. It is the energy value at which the theory must be significantly modified, say be the introduction of new particles or dynamics. An example is Fermi's theory of β-decay. In that case the cutoff is m_W, because at that scale we cannot neglect the effects of the $W_\mu^\pm$ field.

The object $\mathcal{M}_S^2$ is a matrix of all the squared masses of the scalar fields. However, the masses are taken to be functions of the scalars. For example,

suppose our Lagrangian is

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi + \frac{1}{2}\partial_\mu\chi\partial^\mu\chi - \frac{m_\phi^2}{2}\phi^2 - \frac{m_\chi^2}{2}\chi^2 - \frac{\kappa}{2}\phi^2\chi^2 - \frac{\lambda_\phi}{4}\phi^4 - \frac{\lambda_\chi}{4}\chi^4 \ . \quad (31.2)$$

The components of mass squared matrix are defined by taking derivatives with respect to the scalar fields. For instance, the $\phi\phi$ component is

$$\left(\mathcal{M}_S^2\right)_{\phi\phi} = -\frac{\partial^2\mathcal{L}}{\partial^2\phi^2} = m_\phi^2 + \kappa\chi^2 + 3\lambda_\phi\phi^2 \ . \quad (31.3)$$

Exercise 31.1

Show that the off diagonal components of the scalar mass squared matrix are

$$\left(\mathcal{M}_S^2\right)_{\phi\chi} = \left(\mathcal{M}_S^2\right)_{\chi\phi} = -\frac{\partial^2\mathcal{L}}{\partial^2\phi^2} = 2\kappa\phi\chi \ , \quad (31.4)$$

while the remaining diagonal term is

$$\left(\mathcal{M}_S^2\right)_{\chi\chi} = -\frac{\partial^2\mathcal{L}}{\partial^2\chi^2} = m_\chi^2 + \kappa\phi^2 + 3\lambda_\chi\chi^2 \ . \quad (31.5)$$

The trace of this matrix is simply

$$\mathrm{Tr}\,\mathcal{M}_S^2 = \left(\mathcal{M}_S^2\right)_{\phi\phi} + \left(\mathcal{M}_S^2\right)_{\chi\chi} \ . \quad (31.6)$$

Therefore, the contribution from the first term in (31.1) is

$$\frac{\Lambda^2}{16\pi^2}\left[m_\phi^2 + m_\chi^2 + \phi^2\left(\kappa + 3\lambda_\phi\right) + \chi^2\left(\kappa + 3\lambda_\chi\right)\right] \ . \quad (31.7)$$

These are the leading contributions to the scalar potential from quantum effects. The first two terms are constants, so they are typically dropped. The second and third terms contribute to the masses of ϕ and χ, respectively. That is, the total mass term for ϕ, including the classical potential and the leading one-loop corrections, is

$$\frac{M_\phi^2}{2}\phi^2 = \frac{1}{2}\phi^2\left[m_\phi^2 + \frac{\Lambda^2}{8\pi^2}\left(\kappa + 3\lambda_\phi\right)\right] \ . \quad (31.8)$$

There are also contributions from the second term in (31.1), but we neglect these for now.

Suppose that the cutoff for this theory is some very high scale, like $\Lambda \sim 10^{15}$ GeV. The result in (31.8) shows that even if m_ϕ is at some low scale (like 100 GeV) the quantum corrections to the scalar mass completely dominate, and drag the physical mass up to the high scale. In the toy model that we are considering here there are two contributions to this quantum

mass term. One, with coefficient $3\lambda_\phi$ comes from ϕ's interactions with itself. The other, with coefficient κ, comes from its interactions with χ.

We could also include the remaining (leading) quantum corrections. These include traces over

$$\mathcal{M}_S^4 = \mathcal{M}_S^2 \mathcal{M}_S^2 \ , \tag{31.9}$$

and

$$\mathcal{M}_S^4 \ln \frac{\mathcal{M}_S^2}{\Lambda^2} \ . \tag{31.10}$$

Evaluating the trace of the squared matrix is straightforward, but what are we to make of this last object? Its trace is most easily understood in terms of the eigenvalues of the mass squared matrix. The trace of any matrix is independent of basis, and these is some basis in which

$$\mathcal{M}_S^2 = \begin{pmatrix} m_+^2 & 0 \\ 0 & m_-^2 \end{pmatrix} \ , \tag{31.11}$$

where $m_\pm^2$ are the eigenvalues of $\mathcal{M}_S^2$. In this same basis

$$\mathcal{M}_S^4 = \begin{pmatrix} m_+^4 & 0 \\ 0 & m_-^4 \end{pmatrix} \ , \tag{31.12}$$

and

$$\ln \frac{\mathcal{M}_S^2}{\Lambda^2} = \begin{pmatrix} \ln \frac{m_+^2}{\Lambda^2} & 0 \\ 0 & \ln \frac{m_-^2}{\Lambda^2} \end{pmatrix} \ . \tag{31.13}$$

Therefore, in any basis, we know that

$$\mathrm{Tr}\, \mathcal{M}_S^4 \ln \frac{\mathcal{M}_S^2}{\Lambda^2} = m_+^4 \ln \frac{m_+^2}{\Lambda^2} + m_-^4 \ln \frac{m_-^2}{\Lambda^2} \ . \tag{31.14}$$

If the eigenvalues are near the electroweak scale $m_\pm \sim 100$ GeV and the cutoff is near 10^{15} GeV then

$$\ln \frac{m_+^2}{\Lambda^2} \approx -25 \ . \tag{31.15}$$

So, the first contribution to the ϕ mass in the Coleman-Weinberg potential would go like 10^{30} GeV^2 while the second is about 25×10^4 GeV^2, which is clearly a small correction.

Exercise 31.2 Determine the complete quantum corrected mass for χ, including both terms in the Coleman-Weinberg potential (31.1).

This first example has shown that a scalar field can receive large corrections to its mass from other scalars. However, fermions and vectors also

contribute to the quantum potential. From fermions we have

$$V_F(\phi_i) = -\frac{\Lambda^2}{8\pi^2}\mathrm{Tr}\,\mathcal{M}_F^2 - \frac{1}{16\pi^2}\mathrm{Tr}\left[\mathcal{M}_F^4\left(\ln\frac{\mathcal{M}_F^2}{\Lambda^2} - \frac{1}{2}\right)\right] , \tag{31.16}$$

and for vectors we have

$$V_V(\phi_i) = \frac{3\Lambda^2}{16\pi^2}\mathrm{Tr}\,\mathcal{M}_V^2 + \frac{3}{32\pi^2}\mathrm{Tr}\left[\mathcal{M}_V^4\left(\ln\frac{\mathcal{M}_V^2}{\Lambda^2} - \frac{1}{2}\right)\right] . \tag{31.17}$$

Clearly, these are quite similar to the scalar contribution, but the signs and coefficients of the terms can differ.

The contributions to a scalar mass follow a similar form to the above analysis. An important aspect of the fermionic calculation is that for a set of fermions f_i we first calculate the fermion mass matrix

$$(\mathcal{M}_F)_{ij} = -\frac{\partial^2\mathcal{L}}{\partial\overline{f}_i\partial f_j} . \tag{31.18}$$

The mass squared matrix is then given by

$$\mathcal{M}_F^2 = \mathcal{M}_F^\dagger\mathcal{M}_F . \tag{31.19}$$

For example, suppose a theory with two fermions f_1 and f_2 along with a real scalar field ϕ. We take the Lagrangian to have the form

$$\begin{aligned}\mathcal{L} = &\frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{m_\phi^2}{2}\phi^2 - \frac{\lambda_\phi}{4}\phi^4 + \overline{f}_1\left(i\gamma^\mu\partial_\mu - m_1\right)f_1 + \overline{f}_2\left(i\gamma^\mu\partial_\mu - m_2\right)f_2 \\ &- \lambda_1\phi\overline{f}_1 f_1 - \lambda_2\phi\overline{f}_2 f_2 - \lambda_{12}\phi\overline{f}_1 f_2 + \text{H.c.} .\end{aligned} \tag{31.20}$$

Note that the Hermitian conjugation on the last line essentially implies that λ_1 and λ_2 are real numbers. While this argument does not imply that λ_{12} is real, by rephasing f_1 and f_2 we can effectively remove any imaginary part.

Exercise 31.3 Show that

$$\mathcal{M}_F = \begin{pmatrix} m_1 + \lambda_1\phi & \lambda_{12}\phi \\ \lambda_{12}\phi & m_2 + \lambda_2\phi \end{pmatrix} , \tag{31.21}$$

and hence that

$$\mathcal{M}_F^2 = \begin{pmatrix} \left(m_1 + \lambda_1\phi\right)^2 + \lambda_{12}^2\phi^2 & \lambda_{12}\phi\left(m_1 + m_2 + \lambda_1\phi + \lambda_2\phi\right) \\ \lambda_{12}\phi\left(m_1 + m_2 + \lambda_1\phi + \lambda_2\phi\right) & \left(m_2 + \lambda_2\phi\right)^2 + \lambda_{12}^2\phi^2 \end{pmatrix} . \tag{31.22}$$

By taking the trace of $\mathcal{M}_F^2$ we can find the leading quantum corrections to the ϕ potential due to its interactions with the fermions. This does

generate a mass term, but also we find a term linear in ϕ

$$-\phi\frac{\Lambda^2}{4\pi^2}\left(m_1\lambda_1 + m_2\lambda_2\right) \ . \tag{31.23}$$

Such a contribution to the potential is very interesting, because the minimum of the total potential is given by

$$0 = \frac{d}{d\phi}\left(\frac{m_\phi^2}{2}\phi^2 + \frac{\lambda_\phi}{4}\phi^4 + V_S(\phi) + V_F(\phi)\right) \ . \tag{31.24}$$

Before we considered the quantum correction the minimum of the ϕ potential was clearly $\phi = 0$, which led to the vacuum expectation value $\langle\phi\rangle = 0$. By including the quantum effects of the fermions we find that the minimum, and hence the VEV, is shifted away from zero.

Exercise 31.4 Show that the leading contribution to the ϕ mass is

$$\frac{M_\phi^2}{2}\phi^2 = \frac{1}{2}\phi^2\left[m_\phi^2 + 3\lambda_\phi\frac{\Lambda^2}{8\pi^2} - \left(\lambda_1^2 + \lambda_2^2 + 2\lambda_{12}^2\right)\frac{\Lambda^2}{4\pi^2}\right] \ . \tag{31.25}$$

The above exercise illustrates that the effects of the scalar self-interaction and its interactions with the fermions give opposite sign contributions to the scalar mass. One might immediately consider situations in which the two quantum effects cancel. Clearly, they can cancel, but one has to choose all of the couplings just right. This requires "tuning" the parameters, unless there is some other structure in the theory (such as a symmetry) that ensures the two terms cancel.

Having developed some familiarity with the Coleman-Weinberg corrections to scalar potentials, we can apply this to the Higgs field H of the Standard Model. The relevant parts of the Lagrangian are

$$\begin{aligned}\mathcal{L} = &\left(D_\mu H\right)^\dagger D^\mu H + \mu^2|H|^2 - \lambda|H|^4 \\ &+ \left(\lambda_\ell\right)^i{}_j\,\overline{L}_i H E^j + \left(\lambda_d\right)^i{}_j\,\overline{Q}_i H D^j + \left(\lambda_u\right)^i{}_j\,\overline{Q}_i\widetilde{H}U^j + \text{ H.c.} \ ,\end{aligned} \tag{31.26}$$

where

$$D_\mu H = \left(\mathbb{I}\partial_\mu + i\frac{g}{2}\sigma_a A_\mu^a + i\frac{g'}{2}\mathbb{I}B_\mu\right)H \ , \tag{31.27}$$

and the σ_a are the Pauli matrices. In calculating the various mass matrices we must include something that has not come up in the simple examples above. When calculating the terms in the mass matrices, all nonscalar fields should be set to their vacuum expectation value (which is zero for fields with spin) after the derivatives are taken. This is why this process only produces a potential for the scalar fields. This is also why the other parts of the standard model do not play a role in this calculation.

Let us begin with the scalar contribution. We calculate

$$\mathcal{M}_S^2 = -\frac{\partial^2 \mathcal{L}}{\partial H^\dagger \partial H} = -\mu^2 + 4\lambda |H|^2 \ . \tag{31.28}$$

There are not contributions from the vector fields that appear in the covariant derivative because these are set to their VEV, which is zero.

The vector contribution comes entirely from the covariant derivative of the Higgs field. Note that

$$\frac{\partial D_\mu H}{\partial A_\mu^a} = i\frac{g}{2}\sigma_a H \ , \qquad \frac{\partial (D_\mu H)^\dagger}{\partial A_\mu^a} = -i\frac{g}{2} H^\dagger \sigma_a \ . \tag{31.29}$$

Exercise 31.5 Using the result above, show that

$$\mathcal{M}_V^2 = \frac{1}{2}\begin{pmatrix} g^2|H|^2 & 0 & 0 & 0 \\ 0 & g^2|H|^2 & 0 & 0 \\ 0 & 0 & g^2|H|^2 & 0 \\ 0 & 0 & 0 & g'^2|H|^2 \end{pmatrix} \ . \tag{31.30}$$

The fermion mass matrix is the most intricate. In reality, there are a great many fermions in the standard model, but they are usefully organized into representations of gauge groups and into generations. We can define a schematic fermion mass matrix with the rows labeled by the left-handed fields and the columns by the right-handed ones. We find

$$\mathcal{M}_F = \begin{pmatrix} \lambda_\ell H & 0 & 0 \\ 0 & \lambda_d H & \lambda_u \widetilde{H} \end{pmatrix} \ , \tag{31.31}$$

where the top row is the left-handed leptons and first column is the right-handed leptons. The second column is the right-handed down-type quarks. Each of the Yukawa terms is a matrix in the space of generations.

Exercise 31.6 Recall that

$$\widetilde{H}^i = \varepsilon^{ij} H_j^\dagger \ . \tag{31.32}$$

Prove that

$$H_i^\dagger \widetilde{H}^i = 0 \ , \qquad \widetilde{H}_i^\dagger H^i = 0 \ . \tag{31.33}$$

Keeping the results of the above exercise in mind, one find the mass squared matrix is

$$\mathcal{M}_F^2 = \begin{pmatrix} \lambda_\ell^\dagger \lambda_\ell |H|^2 & 0 & 0 \\ 0 & \lambda_d^\dagger \lambda_d |H|^2 & 0 \\ 0 & 0 & \lambda_u^\dagger \lambda_u |H|^2 \end{pmatrix} \ . \tag{31.34}$$

In this matrix each Yukawa matrix is three-by-three, so the full matrix

here is nine-by-nine in generation space. The quark colors add additional numbers of fermions. Recall that the trace of a matrix is simply the sum of its eigenvalues. This means that, for instance,

$$\mathrm{Tr}\,\lambda_\ell^\dagger\lambda_\ell = \lambda_e^2 + \lambda_\mu^2 + \lambda_\tau^2 \ , \tag{31.35}$$

the sum of the squares of the leptons Yukawa couplings.

Putting all of these contributions together we find that the leading contribution to the Higgs potential includes the mass term

$$|H|^2\left[-\mu^2 + \frac{\Lambda^2}{8\pi^2}\left(2\lambda + \frac{9}{4}g^2 + \frac{3}{4}g'^2 - \lambda_e^2 - \lambda_\mu^2 - \lambda_\tau^2 \right.\right. \\ \left.\left. - 3\lambda_d^2 - 3\lambda_s^2 - 3\lambda_b^2 - 3\lambda_u^2 - 3\lambda_c^2 - 3\lambda_t^2\right)\right] \ . \tag{31.36}$$

Of all the quantum contributions, that of the top quark is the largest, because its Yukawa coupling (and hence mass) is the largest. What we have extracted from experiment is the total value, which we might call $\mu^2_{\rm Phy}$, of the term is square brackets. We have also measured the values of all the couplings that appear within the parenthesis. The two quantities we do not know experimentally are $-\mu^2$ and Λ. In fact, we do not even know the sign of the classical potential term so it might be better to label it $\pm\mu^2$.

Suppose that the standard model is the correct model of nature up to very high energies. How high might we imagine? Our analysis of gravity suggests that it is an effective theory, but must be modified at energies near the Planck scale $M_{\rm Pl} \approx 2.435 \times 10^{18}$ GeV. If we take this as our largest possible cutoff then we schematically have

$$\pm\mu^2 - \frac{3}{8\pi^2}M_{\rm Pl}^2 = -\mu^2_{\rm Phy} \ , \tag{31.37}$$

where we have used $\lambda_t \approx 1$. The only way for this to make any sense is to choose the positive sign for the classical potential term and to have the two terms very nearly cancel. However, this is a very mysterious cancellation. Within the standard model (assuming no other structure) the mass term in the classical Higgs potential and the other couplings of the standard model Lagrangian are independent, they have no relation to each other. It seems very odd that they would conspire to cancel with such amazing precision.

Just how amazing is this cancellation? We can get some idea by looking at explicit number in the units of GeV^2. Suppose that

$$\frac{3}{8\pi^2}M_{\rm Pl}^2 \times \ \text{couplings} \approx 855974793291057288616973800412328 80 \ \text{GeV}^2 \tag{31.38}$$

is the number produced by quantum effects (the digits are simply random

numbers used to illustrate a point). Current measurement indicate the standard model Higgs boson mass is about $m_h = 125$ GeV and the Higgs VEV is about $v = 246$ GeV. This implies

$$\lambda = \frac{m_h^2}{2v^2} \approx 0.13 \ , \qquad \mu = \frac{m_h}{\sqrt{2}} \approx 88 \text{ GeV} \ . \tag{31.39}$$

So, we have that

$$\mu_{\text{Phy}} \approx 7813 \text{ GeV}^2 \ . \tag{31.40}$$

The difference in the magnitudes of this number and (31.38) is conspicuous. To satisfy the equation for μ_{Phy} in (31.37) the Lagrangian parameter μ^2 must exactly match the first 31 digits of (31.38). While there is no logical inconsistency with this level of agreement, the fact that these are expected to be independent numbers make this cancellation seem highly improbable.

This apparent fine-tuning of parameters that seems to be required in the above scenario has driven considerable effort in building extensions of the standard model. One way to reduce the fine tuning, is simply to reduce Λ to a much lower scale. For instance, if the Higgs field is not elementary, but rather a composite of tightly bound fermions [75, 77] then the confinement scale of this new force becomes the cut off of the model. Of course, if the Higgs is a bound state, then we might expect other similar bound states at about the same mass, which we have not seen any hint of. A composite Higgs with a much lighter mass than the other bound states is possible, and natural, if the Higgs is a pseudo-Nambu-Goldstone boson, somewhat like the pions. This means that along with a new force to bind new fermions, there is also some new approximate global symmetry.

The cut off may also be much lower than the Planck scale if there are additional spatial dimensions beyond the large ones we are familiar with. We might not presently be aware of these additional spatial dimensions, which may be largely flat [78, 79] or warped [80, 81]. At low energies (which include all the energies we have probed at colliders) we presently see no sign of these additional dimension, though searches continue. It still may be that these dimensions become apparent at a scale much lower than the Planck mass, and have the potential to explain why the Higgs mass parameter is so low.

Finally, symmetries can explain why the Higgs mass parameter is not sensitive to high scales. We saw above that scalar and fermionic contributions to the Higgs potential come with opposite sign. In supersymmetric [71] theories the fermions in the standard model are related by supersymmetry to scalars, which largely cancel the large corrections to the mass parameter. A similar story applies to the pseudo-Nambu-Goldstone boson Higgs theories mentioned above. In order for these models to explain the observed hierarchy there must be symmetry partners of at least some of the standard model fermions, and especially the top quark. That is

why top partners have been carefully sought for experimentally. There are several other interesting ideas that have been explored for explaining the electroweak hierarchy, which are surveyed in [82].

PART

VIII

Physics Prerequisites

This part of the text provides a minimal introduction to variational methods and quantum mechanics. It is meant to be enough to make the rest of the book comprehensible but it is nowhere near a complete or comprehensive treatment of these topics. Additional study is strongly recommended.

Sec. 32. Variational Methods
Sec. 33. Quantum Essentials

Section 32

Variational Methods

The calculus of variations is a powerful set of tools for determining the equations that govern a physical system. Its power comes from the flexibility it allows for choosing coordinates. This makes variational methods the preferred tool for many difficult problems. However, when first introduced to this formalism it can seem unusual and mysterious. We attempt to show that they are no more mysterious than Newton's 2nd Law.

Remember that Newton's 2nd law is often written as

$$m\ddot{\vec{x}} = -\nabla U(\vec{x}) \ , \tag{32.1}$$

where the dot denotes a time derivative and $\vec{x}(t)$ is the path of some object through three-dimensional space as a function of time. The potential energy $U(\vec{x})$ is a function of this trajectory and is related to the typical notion of a force by

$$\vec{F} = -\nabla U(\vec{x}) \ . \tag{32.2}$$

In the index notation described in Sec. 35 Newton's 2nd law is

$$m\,\ddot{x}_i = -\partial_i U(x) \ . \tag{32.3}$$

What is the physical justification for this equation? Simply that it works. In that sense it is a mysterious piece of information that we do not derive, we simply use.

Definition

With this in mind, we define a somewhat abstract quantity called the **action** and often denoted by S. This quantity has dimensions of angular momentum. It is often defined in terms of an integral

$$S = \int_{t_i}^{t_f} dt\, L \ , \tag{32.4}$$

where L is called the Lagrangian.

For the simple system we are considering, that of a particle moving under the influence of a force, the Lagrangian is defined to be the kinetic energy minus the potential energy:

$$\begin{aligned} L =& \frac{1}{2} m \dot{\vec{x}} \cdot \dot{\vec{x}} - U(\vec{x}) \\ =& \frac{m}{2} \dot{x}^i \dot{x}_i - U(x) \ . \end{aligned} \tag{32.5}$$

Notice that S is a number. Any specific function of time $x^i(t)$ produces a number. This means that $S[x^i(t)]$ is a **functional** of $x^i(t)$, that is a mapping from the space of functions to the space of numbers.

Suppose that we are interested in the value of the action at a special trajectory $x_c^i(t)$ which we call the **classical trajectory**. This is the physical path, the actual true path, the object takes from time t_i to t_f. We do not yet know what this path is, but we can give it a name in the abstract. Of course, this simply leads to

$$S[x_c^i(t)] \equiv S_c = \int_{t_i}^{t_f} dt \, \left(\frac{m}{2} \dot{x}_c^i \dot{x}_{ci} - U(x_c) \right) \ . \tag{32.6}$$

As a brief aside, suppose you were investigating some function $f(x)$ and were interested in knowing if a special point x_c was an extremum, either a maximum or a minimum of $f(x)$. You might expand the function about x_c in a Taylor series

$$f(x_c + \delta x) = f(x_c) + \delta x \left. \frac{df}{dx} \right|_{x=x_c} + \frac{1}{2} (\delta x)^2 \left. \frac{d^2 f}{dx^2} \right|_{x=x_c} + \dots \ , \tag{32.7}$$

where δx is some small deviation from x_c. If x_c is an extremum of $f(x)$ then $\left. \frac{df}{dx} \right|_{x=x_c} = 0$. Then, to **leading order** in δx we find that

$$f(x_c + \delta x) = f(x_c) + \mathcal{O}(\delta x^2) \ , \tag{32.8}$$

where the second term denotes all terms that are at least quadratic in δx.

Exercise 32.1 Remind yourself that the form of a Taylor series expansion of a function of a vector $f(x_c^i + \delta x^i)$ is

$$f(x_c^i + \delta x^i) = f(x_c^i) + \delta x^i \partial_i f \Big|_{x=x_c} + \frac{1}{2} \, \delta x^i \delta x^j \partial_i \partial_j f \Big|_{x=x_c} + \dots \ . \tag{32.9}$$

We use this same way of thinking to examine the action. We define a

function $\delta x^i(t)$ that is small, $\delta x^i(t) \ll 1$ for all t. We also assume that

$$\delta x^i(t_i) = 0 \ , \quad \delta x^i(t_f) = 0 \ . \tag{32.10}$$

Then the sum

$$x_c^i(t) + \delta x^i(t) \ , \tag{32.11}$$

is a function that is close to the classical trajectory and equal to it at the initial and final times. In essence we are considering trajectories that begin and end at the same place but in between can vary a little from the classical path. With this we can examine how the action responds to small deviations away from the classical trajectory.

We are interested only in the leading order behavior, so we drop terms of order δx^2 and higher throughout the derivation. We find

$$\begin{aligned} S[x_c^i(t) + \delta x^i(t)] &= \int_{t_i}^{t_f} dt \ \left[\frac{m}{2} \left(\dot{x}_c^i + \dot{\delta x}^i \right) \left(\dot{x}_{ci} + \dot{\delta x}_i \right) - U(x_c + \delta x) \right] \\ &= \int_{t_i}^{t_f} dt \ \left[\frac{m}{2} \left(\dot{x}_c^i \dot{x}_{ci} + 2\dot{x}_c^i \dot{\delta x}_i + \mathcal{O}(\delta x^2) \right) \right. \\ &\qquad \left. -U(x_c) - \delta x^i \partial_i U(x)\Big|_{x=x_c} - \mathcal{O}(\delta x^2) \right] \\ &= S_c + \int_{t_i}^{t_f} dt \ \left(m\dot{x}_c^i \dot{\delta x}_i - \delta x^i \partial_i U(x)\Big|_{x=x_c} \right) + \mathcal{O}(\delta x^2) \ . \end{aligned} \tag{32.12}$$

We can integrate the $\dot{\delta x}_i$ term by parts

$$\int_{t_i}^{t_f} dt \, m\dot{x}_c^i \dot{\delta x}_i = m\dot{x}_c^i \delta x\Big|_{t_i}^{t_f} - \int_{t_i}^{t_f} dt \, m\ddot{x}_c^i \delta x_i \ . \tag{32.13}$$

The boundary term vanishes because of the boundary conditions assumed in Eq. (32.10). This leads to

$$S[x_c^i(t) + \delta x^i(t)] = S_c - \int_{t_i}^{t_f} dt \, \delta x^i \left(m\ddot{x}_c^i + \partial_i U(x)|_{x=x_c} \right) + \mathcal{O}(\delta x^2) \ . \tag{32.14}$$

The first derivative of a function vanishes at an extremum. We are seeking a similar condition on S, that its first order variation vanishes at the classical trajectory x_c^i, or that

$$S[x_c^i(t) + \delta x^i(t)] = S_c \tag{32.15}$$

for any δx^i. This can only be the case for all deviations δx^i if

$$m\ddot{x}_c^i + \partial_i U(x)|_{x=x_c} = 0 \ , \tag{32.16}$$

which is exactly Newton's 2nd law, (32.3).

Let us pause to make clear what this derivation is supposed to mean. We have shown that Newton's 2nd law is equivalent to requiring that the action evaluated at the classical trajectory is stationary, or that its small variations vanish. This requirement is sometimes known as **Hamilton's principle**. This principle can then replace Newton's 2nd law in more complex situations.

Consider we simply have a physical system that we characterize by some kind of coordinates $q^i(t)$. These coordinates can be nearly anything we need them to be, which is where the power of this formalism comes from. Assuming we can determine the correct form of the Lagrangian (which is not obvious) we consider it as a function of these coordinates and their time derivatives

$$L(q^i,\, \dot{q}^i) \ . \tag{32.17}$$

This function can be expanded to linear order in deviations of these coordinates from the classical trajectory $q^i_c + \delta q^i$

$$L(q^i_c + \delta q^i,\, \dot{q}^i_c + \dot{\delta q}^{\,i}) = L(q^i_c,\, \dot{q}^i_c) + \delta q^i \frac{\partial L}{\partial q^i} + \dot{\delta q}^{\,i} \frac{\partial L}{\partial \dot{\delta q}^{\,i}} + \mathcal{O}(\delta q^2) \ . \tag{32.18}$$

Exercise 32.2

Using the methods shown in the derivation above show that

$$S[q^i_c(t) + \delta q^i(t)] = S_c + \int_{t_i}^{t_f} dt\, \delta q^i \left(\frac{\partial L}{\partial q^i} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}^i} \right) + \mathcal{O}(\delta q^2) \ . \tag{32.19}$$

Definition

The exercise above shows that the action is stationary at the classical trajectory $q^i_c(t)$ when

$$\frac{\partial L}{\partial q^i} = \frac{d}{dt} \frac{\partial L}{\partial \dot{q}^i} \ . \tag{32.20}$$

These are known as the **Euler-Lagrange** equations. In practice, we typically do not explicitly vary the action. We simply use these equations to find the equations which determine the evolution of the system we are interested in.

Example

Let us use the Euler-Lagrange equations on the simple Lagrangian defined in Eq. (32.5). To make explicit that this can be written as a function of x^i only (and not x_i) we express the Lagrangian as

$$L = \frac{m}{2} g_{ij} \dot{x}^i \dot{x}^j - U(x) \ , \tag{32.21}$$

where g_{ij} is the metric of three dimensional Euclidean space. We then find that

$$\frac{\partial L}{\partial x^k} = -\partial_k U(x) \ . \tag{32.22}$$

The choice of index should make the next steps a little clearer. Using the product rule we also find

$$\frac{\partial L}{\partial \dot{x}^k} = \frac{m}{2} g_{ij} \left(\delta^i_k \dot{x}^j + \dot{x}^i \delta^j_k \right) = m \dot{x}_k \ . \tag{32.23}$$

The Euler-Lagrange equations then yield

$$-\partial_i U(x) = \frac{d}{dt} m \, \dot{x}_k = m \ddot{x}_k \ . \tag{32.24}$$

This is (again) exactly Newton's 2nd Law.

Exercise 32.3 Work through all the steps of the above example. In particular, make sure that you can reproduce Eq. (32.23).

Example Consider a simple harmonic oscillator, such as a mass m attached to a spring with spring constant k. With x defined as the distance the spring has compressed or extended from equilibrium the usual equation of motion (from Newton's 2nd Law) is

$$m \, \ddot{x}^i = -k \, x^i \tag{32.25}$$

and leads to oscillations with angular frequency

$$\omega = \sqrt{\frac{k}{m}} \ . \tag{32.26}$$

This is simply a specific example of the previous form of the above equations with

$$U(x) = \frac{1}{2} k \, x^i x_i \ , \tag{32.27}$$

so the Lagrangian is

$$L = \frac{m}{2} \dot{x}^i \dot{x}_i - \frac{k}{2} x^i x_i \ . \tag{32.28}$$

However, we can write this in terms of a new corrdinate

$$x^i(t) = \frac{1}{\sqrt{m}} \overline{x}^i(t) \ , \tag{32.29}$$

and find the new Lagrangian

$$L = \frac{1}{2} \dot{\overline{x}}^i \, \dot{\overline{x}}_i - \frac{\omega^2}{2} \overline{x}^i \overline{x}_i \ . \tag{32.30}$$

This leads to exactly the same equations of motion.

Exercise 32.4 Show that the Lagrangian above produces the correct equation for simple harmonic motion with angular frequency ω.

Example Often complicated systems can be modeled as a combination of several simple oscillators. In this example we consider motion of masses in one dimension only. Suppose that between two walls we have a spring of spring constant k_1 attached to the left side of a mass m_1. A second spring with constant k_2 connect the right side of this mass to the left side of a second mass, m_2, which is then connected to the right wall by a spring with constant k_3.

We are interested in modeling the motion of the masses. We define two coordinates $x_1(t)$ and $x_2(t)$ that measure the deviations of m_1 and m_2 from their equilibrium positions, positive x taken to be motion to the right. The kinetic energy of the two masses is given by

$$T = \frac{m_1}{2}\dot{x}_1^2 + \frac{m_2}{2}\dot{x}_2^2 \ . \tag{32.31}$$

The potential energy is similarly the sum of terms encoding how much each spring is stretched or compressed

$$U = \frac{k_1}{2}x_1^2 + \frac{k_2}{2}(x_1 - x_2)^2 + \frac{k_3}{2}x_2^2 \ . \tag{32.32}$$

The only subtle point here is that spring two can be adjusted away from equilibrium by changing the location of either mass.

Exercise 32.5 Find the Lagrangian as given by

$$L = T - U \ , \tag{32.33}$$

and then find the equations of motion for each mass. Explain how these equations agree with characterizing the system by forces and Newton's 2nd Law.

Exercise 32.6 Show that the Lagrangian can be written in matrix form

$$L = \frac{1}{2}\begin{pmatrix}\dot{x}_1 & \dot{x}_2\end{pmatrix}\begin{pmatrix}m_1 & 0\\ 0 & m_2\end{pmatrix}\begin{pmatrix}\dot{x}_1\\ \dot{x}_2\end{pmatrix} - \frac{1}{2}\begin{pmatrix}x_1 & x_2\end{pmatrix}\begin{pmatrix}k_1+k_2 & -k_2\\ -k_2 & k_2+k_3\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix} \ . \tag{32.34}$$

Note that the mass matrix M and the spring constant matrix K do not commute

$$[M,\, K] \neq 0 \ . \tag{32.35}$$

This means that they cannot both be made diagonal in the same basis.

However, make the transformations

$$x_1 = \frac{1}{\sqrt{m_1}}\overline{x}_1 \ , \qquad x_2 = \frac{1}{\sqrt{m_2}}\overline{x}_2 \ , \tag{32.36}$$

to find

$$\begin{aligned} L =& \frac{1}{2}\left(\dot{\overline{x}}_1 \ \dot{\overline{x}}_2\right)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \dot{\overline{x}}_1 \\ \dot{\overline{x}}_2 \end{pmatrix} \\ & - \frac{1}{2}\left(\overline{x}_1 \ \overline{x}_2\right)\begin{pmatrix} \omega_1^2 + \omega_{2,1}^2 & -\omega_{2,1}\,\omega_{2,2} \\ -\omega_{2,1}\,\omega_{2,2} & \omega_{2,2}^2 + \omega_3^2 \end{pmatrix}\begin{pmatrix} \overline{x}_1 \\ \overline{x}_2 \end{pmatrix} \ , \end{aligned} \tag{32.37}$$

where

$$\omega_1 = \sqrt{\frac{k_1}{m_1}} \ , \quad \omega_3 = \sqrt{\frac{k_3}{m_2}} \ , \quad \omega_{2,1} = \sqrt{\frac{k_2}{m_1}} \ , \quad \omega_{2,2} = \sqrt{\frac{k_2}{m_2}} \ . \tag{32.38}$$

These two matrices must commute, since the identity matrix commutes with everything. This means that we can diagonalize the frequency matrix.

Example

Coupled Oscillators Continued: By going through the usual diagonalization procedure we find the transformation matrix

$$\begin{aligned} S =& \frac{1}{\sqrt{2}\sqrt{\omega_+^2 - \omega_-^2}\sqrt{\omega_+^2 - \omega_-^2 + \omega_1^2 - \omega_3^2 + \omega_{2,1}^2 - \omega_{2,2}^2}} \times \\ & \begin{pmatrix} \omega_+^2 - \omega_-^2 + \omega_1^2 - \omega_3^2 + \omega_{2,1}^2 - \omega_{2,2}^2 & 2\omega_{2,1}\omega_{2,2} \\ -2\omega_{2,1}\omega_{2,2} & \omega_+^2 - \omega_-^2 + \omega_1^2 - \omega_3^2 + \omega_{2,1}^2 - \omega_{2,2}^2 \end{pmatrix} \end{aligned} \tag{32.39}$$

where

$$\omega_\pm^2 = \frac{1}{2}\left(\omega_1^2 + \omega_3^2 + \omega_{2,1}^2 + \omega_{2,2}^2 \pm \sqrt{\left(\omega_1^2 - \omega_3^2 + \omega_{2,1}^2 - \omega_{2,2}^2\right)^2 + 4\omega_{2,1}^2\omega_{2,2}^2}\right) \ . \tag{32.40}$$

Note that S is orthogonal, that is $S^T = S^{-1}$. This means we can write

$$\begin{pmatrix} \overline{x}_1 \\ \overline{x}_2 \end{pmatrix} = SS^T\begin{pmatrix} \overline{x}_1 \\ \overline{x}_2 \end{pmatrix} \equiv S\begin{pmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{pmatrix} \ , \tag{32.41}$$

and therefore

$$L = \frac{1}{2}\left(\dot{\tilde{x}}_1 \ \dot{\tilde{x}}_2\right)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \dot{\tilde{x}}_1 \\ \dot{\tilde{x}}_2 \end{pmatrix} - \frac{1}{2}\left(\tilde{x}_1 \ \tilde{x}_2\right)\begin{pmatrix} \omega_+^2 & 0 \\ 0 & \omega_-^2 \end{pmatrix}\begin{pmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{pmatrix} \ . \tag{32.42}$$

The point of all these manipulations is that we began with a set of

coupled oscillators and have ended with a set of independent oscillators. To be sure, these oscillating modes have a complicated (though straightforward) relationship to the coordinates that describe the motion of the masses. But this illustrates the power of the Lagrangian formalism. It is flexible in what it can use as coordinates, allowing us to algebraically reach a point where the differential equations are simple to solve.

Subsection 32.1

Hamiltonian Dynamics

Another manifestation of the variational techniques uses a function called the Hamiltonian rather than the Lagrangian. We do not spend much time on this rich subject, only introduce a few essentials. The Hamiltonian formulation of classical physics is the language that most easily transitions into quantum mechanics.

Definition Given a Lagrangian $L(q^i, \dot{q}^i)$, the **conjugate momentum** p_i of the coordinate q^i is defined by

$$p_i = \frac{\partial L}{\partial \dot{q}^i} \ . \tag{32.43}$$

Note that while q^i is a vector its conjugate momentum p_i is a dual vector.

Example For a simple harmonic oscillator the Lagrangian is

$$L = \frac{m}{2}\dot{x}^i\dot{x}_i - \frac{k}{2}x^i x_i \ . \tag{32.44}$$

The momentum conjugate to x^i is

$$p_k = \frac{\partial L}{\partial \dot{x}^k} = m\,\dot{x}_k \ . \tag{32.45}$$

In this simple case momentum is just the mass times the velocity.

Definition Given a Lagrangian $L(q^i, \dot{q}^i)$, the **Hamiltonian** $H(q^i, p_i)$ is defined by

$$H = \dot{q}^i p_i - L \ . \tag{32.46}$$

The first term above does imply a sum over the repeated index. Because we are considering the Hamiltonian to be a function of q^i and p_i we must understand $\dot{q}^i$ to be a function of these variables.

Exercise 32.7 Show that the Hamiltonian for the simple harmonic oscillator is

$$H = \frac{1}{2m} p^i p_i + \frac{k}{2} x^i x_i \ . \tag{32.47}$$

We can now obtain Hamilton's equations from this definition and the Euler-Lagrange equations given in Eq. (32.20). In these derivations we take $\dot{q}^i(q^i, p_i)$ which requires using the chain rule. First,

$$\begin{aligned} \frac{\partial H}{\partial q^i} =& \frac{\partial \dot{q}^j}{\partial q^i} p_j - \frac{\partial L}{\partial q^i} - \frac{\partial \dot{q}^j}{\partial q^i} \frac{\partial L}{\partial \dot{q}^j} \\ =& \frac{\partial \dot{q}^j}{\partial q^i} p_j - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}^i} - \frac{\partial \dot{q}^j}{\partial q^i} p_j \\ =& - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}^i} = -\dot{p}_i \ . \end{aligned} \tag{32.48}$$

Next,

$$\begin{aligned} \frac{\partial H}{\partial p_i} =& \dot{q}^j \frac{\partial p_j}{\partial p_i} + p_j \frac{\partial \dot{q}^j}{\partial p_i} - \frac{\partial \dot{q}^j}{\partial p_i} \frac{\partial L}{\partial \dot{q}^j} \\ =& \dot{q}^j \delta^i_j + p_j \frac{\partial \dot{q}^j}{\partial p_i} - \frac{\partial \dot{q}^j}{\partial p_i} p_j \\ =& \dot{q}^i \ . \end{aligned} \tag{32.49}$$

Whereas the Euler-Lagrange equations are second order, Hamilton's equations are a coupled set of first order equations. Hamiltonians have all of the coordinate flexibility of Lagrangians and a little bit more. So-called **canonical transformations** can mix the q's and p's which can be exploited to further simplify some systems. Essentially, any transformation is allowed the preserves the form of the Hamiltonian equations above.

It can be shown that area preserving transformations are exactly what is needed. If new coordinates $Q(q, p)$ and $P(q, p)$ are to be used then one can construct the Jacobian matrix

$$J = \begin{pmatrix} \frac{\partial Q}{\partial q} & \frac{\partial Q}{\partial p} \\ \frac{\partial P}{\partial q} & \frac{\partial P}{\partial p} \end{pmatrix} \ . \tag{32.50}$$

To be area preserving, and hence canonical, we require

$$\mathrm{Det}(J) = 1 \ . \tag{32.51}$$

Example A useful example of using canonical transformations can be applied to the simple harmonic oscillator. We found above that the Hamiltonian is

given by

$$H = \frac{1}{2m}p^2 + \frac{k}{2}x^2 \ , \tag{32.52}$$

where we are keeping to one dimension for simplicity. Notice that we could rewrite this as

$$H = \frac{1}{2}\sqrt{\frac{k}{m}}\left(\frac{p^2}{\sqrt{km}} + x^2\sqrt{km}\right) \ . \tag{32.53}$$

This has a common factor out front and is then the sum of two squared quantities, kind of like the radius of polar coordinates in two dimensions

$$x = r\cos\theta \ , \qquad y = r\sin\theta \ . \tag{32.54}$$

If we can find the right canonical transformation, the hope is that we can write the Hamiltonian in terms of a single variable that is kind of like the radius in the space of x and p. The first transformation we might try is $(x, p) \to (r, \theta)$ where

$$r = \sqrt{\frac{p^2}{\sqrt{km}} + x^2\sqrt{km}} \ , \qquad \theta = \tan^{-1}\left(\frac{p/(km)^{1/4}}{x(km)^{1/4}}\right) \ . \tag{32.55}$$

Exercise 32.8 Show that the transformation given in (32.55) is not canonical. Then, show that the related transformation $(x, p) \to (N, \theta)$

$$N = \frac{1}{2}\left(\frac{p^2}{\sqrt{km}} + x^2\sqrt{km}\right) \ , \qquad \theta = \tan^{-1}\left(\frac{p}{x\sqrt{km}}\right) \ , \tag{32.56}$$

is canonical. Finally, show that the inverse of this transformation is

$$x = \frac{\sqrt{2N}}{(km)^{1/4}}\cos\theta \ , \qquad p = \sqrt{2N}(km)^{1/4}\sin\theta \ . \tag{32.57}$$

Example **Simple Harmonic Oscillator continued:** The above exercise confirms that (32.56) is a canonical transformation and hence is preserves the form of Hamilton's equations. It is easy to see that the Hamiltonian takes the form

$$H = \omega N \ , \tag{32.58}$$

where $\omega = \sqrt{k/m}$ as usual. This is clearly a simple Hamiltonian. The

resulting equations are

$$\dot{N} = \frac{\partial H}{\partial \theta} = 0 \ , \qquad \dot{\theta} = -\frac{\partial H}{\partial N} = -\omega \ . \tag{32.59}$$

This results show that N is conserved, which also show that the Hamiltonian (and hence the energy $E = \omega N$) is conserved. The second equation can be solved

$$\theta(t) = \theta_0 - \omega t \ . \tag{32.60}$$

This, in turn, implies that

$$x(t) = \frac{\sqrt{2N(t)}}{(km)^{1/4}} \cos\left[\theta(t)\right] = \sqrt{\frac{2E}{k}} \cos(\theta_0 - \omega t) \ . \tag{32.61}$$

We have found the standard motion for the simply harmonic oscillator. This may feel like more work than simply solving the initial differential equation, but the point was to show the method of canonical transformations. The form of the Hamiltonian also anticipates our quantum mechanical analysis of the simple harmonic oscillator in Sec. 33.4.

In whatever coordinates we choose, as long as the produce the standard Hamilton equations, we can also calculate the time derivative of the Hamiltonian

$$\begin{aligned} \frac{dH}{dt} =& \frac{\partial H}{\partial q^i}\frac{dq^i}{dt} + \frac{\partial H}{\partial p_i}\frac{dp_i}{dt} + \frac{\partial H}{\partial t} \\ =& (-\dot{p}_i)\dot{q}^i + \dot{q}^i\dot{p}_i + \frac{\partial H}{\partial t} \\ =& \frac{\partial H}{\partial t} \ . \end{aligned} \tag{32.62}$$

This means that if the Hamiltonian does not depend on time explicitly, meaning there are factors of t other than within q^i and p_i, then it is a conserved quantity. This is usually equal to the energy of the system.

Definition

There is one more structure related to the Hamiltonian formalism that anticipates the form of quantum mechanics. The **Poisson Bracket** of two functions $f(q,p)$ and $g(q,p)$ of the coordinates and their conjugate momenta is given by

$$\{f,\, g\}_{\rm P} = \frac{\partial f}{\partial q^i}\frac{\partial g}{\partial p_i} - \frac{\partial g}{\partial q^i}\frac{\partial f}{\partial p_i} \ . \tag{32.63}$$

Notice that the bracket is antisymmetric

$$\{f,\, g\}_{\rm P} = -\left\{g,\, f\right\}_{\rm P} \ . \tag{32.64}$$

Exercise 32.9 Use the definition of the Poisson bracket to show that

$$\left\{q^i,\, q^j\right\}_{\mathrm{P}} = 0 \ , \quad \{p_i,\, p_j\}_{\mathrm{P}} = 0 \ , \quad \left\{q^i,\, p_j\right\}_{\mathrm{P}} = \delta^i_j \ . \tag{32.65}$$

The Poisson bracket reveals some interesting structure of Hamiltonian mechanics. Note that for any function $f(q,p,t)$, its total time derivative is

$$\begin{aligned}\frac{df}{dt} =& \frac{\partial f}{\partial q^i}\frac{dq^i}{dt} + \frac{\partial f}{\partial p_i}\frac{dp_i}{dt} + \frac{\partial f}{\partial t} \\ =& \frac{\partial f}{\partial q^i}\frac{\partial H}{\partial p_i} + \frac{\partial f}{\partial p_i}\left(-\frac{\partial H}{\partial q^i}\right) + \frac{\partial f}{\partial t} \\ =& \{f,\, H\}_{\mathrm{P}} + \frac{\partial f}{\partial t} \ .\end{aligned} \tag{32.66}$$

This implies that any conserved quantity F must satisfy

$$\{H,\, F\}_{\mathrm{P}} = \frac{\partial F}{\partial t} \ . \tag{32.67}$$

In particular, if the conserved quantity does not depend on time explicitly, then its Poisson bracket with the Hamiltonian vanishes. On the other hand, for any quantity $f(q,p)$ that does not depend explicitly on time, its time evolution is encoded in its bracket with the Hamiltonian. The idea to remember is that there is a close connection between the Hamiltonian and time evolution.

SUBSECTION 32.2

Field Theories

So far we have discussed the trajectories of local objects, something like a particle or a planet. However, much of modern physics is more accurately described in the language of fields.

Definition In contrast to a localized object which can be tracked through space and time by a trajectory $x^i(t)$, a **field** $\phi(t, x^i)$ takes values throughout all space and time. In the local object case we specify a specific time t_0 and the trajectory tells us the location of the object in space $x^i(t_0)$. Mathematically this is a map from time to physical space, for example $\mathbb{R} \to \mathbb{R}^3$. In the case of the field both a time and a location in space need to be specified t_0 and x^i_0, then the value of the field is given $\phi(t_0, x^i_0)$. This a mapping from time and space to numbers, for example $\mathbb{R}^4 \to \mathbb{R}$.

Example The electric potential $\phi(t, x^i)$ is an example of a **scalar field**. If we are looking at the potential due to a fixed electric charge, then the field

does not depend on time. However, the further away from the charge we measure the smaller value of the potential we find. At any point in space there would be some specific value.

Example The electric field $E^i(t, x^i)$ is an example of a **vector field**. In this case, once a time and location in space are specified the field does not return just one number, but three, a vector of numbers. Again, considering the case of a single charged particle, the field would not change in time, but at any point in space one can measure the three components of the electric field.

When using the language of fields we do not use the Lagrangian L, but the Lagrangian density $\mathcal{L}$ which are related by

$$L = \int d^3x \, \mathcal{L} \ . \tag{32.68}$$

As its name might suggest, the Lagrangian density is something like the Lagrangian per unit volume. This means that the action is

$$S = \int dt \, d^3x \, \mathcal{L} \equiv \int d^4x \, \mathcal{L} \ . \tag{32.69}$$

The Lagrangian density is a function of fields and their derivatives in both time and space

$$\mathcal{L}(\phi, \dot{\phi}, \partial_i \phi) \ . \tag{32.70}$$

We can find the correct Euler-Lagrange equations by considering

$$\phi = \phi_c + \delta\phi \ , \tag{32.71}$$

where ϕ_c is the classical field configuration and $\delta\phi$ is a small deviation. We choose this deviation to vanish at the initial and final times (which may be infinite) as well as the boundaries in space (which may be at infinity)

$$\delta\phi(t_{f,i}, x^i) = 0 \ , \quad \delta\phi(t, \pm\infty) = 0 \ . \tag{32.72}$$

The action can be expanded to first order in $\delta\phi$ to find

$$S = \int d^4x \left[\mathcal{L}(\phi_c, \dot{\phi}_c, \partial_i \phi_c) + \left.\frac{\partial \mathcal{L}}{\partial \phi}\right|_{\phi=\phi_c} \delta\phi + \left.\frac{\partial \mathcal{L}}{\partial \dot{\phi}}\right|_{\phi=\phi_c} \dot{\delta\phi} + \left.\frac{\partial \mathcal{L}}{\partial(\partial_i \phi)}\right|_{\phi=\phi_c} \partial_i \delta\phi \right] . \tag{32.73}$$

The classical action is simply

$$S_c = \int d^4x \ \mathcal{L}(\phi_c, \dot{\phi}_c, \partial_i \phi_c) \ , \tag{32.74}$$

and we can integrate by parts in time to find

$$\int_{t_i}^{t_f} dt\, \frac{\partial \mathcal{L}}{\partial \dot\phi}\bigg|_{\phi=\phi_c} \dot{\delta\phi} = \frac{\partial \mathcal{L}}{\partial \dot\phi}\bigg|_{\phi=\phi_c} \delta\phi \bigg|_{t_i}^{t_f} - \int_{t_i}^{t_f} dt\, \delta\phi\, \partial_t \frac{\partial \mathcal{L}}{\partial \dot\phi}\bigg|_{\phi=\phi_c} \ , \tag{32.75}$$

where the boundary term vanishes. The spatial integration by parts is essentially the same, but has a slightly more complicated form

$$\int d^3x\ \frac{\partial \mathcal{L}}{\partial(\partial_i\phi)}\bigg|_{\phi=\phi_c} \partial_i\delta\phi = \int_\infty d^2\sigma_i\ \frac{\partial \mathcal{L}}{\partial(\partial_i\phi)}\bigg|_{\phi=\phi_c} \delta\phi - \int d^3x\, \delta\phi\ \partial_i \frac{\partial \mathcal{L}}{\partial(\partial_i\phi)}\bigg|_{\phi=\phi_c} \ , \tag{32.76}$$

where the boundary term is an integral over the two dimensional sphere at spatial infinity. The integration measure is a vector that points orthogonally outward from the surface of the sphere. This integral is zero because we have chosen $\delta\phi$ to vanish at spatial infinity. Thus, we find

$$S = S_c + \int d^4x\, \delta\phi \left(\frac{\partial \mathcal{L}}{\partial \phi}\bigg|_{\phi=\phi_c} - \partial_t \frac{\partial \mathcal{L}}{\partial \dot\phi}\bigg|_{\phi=\phi_c} - \partial_i \frac{\partial \mathcal{L}}{\partial(\partial_i\phi)}\bigg|_{\phi=\phi_c} \right) \ . \tag{32.77}$$

If we require the action to be stationary for any $\delta\phi$ then it must be that

$$\frac{\partial \mathcal{L}}{\partial \phi} = \partial_t \frac{\partial \mathcal{L}}{\partial \dot\phi} + \partial_i \frac{\partial \mathcal{L}}{\partial(\partial_i\phi)} \ , \tag{32.78}$$

which are the Euler-Lagrange equations. The conjugate momentum fields are defined to be

$$\pi(t, x^i) = \frac{\partial \mathcal{L}}{\partial \dot\phi} \ , \tag{32.79}$$

and the Hamiltonian density is then

$$\mathcal{H} = \pi\dot\phi - \mathcal{L} \ . \tag{32.80}$$

Example

A simple example in one spatial dimension is the Lagrangian density

$$\mathcal{L} = \frac{1}{2}\dot u^2 - \frac{c}{2}\left(\partial_x u\right)^2 \ , \tag{32.81}$$

where $u(t, x)$ is a function of time and one spatial dimension. The Euler-Lagrange equations given in (32.78) lead to

$$0 = \partial_t \dot u - c\, \partial_x \partial_x u \tag{32.82}$$

or

$$\ddot u - c\, \partial_x^2 u = 0 \ . \tag{32.83}$$

This is simple the wave equation for a wave of speed c.

The conjugate momentum of u is

$$\pi = \frac{\partial \mathcal{L}}{\partial \dot{u}} = \dot{u} \ , \tag{32.84}$$

so the Hamiltonian density is

$$\mathcal{H} = \pi^2 - \frac{1}{2}\pi^2 + \frac{c}{2}\left(\partial_x u\right)^2 = \frac{1}{2}\pi^2 + \frac{c}{2}\left(\partial_x u\right)^2 \ . \tag{32.85}$$

This is associated with the energy density of the field u. Note that there are two positive contributions: one from the field changing in time and one from changes in u along the x direction. This means that a crinkled field has more energy stored in it than a uniform field.

Exercise 32.10 Show that the Lagrangian density of the field $\phi(t,\, x^i)$ in three dimensions

$$\mathcal{L} = \frac{1}{2}\dot{\phi}^2 - \frac{c}{2}\left(\partial_i \phi\right)\partial^i \phi \ , \tag{32.86}$$

leads to the three dimensional wave equation

$$\ddot{\phi} - c\,\partial_i \partial^i \phi = 0 \ . \tag{32.87}$$

Find the Hamiltonian density and comment on the various contributions to the energy of a given field configuration.

SUBSECTION 32.3

Symmetries and Noether's Theorem

One reason that the Lagrangian formalism is so powerful was elucidated by Emmy Noether. We have seen that the flexibility under coordinate transformations (or field transformations) allows us to re-express the Lagrangian in ways that may be simpler to solve. What if we find a class of nontrivial transformations that leave the form of the Lagrangian unchanged?

Example Suppose we have a Lagrangian density with two fields ϕ_1 and ϕ_2 with

$$\mathcal{L} = \frac{1}{2}\dot{\phi}_1\dot{\phi}_1 - \frac{c}{2}\left(\partial_i \phi_1\right)\partial^i \phi_1 + \frac{1}{2}\dot{\phi}_2\dot{\phi}_2 - \frac{c}{2}\left(\partial_i \phi_2\right)\partial^i \phi_2 \ . \tag{32.88}$$

We then transform to new fields ϕ_1' and ϕ_2' according to

$$\phi_1 = \cos\theta\phi_1' - \sin\theta\phi_2' \ , \qquad \phi_2 = \sin\theta\phi_1' + \cos\theta\phi_2' \ . \tag{32.89}$$

We find that the Lagrangian now takes the form

$$\mathcal{L} = \frac{1}{2}\dot{\phi}_1'\dot{\phi}_1' - \frac{c}{2}\left(\partial_i\phi_1'\right)\partial^i\phi_1' + \frac{1}{2}\dot{\phi}_2'\dot{\phi}_2' - \frac{c}{2}\left(\partial_i\phi_2'\right)\partial^i\phi_2' \ . \tag{32.90}$$

We see that the functional form of the two equations is identical. For any real number θ the two Lagrangians lead to exactly the same equations of motion. We say that this Lagrangian has a **continuous symmetry** because θ can take a continuum of values.

Exercise 32.11 Fill in the steps that lead to the Lagrangian in Eq. (32.90) in the previous example.

In general, suppose that we have a Lagrangian density $\mathcal{L}(\phi, \dot{\phi}, \partial_i\phi)$ and a transformation to a new set of fields

$$\phi'(t, x^i, \theta) \ , \tag{32.91}$$

such that $\phi'(t, x^i, 0) = \phi(t, x^i)$. That is, the transformation from ϕ to ϕ' depends on a continuous parameter θ such that when $\theta = 0$ the fields are identical. If this transformation is a symmetry of the Lagrangian, then we have

$$\mathcal{L}(\phi, \dot{\phi}, \partial_i\phi) = \mathcal{L}(\phi', \dot{\phi}', \partial_i\phi') \ , \tag{32.92}$$

for all θ. If we expand in a series expansion of θ then we would find coefficients for every power. Because the symmetry holds for all θ we can vary it arbitrarily and the symmetry still has to hold true. This implies that coefficient of each power of θ must preserve the symmetry. In particular, we can consider small values of theta, those with $\theta \ll 1$, to find constraints from the θ^1 term. To leading order in θ this means

$$\begin{aligned}\phi'(\theta) =& \phi'(0) + \theta\frac{d\phi'}{d\theta}\bigg|_{\theta=0} + \mathcal{O}(\theta^2) \\ =& \phi + \theta\frac{d\phi'}{d\theta}\bigg|_{\theta=0} + \mathcal{O}(\theta^2) \ ,\end{aligned} \tag{32.93}$$

where we have suppressed the field dependence on space and time for brevity. Of course, if our continuous symmetry is invertible, and near to $\theta = 0$ it certainly is, then we can just as correctly write

$$\begin{aligned}\phi(\theta) =& \phi(0) + \theta\frac{d\phi}{d\theta}\bigg|_{\theta=0} + \mathcal{O}(\theta^2) \\ =& \phi' + \theta\frac{d\phi}{d\theta}\bigg|_{\theta=0} + \mathcal{O}(\theta^2) \ .\end{aligned} \tag{32.94}$$

The expansions in Eqs. (32.93) and (32.94) must agree at linear order in θ.

This implies that

$$\left.\frac{d\phi'}{d\theta}\right|_{\theta=0} = -\left.\frac{d\phi}{d\theta}\right|_{\theta=0} . \tag{32.95}$$

We can then expand the Lagrangian to linear order in θ

$$\begin{aligned}\mathcal{L}(\phi,\dot{\phi},\partial_i\phi) =& \mathcal{L}(\phi',\dot{\phi}',\partial_i\phi') \\ &+ \theta\left.\frac{d\phi}{d\theta}\right|_{\theta=0}\frac{\partial\mathcal{L}}{\partial\phi} + \theta\partial_t\left.\frac{d\phi}{d\theta}\right|_{\theta=0}\frac{\partial\mathcal{L}}{\partial\dot{\phi}} + \theta\partial_i\left.\frac{d\phi}{d\theta}\right|_{\theta=0}\frac{\partial\mathcal{L}}{\partial(\partial_i\phi)} .\end{aligned} \tag{32.96}$$

But, from our assumption that the field transformation is a symmetry we have Eq. (32.92). This means that

$$\left.\frac{d\phi}{d\theta}\right|_{\theta=0}\frac{\partial\mathcal{L}}{\partial\phi} + \partial_t\left.\frac{d\phi}{d\theta}\right|_{\theta=0}\frac{\partial\mathcal{L}}{\partial\dot{\phi}} + \partial_i\left.\frac{d\phi}{d\theta}\right|_{\theta=0}\frac{\partial\mathcal{L}}{\partial(\partial_i\phi)} = 0 . \tag{32.97}$$

Although it may not look like much, this result has some important consequences. First, we define the following quantities

$$\rho = \left.\frac{d\phi}{d\theta}\right|_{\theta=0}\frac{\partial\mathcal{L}}{\partial\dot{\phi}} , \qquad J^i = \left.\frac{d\phi}{d\theta}\right|_{\theta=0}\frac{\partial\mathcal{L}}{\partial(\partial_i\phi)} . \tag{32.98}$$

Note that, similar to the continuity equation you may have run across in electrodynamics, we find

$$\begin{aligned}\partial_t\rho + \partial_i J^i =& \partial_t\left.\frac{d\phi}{d\theta}\right|_{\theta=0}\frac{\partial\mathcal{L}}{\partial\dot{\phi}} + \left.\frac{d\phi}{d\theta}\right|_{\theta=0}\partial_t\frac{\partial\mathcal{L}}{\partial\dot{\phi}} \\ &+ \partial_i\left.\frac{d\phi}{d\theta}\right|_{\theta=0}\frac{\partial\mathcal{L}}{\partial(\partial_i\phi)} + \left.\frac{d\phi}{d\theta}\right|_{\theta=0}\partial_i\frac{\partial\mathcal{L}}{\partial(\partial_i\phi)} .\end{aligned} \tag{32.99}$$

Then, using Eq. (32.97) we find

$$\begin{aligned}\partial_t\rho + \partial_i J^i =& \left.\frac{d\phi}{d\theta}\right|_{\theta=0}\left(-\frac{\partial\mathcal{L}}{\partial\phi} + \partial_t\frac{\partial\mathcal{L}}{\partial\dot{\phi}} + \partial_i\frac{\partial\mathcal{L}}{\partial(\partial_i\phi)}\right) \\ =& 0 ,\end{aligned} \tag{32.100}$$

where we have used the Euler-Lagrange equations (32.78). In four-vector notation

Let's pause for a moment to figure out what this means. We have shown that if there is a continuous set of transformations that leave the functional form of the Lagrangian unchanged (which we refer to as a symmetry of the Lagrangian), there are quantities analogous to a charge density and vector current density that satisfy the continuity equation as above. We define

the total charge Q by integrating over the charge density

$$Q = \int d^3x\, \rho \ . \tag{32.101}$$

The essential characteristic of this charge is that it is **conserved**. That is,

$$\frac{dQ}{dt} = \int d^3x\, \partial_t \rho = -\int d^3x\, \partial_i J^i = \int_\infty d^2\sigma_i J^i = 0 \ , \tag{32.102}$$

where we have used the continuity equation in the second equality and Gauss' Law in the third. The final equality holds under the assumption that the fields go to zero at infinity.

Example

We return to the Lagrangian density given in Eq. (32.88). It is straightforward to show that

$$\phi_1 = \cos\theta\phi_1' - \sin\theta\phi_2' \ , \qquad \phi_2 = \sin\theta\phi_1' + \cos\theta\phi_2' \ , \tag{32.103}$$

so we find

$$\left.\frac{d\phi_1}{d\theta}\right|_{\theta=0} = -\phi_2 \ , \qquad \left.\frac{d\phi_2}{d\theta}\right|_{\theta=0} = \phi_1 \ . \tag{32.104}$$

Therefore, we have

$$\begin{aligned}\rho =& \left.\frac{d\phi_1}{d\theta}\right|_{\theta=0} \frac{\partial\mathcal{L}}{\partial\dot{\phi}_1} + \left.\frac{d\phi_2}{d\theta}\right|_{\theta=0} \frac{\partial\mathcal{L}}{\partial\dot{\phi}_2} \\ =& \phi_2\dot{\phi}_1 - \phi_1\dot{\phi}_2 \ ,\end{aligned} \tag{32.105}$$

and

$$\begin{aligned}J^i =& \left.\frac{d\phi_1}{d\theta}\right|_{\theta=0} \frac{\partial\mathcal{L}}{\partial(\partial_i\phi_1)} + \left.\frac{d\phi_2}{d\theta}\right|_{\theta=0} \frac{\partial\mathcal{L}}{\partial(\partial_i\phi_2)} \\ =& \phi_2\partial^i\phi_1 - \phi_1\partial^i\phi_2 \ .\end{aligned} \tag{32.106}$$

Though it is not obvious from these results, the conserved charge determined here is like physical charges such as baryon number or lepton number. This connection is made clearer later in the text.

As we show later, electric charge is also related to a symmetry of the Lagrangian density that describes quantum electrodynamics. Other conserved quantities, like baryon number and lepton number are also related to symmetries. Even when the equations of motion that result from a Lagrangian are difficult (seemingly impossible) to solve exactly, determining the symmetries of that system lead to nontrivial predictions about the resulting phenomena that can be checked experimentally.

To recap, we have shown (following Emmy Noether) that symmetries of the Lagrangian correspond to conserved quantities in the description of

the physical system. This a deep result and very powerful. Although we do not take the time to show it here, this result allows us to connect familiar conserved quantities to assumptions about physical laws. For instance, if we assume that the laws of physics are independent of time, meaning that no matter when a process occurs it will always occur according to the same rules, then there is a conserved quantity, which we call energy. If physics is the same everywhere along a direction in physical space we find a conserved quantity, called momentum, along that direction in space. If it does not matter how objects are oriented, that is, if system can be rotated in space without changing the dynamics and results, then we find there is a conserved quantity, called angular momentum.

What about the boost symmetry of special relativity? As shown in [15], the conservation law that follows from this symmetry is the steady motion of the center-of-energy. In essence, this is Newton's 1st Law, the law of inertia. You may remember that the conservation of momentum is equivalent to Newton's 3rd Law, that of equal and opposite reaction. This means that Hamilton's principle of stationary action, when combined with Noether's theorem, produces all three of Newton's laws, including their rotational forms, as well as the conservation of energy.

In fact, all of these spacetime symmetries can be neatly packaged in terms of some a quantity called the stress-energy or energy-momentum tensor $T_{\mu\nu}$, which is typically defined by

$$T^{\mu}{}_{\nu} = \sum_{\phi} \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \partial_\nu \phi - \delta^{\mu}_{\nu} \mathcal{L} \ , \tag{32.107}$$

where the sum is over all fields that the Lagrangian density depends on. One then finds that the energy is given by

$$E = \int d^3x \, T^{00} \ , \tag{32.108}$$

and the conserved linear momenta are given by

$$P^i = \int d^3x \, T^{0i} \ . \tag{32.109}$$

Section 33

Quantum Essentials

Quantum mechanics is a formalism for describing nature that is amazingly consistent with experiment. While some of the predictions do not agree with our day-to-day intuition about the natural world, those who study nature on the smallest scales find agreement with these predictions time and again. In this text we simply assume that nature is quantum mechanical.

This introduction is only a brief one. There are many good undergraduate texts that provide a more complete treatment, including [83–87]. More advanced, but still quite readable, texts include [88–90].

The mathematical language for describing the state of a physics system is taken to be a complex inner product space, more particularly a Hilbert space, but the distinctions are not essential to this discussion. This inner-product space is often infinite dimensional, so matrix representations can be cumbersome. Vectors are often denoted by

$$|\psi\rangle \ , \tag{33.1}$$

and are some times called Kets, being the latter part of a bracket. Dual vectors are represented by

$$|\psi\rangle^\dagger = \langle\psi| \ , \tag{33.2}$$

and are known as Bras, the first part of a bracket. The inner product between two vectors $|\psi\rangle$ and $|\phi\rangle$ is given by

$$\langle\psi|\phi\rangle = \langle\phi|\psi\rangle^* \ , \tag{33.3}$$

and is a complex number. However, the inner product of any vector with itself satisfies

$$\langle\psi|\psi\rangle = \langle\psi|\psi\rangle^* \ , \tag{33.4}$$

and so must be a real number.

Complex numbers are useful mathematical objects but they cannot be physical observables like the energy or momentum of a state. Consequently, the inner product of two vectors cannot be directly interpreted as a physical quantity. However, a Hermitian operator H, $(H^\dagger = H)$ always has real eigenvalues and these can be associated with the results of physical measurements. [2]

To see that Hermitian operators have real eigenvalues, consider an eigenvector $|\lambda\rangle$ of H with eigenvalue λ

$$H|\lambda\rangle = \lambda|\lambda\rangle \ . \tag{33.5}$$

Take the inner product of this equation with $\langle\lambda|$ to find

$$\langle\lambda|H|\lambda\rangle = \lambda\langle\lambda|\lambda\rangle \ . \tag{33.6}$$

On the other hand, the Hermitian conjugate of Eq. (33.5) produces

$$\langle\lambda|H^\dagger = \langle\lambda|H = \lambda^*\langle\lambda| \ . \tag{33.7}$$

[2] In many quantum mechanics textbooks operators are distinguished by a hat, such as $\widehat{H}$. We do not use this notation in this text because it is typical in quantum field theory texts not to put hats on operators.

Taking the inner product of this equation with $|\lambda\rangle$ leads to

$$\langle\lambda|H|\lambda\rangle = \lambda^*\langle\lambda|\lambda\rangle \ . \tag{33.8}$$

The only way both (33.6) and (33.8) can be true is if λ is a real number.

The eigenvectors of H with distinct eigenvalues are also orthogonal. Consider two eigenvectors of H labeled by their eigenvalues $|\lambda_1\rangle$ and $|\lambda_2\rangle$. Then consider that

$$\langle\lambda_1|H|\lambda_2\rangle = \lambda_2\langle\lambda_1|\lambda_2\rangle \tag{33.9}$$

$$= \lambda_1\langle\lambda_1|\lambda_2\rangle \ , \tag{33.10}$$

where in the top line we consider H acting on the vector to the right and in the bottom line we consider H acting on the dual vector to the left. We assumed at the outset that $\lambda_1 \neq \lambda_2$. in this case, the above equalities can only be satisfied if

$$\langle\lambda_1|\lambda_2\rangle \ = 0 \ . \tag{33.11}$$

Thus, if two eigenvectors of H have unequal eigenvalues then the two vectors are orthogonal.

The upshot of these characteristics of Hermitian operators is that the eigenvectors of such an operator can be used as an orthonormal basis for at least a part of the inner-product space. This means that any vector $|\psi\rangle$ that is taken to describe a physical system can be expressed as a linear combination of the eigenvectors of H

$$|\psi\rangle = \sum_i c_i|\lambda_i\rangle \ . \tag{33.12}$$

Exercise 33.1 Find a formula for determining the c_i coefficients by taking the inner product of Eq. (33.12) with $\langle\lambda_j|$.

Then the inner product is

$$\begin{aligned}\langle\psi|H|\psi\rangle &= \sum_j\sum_i \lambda_i c_j^* c_i\langle\lambda_j|\lambda_i\rangle \\ &= \sum_i \lambda_i|c_i|^2 \ .\end{aligned} \tag{33.13}$$

Exercise 33.2 Supply the intermediate steps that lead to Eq. (33.13).

The interpretation of this result, initially put forward by Max Born, is that if $|\psi\rangle$ describes some physical system and H corresponds to some observable quantity, energy for example, then when the energy of the state is measured any of the values λ_i are possible, but no values other than the λ_i. The *probability* of measuring the state to have energy λ_i is equal to

$|c_i|^2$. Because the sum of all probabilities must equal one, we find that

$$\sum_i |c_i|^2 = 1. \tag{33.14}$$

Exercise 33.3 Show that the above interpretation requires

$$\langle\psi|\psi\rangle = 1 \ . \tag{33.15}$$

This implies that we do not use the full inner-product space to describe physical systems. Only normalized states are employed.

With this understanding in mind we can make a simple statement about measurement. Suppose a physical system is described by a state $|\psi\rangle$. Suppose further that we are interested in making a measurement of a quantity that is associated with an operator O, which has eigenstates $|\lambda\rangle$ labeled by the eigenvalues λ. The probability of measuring the state $|\psi\rangle$ and finding λ is

$$|\langle\lambda|\psi\rangle|^2 \ . \tag{33.16}$$

Perhaps the oddest aspect of quantum mechanics is that measurement can drastically change the state. If the state is measured and returns λ then the state is changed into the state $|\lambda\rangle$, assuming there is only one state with eigenvalue λ. If there is more than one such state, meaning there is a degeneracy in the eigenvalue spectrum, then after a measurement of λ the original state is changed into a linear combination of the states with eigenvalue λ.

Because the probabilities are given by the $|c_i|^2$, two states that are related by a simple rephasing

$$|\phi\rangle = e^{i\theta}|\psi\rangle \ , \tag{33.17}$$

lead to the same probabilities. Therefore, we find that a given physical state can be described just as well by either vector. This leads us to say that we can always **rephase** a state without changing any of its physical qualities. Note, however, that the relative phase between two parts of a state is physical. It is only a phase that is applied to the whole state at once is not determined.

Exercise 33.4 Consider a simple quantum state

$$|\psi\rangle = r_1 e^{-i\theta_1}|\psi_1\rangle + r_2 e^{-i\theta_2}|\psi_2\rangle \ , \tag{33.18}$$

where $|\psi_1\rangle$ and $|\psi_2\rangle$ are *not* orthogonal. Show that $\langle\psi|\psi\rangle$ depends only on $\theta_1 - \theta_2$.

Subsection 33.1

Position and Momentum Operators

The Hermitian operators that correspond to physical observables may also reveal important structure. In particular, the operators that represent two different observables may not commute. If they do not commute then we cannot find a basis in which both operators are diagonal. The prime example the operators $\widehat{X}^i$ and $\widehat{P}_i$ what correspond to position and momentum, respectively. That is, a state of definite position in three dimensional space satisfies

$$\widehat{X}^i|x\rangle = x^i|x\rangle \ . \tag{33.19}$$

Here, $|x\rangle$ is a shorthand for $|x^{(1)}, x^{(2)}, x^{(3)}\rangle$, a vector in the inner product space that is labeled by three components of position. A similarly defined state of definite spatial momentum satisfies

$$\widehat{P}_i|p\rangle = p_i|p\rangle \ . \tag{33.20}$$

Note that $|x\rangle$ is a vector in the inner-product space and hence carries no index, while x^i is a vector in physical space.

Recall from Eq. (32.65) that in the Hamiltonian language the coordinates and momenta have a nonvanishing Poisson bracket. In analogy to this, we take the commutator of the position and momentum operators to be

$$\left[\widehat{X}^i,\, \widehat{P}_j\right] = i\hbar\, \delta^i_j \widehat{\mathbb{I}} \ , \tag{33.21}$$

where $\widehat{\mathbb{I}}$ is the identity operator on the Hilbert space. (Identity operators such as this are often dropped for brevity.) By making this assumption we can find the form of $\widehat{X}^i$ in the momentum basis and vice versa. First, we note that the normalization condition for a continuous basis set like position or momentum is, for example,

$$\langle x|x'\rangle = \delta^3(x - x') \ , \tag{33.22}$$

where the delta function's prime characteristic is that

$$\int d^3x\, \delta^3(x - x')\, f(x) = f(x') \ . \tag{33.23}$$

There are many ways to define the delta function, a useful example for a one dimensional delta function is

$$\delta(x - x') = \frac{1}{2\pi} \int dp\, e^{ip(x-x')} \ . \tag{33.24}$$

Exercise 33.5 Using the definition of a delta function in one dimension, show that

$$\delta^3(x - x') = \frac{1}{(2\pi)^3} \int d^3p \, e^{ip_i(x^i - x'^i)} \ . \tag{33.25}$$

Formally we have that

$$\delta(x) = \begin{cases} \infty & x = 0 \\ 0 & x \neq 0 \end{cases} \ . \tag{33.26}$$

This means that states of definite position (or momentum) cannot be physical because they do not have a norm of one. They have infinite norm

$$\langle x|x\rangle = \infty \ . \tag{33.27}$$

Nevertheless, these states are useful as a basis that we can use to describe physical states.

Beginning from Eq. (33.21) we find

$$\begin{aligned} \langle x| \left[\widehat{X}^i, \, \widehat{P}_j\right] |x'\rangle &= i\hbar \, \delta^i_j \, \langle x|x'\rangle \\ \langle x| \left(\widehat{X}^i \widehat{P}_j - \widehat{P}_j \widehat{X}^i\right) |x'\rangle &= i\hbar \, \delta^i_j \, \delta^3(x - x') \\ \langle x| \left(x^i \widehat{P}_j - \widehat{P}_j x'^i\right) |x'\rangle &= i\hbar \, \delta^i_j \, \delta^3(x - x') \\ (x^i - x'^i) \langle x|\widehat{P}_j|x'\rangle &= i\hbar \, \delta^i_j \, \delta^3(x - x') \ . \end{aligned} \tag{33.28}$$

To understand this result we need to review some properties of delta functions. In particular, we review how to understand the derivative of a delta function. The essential observation is that because the delta function acts extremely locally we can integrate by parts

$$\int d^3x \, \partial_j \delta^3(x - x') f(x) = - \int d^3x \, \delta^3(x^i - x'^i) \partial_j f(x) = \partial_j f(x') \ . \tag{33.29}$$

This implies that

$$\begin{aligned} &\int d^3x \, \partial_j \delta^3(x^i - x'^i)(x^i - x'^i) f(x) \\ &= - \int d^3x \, \delta^3(x - x') \left[\delta^i_j f(x) + (x^i - x'^i) \partial_j f(x)\right] \\ &= - \delta^i_j \int d^3x \, \delta^3(x - x') f(x) \ . \end{aligned} \tag{33.30}$$

In short, we find that

$$(x^i - x'^i) \partial_j \delta^3(x - x') = -\delta^i_j \, \delta^3(x - x') \tag{33.31}$$

Exercise 33.6 Work through the steps above.

When compared with Eq. (33.28), this result implies that

$$\langle x|\widehat{P}_j|x'\rangle = -i\hbar\,\partial_j\delta^3(x-x')\ . \tag{33.32}$$

Exercise 33.7 Show that

$$\langle x|\widehat{P}_j|x'\rangle = i\hbar\,\partial'_j\delta^3(x-x')\ , \tag{33.33}$$

where

$$\frac{\partial}{\partial x'^j} \equiv \partial'_j\ . \tag{33.34}$$

Exercise 33.8 Using the normalization of the momentum basis

$$\langle p|p'\rangle = \delta^3(p-p')\ , \tag{33.35}$$

show that

$$\langle p|\widehat{X^j}|p'\rangle = i\hbar\,\frac{\partial}{\partial p_j}\delta^3(p-p') = -i\hbar\,\frac{\partial}{\partial p'_j}\delta^3(p-p')\ . \tag{33.36}$$

We can also determine the overlap between position and momentum eigenstates. Before we do we make one additional observation from linear algebra. Given a complete orthonormal basis of states $|\lambda_i\rangle$ we can write the identity operator as

$$\widehat{\mathbb{I}} = \sum_i |\lambda_i\rangle\langle\lambda_i|\ . \tag{33.37}$$

Exercise 33.9 Given an arbitrary vector

$$|\psi\rangle = \sum_i c_i|\lambda_i\rangle\ , \tag{33.38}$$

show that Eq. (33.37) is the identity operator by showing that

$$\widehat{\mathbb{I}}|\psi\rangle = |\psi\rangle\ . \tag{33.39}$$

If the basis is continuous, as with the position basis, then the sum becomes an integral

$$\widehat{\mathbb{I}} = \int dx|x\rangle\langle x|\ . \tag{33.40}$$

Exercise 33.10 Given an arbitrary vector

$$|\psi\rangle = \int dx' c(x')|x'\rangle\ , \tag{33.41}$$

show that Eq. (33.40) is the identity operator by showing that

$$\widehat{\mathbb{I}}|\psi\rangle = |\psi\rangle \ . \tag{33.42}$$

We are now ready to determine the overlap between position states $|x\rangle$ and momentum states $|p\rangle$. We begin with the commutation relation given in Eq. (33.21) and write

$$\begin{aligned}
\langle x|\widehat{X^i}\widehat{P}_j - \widehat{P}_j\widehat{X^i}|p\rangle =& i\hbar\,\delta^i_i\,\langle x|p\rangle \\
x^i p_j\langle x|p\rangle - \langle x|\widehat{P}_j\widehat{X^i}|p\rangle =& i\hbar\,\delta^i_i\,\langle x|p\rangle \\
\langle x|p\rangle\left(x^i p_j - i\hbar\,\delta^i_j\right) =& \langle x|\widehat{P}_j\widehat{X^i}|p\rangle \ .
\end{aligned} \tag{33.43}$$

We focus on the right-hand side. By inserting the identity operator we find

$$\begin{aligned}
\langle x|\widehat{P}_j\widehat{X^i}|p\rangle =& \langle x|\widehat{P}_j\widehat{\mathbb{I}}\widehat{X^i}|p\rangle \\
=& \int dx'\langle x|\widehat{P}_j|x'\rangle\langle x'|\widehat{X^i}|p\rangle \\
=& i\hbar\int dx'\partial'_j\delta^3(x-x')\langle x'|x'^i|p\rangle \\
=& -i\hbar\int dx'\delta^3(x-x')\partial'_j\left(x'^i\langle x'|p\rangle\right) \\
=& -i\hbar\left(\delta^i_j\langle x|p\rangle + x^i\partial_j\langle x|p\rangle\right) \ .
\end{aligned} \tag{33.44}$$

We can then insert the result in Eq. (33.44) into Eq. (33.43) to find

$$\langle x|p\rangle\left(x^i p_j - i\hbar\,\delta^i_j\right) = -i\hbar\left(\delta^i_j\langle x|p\rangle + x^i\partial_j\langle x|p\rangle\right) \ . \tag{33.45}$$

This is simplifies to

$$x^i\partial_j\langle x|p\rangle = \frac{i}{\hbar}x^i p_j\langle x|p\rangle \ . \tag{33.46}$$

Exercise 33.11 Show that this result implies that

$$\langle x|p\rangle = c\,e^{\frac{i}{\hbar}x^i p_i} \ , \tag{33.47}$$

where we can use rephasing to ensure c is a real number. Then, using the fact that

$$\delta^3(x-x') = \langle x|x'\rangle = \int d^3p\langle x|p\rangle\langle p|x'\rangle \ , \tag{33.48}$$

show that

$$c = \frac{1}{(2\pi\hbar)^{3/2}} \ . \tag{33.49}$$

Subsection 33.2
Quantum Symmetries

This requirement that physical states are normalized has some important consequences. First, it constrains how symmetries are represented in the inner-product space. We typically think of symmetries as being related to some mathematical group G, see Sec. 36 for a review of what groups are. The elements of G must be represented in some way in the inner-product space, which we may write

$$\widehat{R}(g) \ . \tag{33.50}$$

Then the action of a symmetry on a physical state $|\psi\rangle$ is

$$\widehat{R}(g)|\psi\rangle \ . \tag{33.51}$$

The action on the dual vector is

$$\langle\psi|\widehat{R}(g)^\dagger \ . \tag{33.52}$$

Then

$$1 = \langle\psi|\psi\rangle \ , \tag{33.53}$$

is mapped to

$$\langle\psi|\widehat{R}(g)^\dagger\widehat{R}(g)|\psi\rangle \ . \tag{33.54}$$

We are interested in representations that preserve the unit norm of these states so that probabilities are preserved by symmetry operations. This is guaranteed if

$$\widehat{R}(g)^\dagger = \widehat{R}(g)^{-1} \ , \tag{33.55}$$

meaning that the representation is unitary.

Now, consider another state $|\phi\rangle$ which is related to $|\psi\rangle$ by a Hermitian operator

$$|\phi\rangle = \widehat{H}|\psi\rangle \ . \tag{33.56}$$

Because both $|\phi\rangle$ and $|\psi\rangle$ are physical states they must be affected by a symmetry transformation according to Eq. (33.51). This means that after the symmetry transformation we find

$$\widehat{R}(g)|\phi\rangle = \widehat{H}'\widehat{R}(g)|\psi\rangle \tag{33.57}$$

where $\widehat{H}'$ is the operator that $\widehat{H}$ is mapped to by the symmetry transformation.

Exercise 33.12 Show that

$$\widehat{H}' = \widehat{R}(g)\widehat{H}\widehat{R}(g)^{-1} \ . \tag{33.58}$$

Then $\widehat{H}'^{\dagger}$ is

$$\widehat{H}'^{\dagger} = \widehat{R}(g)^{-1\dagger}\widehat{H}\widehat{R}(g)^{\dagger} \tag{33.59}$$

Symmetry transformations should leave the physical predictions, that is the eigenvalues of the operators, invariant. A basic first step is that $\widehat{H}'$ should be a Hermitian operator. Comparing (33.58) and (33.59) we see that this also requires that $\widehat{R}(g)$ be a unitary representation.

The groups are want to represent in this way are typically Lie groups, which are reviewed in Sec. 37. Consequently, their elements have an exponential form

$$g = e^{i\theta X} \ , \tag{33.60}$$

where X is an element of the associated Lie algebra. This form is also inherited by the representations of the group and the representations of the algebra.

Exercise 33.13 Show that if a unitary matrix has the form

$$U = e^{-i\theta H} \ , \tag{33.61}$$

that H must be a Hermitian matrix. You may find the discussion surrounding Eq. (38.18) useful.

We have argued that Hermitian matrices play an important role in quantum mechanics. These operators can be associated with observable quantities. It is natural to consider what unitary operators the exponentiation of these Hermitian operators lead to.

Exercise 33.14 Prove, using the power series expansion of the exponential, that

$$e^{-\frac{i}{\hbar}\alpha^i \widehat{P}_i}|p\rangle = e^{-\frac{i}{\hbar}\alpha^i p_i}|p\rangle \ , \qquad \langle p|e^{-\frac{i}{\hbar}\alpha^i \widehat{P}_i} = \langle p|e^{-\frac{i}{\hbar}\alpha^i p_i} \ , \tag{33.62}$$

where α^i is a constant vector.

A simple example of this exponentiation is to use the momentum operator. We consider acting on a state of definite position with this exponentiated operator

$$e^{-\frac{i}{\hbar}\alpha^i \widehat{P}_i}|x\rangle \ , \tag{33.63}$$

where α^i is a constant vector with dimensions of length. We can then

express the identity as an integral over momentum states to find

$$\begin{aligned} e^{-\frac{i}{\hbar}\alpha^i \widehat{P}_i}|x\rangle &= \int d^3p\,|p\rangle\langle p|e^{-\frac{i}{\hbar}\alpha^i \widehat{P}_i}|x\rangle \\ &= \int d^3p\,|p\rangle\langle p|e^{-\frac{i}{\hbar}\alpha^i p_i}|x\rangle \\ &= \int d^3p\,|p\rangle e^{-\frac{i}{\hbar}\alpha^i p_i}\langle p|x\rangle \\ &= \int d^3p\,|p\rangle e^{-\frac{i}{\hbar}\alpha^i p_i}\frac{1}{(2\pi\hbar)^{3/2}}e^{-\frac{i}{\hbar}x^i p_i} \\ &= \int d^3p\,|p\rangle \frac{1}{(2\pi\hbar)^{3/2}}e^{-\frac{i}{\hbar}(x^i+\alpha^i)p_i} \\ &= \int d^3p\,|p\rangle\langle p|x+\alpha\rangle \\ &= |x+\alpha\rangle\ . \end{aligned} \tag{33.64}$$

So, we find that exponentiating the momentum operator produces translation in space, the position x^i was shifted by the vector α^i.

Exercise 33.15 Act with the exponentiation of the position operator on a state of definite momentum to show that

$$e^{-\frac{i}{\hbar}\widehat{X}^i k_i}|p\rangle = |p+k\rangle\ , \tag{33.65}$$

where k_i is a constant dual vector with dimensions of momentum. This shows that the exponentiation of the position operator produces an operator that shifts the momentum by a constant amount.

Subsection 33.3

Time Evolution

We have just seen that there is a reciprocal relationship between the momentum and position operators. The exponentiation of one produces a shift in the other. These operators are also related by the commutation relations given in Eq. (33.21) that were related to the action of the Poisson bracket of Hamiltonian mechanics Eq. (32.65). The Hamiltonian formulation also relates the Poisson bracket of operators to their evolution in time, see Eq. (32.66).

While not a proof, this motivates the form of the time evolution operator

$$e^{-\frac{i}{\hbar}\widehat{H}\,t}\ , \tag{33.66}$$

where $\widehat{H}$ is the Hamiltonian operator, some function of the position and momentum operators. This operator shifts a vector at some time t_0 to the

time $t_0 + t$, or if we take $t_0 = 0$

$$|\psi(t)\rangle = e^{-\frac{i}{\hbar}\widehat{H}\,t}|\psi(0)\rangle \ . \tag{33.67}$$

This form turns out to be correct as long as the Hamiltonian itself does not depend on time.

Exercise 33.16 Use the definition of quantum time evolution to show that

$$i\hbar\frac{d}{dt}|\psi(t)\rangle = \widehat{H}|\psi(t)\rangle \ . \tag{33.68}$$

This makes clear the connection between time evolution and the Hamiltonian. When the state vector is expressed in position space this is called the Schrödinger equation. In fact, this equation is typically taken as an axiom of quantum mechanics and the results discussed above follow from this assumption.

The eigenstates of the Hamiltonian can be denoted $|E_i\rangle$ so that

$$\widehat{H}|E_i\rangle = E_i|E_i\rangle \ . \tag{33.69}$$

We can then express any state vector (at what we define as $t = 0$) to be

$$|\psi(0)\rangle = \sum_i c_i|E_i\rangle \ , \tag{33.70}$$

where the c_i are given by

$$c_i = \langle E_i|\psi(0)\rangle \ . \tag{33.71}$$

Acting with the time evolution operator leads to

$$|\psi(t)\rangle = \sum_i c_i e^{-\frac{i}{\hbar}tE_i}|E_i\rangle \ . \tag{33.72}$$

This means that once we know the Hamiltonian of a system we can, in principle, determine the energy eigenstates. Then any vector can be expressed as a linear combination of those eigenstates. The time evolution of any such vector is simple to understand, in the energy basis. Each component evolves with a simple phase determined by the energy of the eigenstate it corresponds to.

Subsection 33.4

Simple Harmonic Oscillator

An important example of many of these principles is the **simple harmonic oscillator**. Recall from Sec. 32.1 that the classical Hamiltonian for simple

harmonic oscillation in one dimension is given by

$$H = \frac{1}{2m}p^2 + \frac{k}{2}x^2 \ . \tag{33.73}$$

In the quantum mechanical treatment we simply take the position and momentum to be operators that satisfy the standard commutation relations:

$$\widehat{H} = \frac{1}{2m}\widehat{P}^2 + \frac{k}{2}\widehat{X}^2 \ , \qquad \left[\widehat{X},\, \widehat{P}\right] = i\hbar\widehat{\mathbb{I}} \ . \tag{33.74}$$

We are most interested in determining the eigenvalues and eigenvectors of the Hamiltonian. To do this we define the operators

$$\widehat{a} = \sqrt{\frac{m\omega}{2\hbar}}\left(\widehat{X} + i\widehat{P}\frac{1}{m\omega}\right) \ , \qquad \widehat{a}^\dagger = \sqrt{\frac{m\omega}{2\hbar}}\left(\widehat{X} - i\widehat{P}\frac{1}{m\omega}\right) \tag{33.75}$$

where we have used the standard definition for the angular frequency of the oscillator

$$\omega = \sqrt{\frac{k}{m}} \ . \tag{33.76}$$

While the names may not be clear at this point, we call $\widehat{a}$ a **lowering operator** or **annihilation operator** and $\widehat{a}^\dagger$ a **raising operator** or **creation operator**.

Exercise 33.17 To see the connection between these operators and the Hamiltonian, show that

$$\widehat{a}^\dagger\widehat{a} = \widehat{H}/(\hbar\omega) - \frac{1}{2}\widehat{\mathbb{I}} \ . \tag{33.77}$$

Exercise 33.18 Using the commutation relations amongst the position and momentum operators show that

$$\left[\widehat{a},\, \widehat{a}^\dagger\right] = \widehat{\mathbb{I}} \ . \tag{33.78}$$

We can now write the Hamiltonian as

$$\widehat{H} = \hbar\omega\left(\widehat{a}^\dagger\widehat{a} + \frac{1}{2}\widehat{\mathbb{I}}\right) \ , \tag{33.79}$$

which is similar to the simple form we found after making a canonical transformation of the classical Hamiltonian (32.58). We define the energy eigenstates of the Hamiltonian by

$$\widehat{H}|E\rangle = E|E\rangle \ . \tag{33.80}$$

We quickly find that

$$\hat{a}^\dagger \hat{a}|E\rangle = \left(\frac{E}{\hbar\omega} - \frac{1}{2}\right)|E\rangle \ . \tag{33.81}$$

In other words, the eigenvectors of $\widehat{H}$ are also the eigenvectors of $\hat{a}^\dagger\hat{a}$, just with a shifted eigenvalue.

We label the eigenvalues of $\hat{a}^\dagger\hat{a}$ by

$$\hat{a}^\dagger \hat{a}|\lambda\rangle = \lambda|\lambda\rangle \ . \tag{33.82}$$

Note that because the vector

$$\hat{a}|\lambda\rangle \equiv |a\lambda\rangle \ , \tag{33.83}$$

is related to

$$\langle\lambda|\hat{a}^\dagger \ , \tag{33.84}$$

by Hermitian conjugation we have

$$\langle\lambda|\hat{a}^\dagger\hat{a}|\lambda\rangle = \langle a\lambda|a\lambda\rangle \geq 0 \ , \tag{33.85}$$

where the last inequality follows from the definition of the inner-product. Because

$$\langle a\lambda|a\lambda\rangle = \lambda\langle\lambda|\lambda\rangle \ , \tag{33.86}$$

we see that

$$\lambda \geq 0 \ . \tag{33.87}$$

Therefore, all the eigenvalues of $\hat{a}^\dagger\hat{a}$ are non-negative.

We now consider the action of $\hat{a}^\dagger\hat{a}$ on the vector $\hat{a}|\lambda\rangle$. Using the commutation relations amongst the creation and annihilation operators we find

$$\begin{aligned}\hat{a}^\dagger\hat{a}\hat{a}|\lambda\rangle &= \left(-\hat{a} + \hat{a}\hat{a}^\dagger\hat{a}\right)|\lambda\rangle \\ &= (-1+\lambda)\,\hat{a}|\lambda\rangle \ .\end{aligned} \tag{33.88}$$

In short, $\hat{a}|\lambda\rangle$ is also an eigenvector of $\hat{a}^\dagger\hat{a}$, but with eigenvalue $\lambda - 1$. This is why it is called a lowering operator. However, because all eigenvalues must be nonnegative, if $\lambda_{\min} < 1$ it must be the case that

$$\hat{a}|\lambda_{\min}\rangle = |\mathbf{0}\rangle \ , \tag{33.89}$$

where $|\mathbf{0}\rangle$ is the zero vector of the vector space. That is, there is a minimum eigenvalue. Because

$$\hat{a}^\dagger\hat{a}|\lambda_{\min}\rangle = \lambda_{\min}|\lambda_{\min}\rangle = |\mathbf{0}\rangle \ , \tag{33.90}$$

we see that this minimum eigenvalue is $\lambda_{\min} = 0$, because any vector multi-

plied by the number zero produces the zero vector. In other words, we find that the minimum energy eigenstate is $|0\rangle$, which is not the zero vector but a vector labeled by the number zero.

Exercise 33.19 Show that

$$\hat{a}^\dagger \hat{a} \hat{a}^\dagger |\lambda\rangle = (\lambda + 1)\, \hat{a}^\dagger |\lambda\rangle \ . \tag{33.91}$$

This shows that $\hat{a}^\dagger|\lambda\rangle$ is an eigenvector of $\hat{a}^\dagger\hat{a}$ with eigenvalue $\lambda+1$. This is why it is called a raising operator.

Now that we have determined the minimum eigenvalue we can use the raising operator to find all the states with larger λ. For instance,

$$\hat{a}^\dagger|0\rangle \propto |1\rangle \ . \tag{33.92}$$

We only know that these states are proportional because we are assuming all of the eigenvectors have been normalized. We determine this normalization by defining

$$\langle n-1|\hat{a}|n\rangle = N_n \ . \tag{33.93}$$

By taking the Hermitian conjugate we find

$$\langle n|\hat{a}^\dagger|n-1\rangle = N_n^* \ . \tag{33.94}$$

We also know from the eigenvalue relation that

$$\langle n|\hat{a}^\dagger\hat{a}|n\rangle = n \ . \tag{33.95}$$

As an aside, note that the operator

$$\widehat{N} \equiv \hat{a}^\dagger\hat{a} \ , \tag{33.96}$$

is often referred to as the **number operator**, because it simply returns the number n that defines an energy eigenstate

$$\widehat{N}|n\rangle = n|n\rangle \ . \tag{33.97}$$

We can insert the identity (as a sum over the complete basis of states) in the middle of Eq. (33.95) to find

$$\begin{aligned} \langle n|\hat{a}^\dagger\hat{a}|n\rangle &= \langle n|\hat{a}^\dagger\widehat{\mathbb{I}}\,\hat{a}|n\rangle \\ &= \sum_{n'} \langle n|\hat{a}^\dagger|n'\rangle\langle n'|\hat{a}|n\rangle \ . \end{aligned} \tag{33.98}$$

However, because we know the raising and lowering operators only change the eigenvalues by one we have

$$\langle n'|\hat{a}|n\rangle = \delta_{n',n-1}\langle n-1|\hat{a}|n\rangle \ . \tag{33.99}$$

This implies that

$$\begin{aligned}\langle n|\hat{a}^\dagger\hat{a}|n\rangle &= \sum_{n'}\langle n|\hat{a}^\dagger|n'\rangle\langle n'|\hat{a}|n\rangle \\ &= \langle n|\hat{a}^\dagger|n-1\rangle\langle n-1|\hat{a}|n\rangle \\ &= |N_n|^2 \ . \end{aligned} \tag{33.100}$$

From Eqs. (33.95) and (33.100) we find

$$|N_n|^2 = n \ . \tag{33.101}$$

Exercise 33.20 From Eq. (33.93) show that

$$\langle n+1|\hat{a}^\dagger|n\rangle = N_{n+1}^* \ . \tag{33.102}$$

The result from Eq. (33.101) does not quite determine the normalization factor. In principle we can have

$$N_n = \sqrt{n}e^{i\theta_n} \ , \tag{33.103}$$

for real phases θ_n. This implies that

$$\hat{a}^\dagger|0\rangle = \sqrt{1}e^{-i\theta_1}|1\rangle \ , \tag{33.104}$$

$$\left(\hat{a}^\dagger\right)^2|0\rangle = \sqrt{2}e^{-i(\theta_1+\theta_2)}|2\rangle \ , \tag{33.105}$$

$$\left(\hat{a}^\dagger\right)^3|0\rangle = \sqrt{6}e^{-i(\theta_1+\theta_2+\theta_3)}|3\rangle \ , \tag{33.106}$$

and so on for any finite n. Keeping track of these phases is irritating and they do not actually have any physical significance. The actual eigenvectors $|n\rangle$ are unchanged. To remove them we simply assume each raising operator does not introduce a phase relative to the eigenstate. Having made this choice we can write any element of the energy eigenbasis basis as

$$|n\rangle = \frac{1}{\sqrt{n!}}\left(\hat{a}^\dagger\right)^n|0\rangle \ . \tag{33.107}$$

To summarize, we have used purely algebraic means to completely determine the spectrum of the $a^\dagger a$ operator. This spectrum of the simple harmonic oscillator Hamiltonian is similar. We have

$$\widehat{H}|n\rangle = \hbar\omega\left(n+\frac{1}{2}\right)|n\rangle \ . \tag{33.108}$$

We can express any vector in terms of this same basis

$$|\psi\rangle = \sum_n c_n|n\rangle \ , \tag{33.109}$$

with

$$c_n = \langle n|\psi\rangle \ . \tag{33.110}$$

Time evolution of this state is given by Eq. (33.72), leading to

$$|\psi(t)\rangle = \sum_n c_n e^{-i\omega t(\frac{1}{2}+n)}|n\rangle \ . \tag{33.111}$$

In essence, once ω has been specified, we can completely predict the quantum mechanical evolution of any state.

Mathematical Resources

PART IX

Sec. 34. Vector Spaces
Sec. 35. Index Notation
Sec. 36. Groups
Sec. 37. Lie Groups
Sec. 38. Lie Algebras
Sec. 39. Representations
Sec. 40. Contour Integrals
Sec. 41. Green Functions
Sec. 42. Statistics

This part of the text introduces mathematical structures that you may be less familiar with. While you may be familiar with some of the topics here it may be useful read over them quickly and see what is emphasized.

Section 34

Vector Spaces

The concept of a vector space may be the most widely used mathematical construction in physics. We use finite-dimensional vector spaces to model the physical space we observe and we use infinite-dimensional vector spaces to describe the the solutions to the differential equations that describe how various systems change in time or space. In particle physics we need to keep track of several distinct vector spaces at once in order to accurately describe nature's structure. Because of this we rely heavily on index notation (described in the following section) in order to keep track of these various spaces.

We first give an abstract definition of a vector space and then examine some examples.

Definition

A **vector space** is a set V together with an Abelian group operation which is denoted by $+$. The identity element is often denoted as $\mathbf{0}$. This set is associated with a **field** F (for most of our purposes the fields of interest are either the real numbers $\mathbb{R}$ or the complex numbers $\mathbb{C}$) with a multiplicative identity element typically denoted as 1 and an additive identity denoted as 0. Elements of F act on elements of V, or there is a map $F \times V \to V$. For all $f \in F$ and $v \in V$ this mapping implies that $fv \in V$.

This mapping satisfies certain properties.

1. $(f_1 + f_2)v = f_1 v + f_2 v$
2. $(f_1 f_2)v = f_1(f_2 v)$
3. $f(v_1 + v_2) = f v_1 + f v_2$
4. $1v = v$
5. $f\mathbf{0} = \mathbf{0}$
6. $0v = \mathbf{0}$

Example

Euclidean Space in Two Dimensions: This vector space can be represented by elements $v \in \mathbb{E}^2$ of the form

$$v = \begin{pmatrix} a \\ b \end{pmatrix} \ , \tag{34.1}$$

where a, $b \in \mathbb{R}$. This is a real vector space and the associated field is the real numbers. The identity element is

$$\mathbf{0} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \ . \tag{34.2}$$

It should be clear that the two dimensional Euclidean vector space can be easily generalized to N dimensions. For convenience in writing, however, we often focus on the simple two dimensional case.

Exercise 34.1

Check that $\mathbb{E}^2$ satisfies all the conditions of a vector space.

Exercise 34.2

Generalize $\mathbb{E}^2$ to the complex vector space $\mathbb{C}^2$ where the elements of the vectors are complex numbers. What is the associated field?

Example

Polynomials: This vector space can be represented by elements of the form

$$r_0 + r_1 x + r_2 x^2 + \dots \ , \tag{34.3}$$

where $a_i \in \mathbb{R}$. This is a real vector space and the associated field is the real numbers. This space has infinite dimensions. Many systems in the physical sciences use this vector space. In other words, many physical systems can be described by polynomials.

Example

Fourier Series: This vector space can be represented by elements of the form

$$a_0 + a_1 \cos(\theta) + b_1 \sin(\theta) + a_2 \cos(2\theta) + b_2 \sin(2\theta) + \dots \ , \tag{34.4}$$

where a_i, $b_i \in \mathbb{R}$. This is a real vector space and the associated field is the real numbers. This space has infinite dimensions. Many systems in the physical sciences use this vector space. In other words, many physical systems can be usefully described by series of sines and cosines. There are many families of functions (Bessel, Legendre, etc) that similarly define a useful vector space for describing the solutions to certain families of problems.

An important property of vector spaces is the concept of linear independence. A related concept is notion of a basis for the space.

Definition A subset S of a vector space V is said to be **linearly independent** if for $f_i \in F$ and $v_i \in S$ the equation

$$f_1 v_1 + f_2 v_2 + \ldots + f_n v_n = \mathbf{0} \ , \tag{34.5}$$

implies that

$$f_1 = f_2 = \ldots = f_n = 0 \ . \tag{34.6}$$

That is the only linear combination of the vectors in S that produce the zero vector is the trivial case where each element of S is multiplied by the 0 of F.

Definition A subset S of a vector space V is said to **span** V if for all $v \in V$ there exist $f_i \in F$ such that for vectors $s_i \in S$ it is true that

$$v = f_1 s_1 + f_2 s_2 + \ldots \tag{34.7}$$

Definition A subset S of a vector space V is said to be a **basis** for V if it is linearly independent and spans V.

Example A basis for $\mathbb{E}^2$ is

$$\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix} , \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\} \ . \tag{34.8}$$

Another is

$$\left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix} , \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\} \ . \tag{34.9}$$

Exercise 34.3 Check that both of the above sets are indeed bases for $\mathbb{E}^2$.

Exercise 34.4 Find a basis for the space of polynomials and a basis for the Fourier series space.

A given element v of a vector space V will be expressed in different ways, depending on what basis is chosen. For instance

$$\begin{pmatrix} a \\ b \end{pmatrix} = a \begin{pmatrix} 1 \\ 0 \end{pmatrix} + b \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{a+b}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \frac{a-b}{2} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \ . \tag{34.10}$$

Any choice of basis is equally correct. Different bases may be more useful for different analyses, just as some problems are simpler in polar coordinates versus Cartesian coordinates. The physical system and its solution are the same in any basis, but some are more convenient. This also makes basis-independent quantities particularly important and useful.

If we are using a vector space to describe something about a physical

system it is often important that we can model a process that takes vectors in the space and maps them to other vectors in the space. For instance, suppose we describe the position of a particle by its position in three dimensional space. If the particle moves in space as time evolves then we must model time evolution as a process that maps one location (a three vector) to another. Often, this can be done with linear transformations, which we often refer to as operators on the vector space.

Definition A **linear transformation** M is a map from a vector space V to the same vector space: $M : V \to V$ such that for $v_i \in V$ and $f_i \in F$

$$M(f_1 v_1 + f_2 v_2) = f_1 M v_1 + f_2 M v_2 \ . \tag{34.11}$$

Example A trivial example for $\mathbb{E}^2$ is

$$M = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \ . \tag{34.12}$$

This simply maps a vector back to itself. Some slightly less trivial examples are

$$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \ , \text{ and } \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \ . \tag{34.13}$$

Exercise 34.5 Check that the above matrices are indeed linear transformations. What do each of these transformations do to a vector (or what vector do they map to).

Example A linear transformation on the space of polynomials is the differential operator $\frac{d}{dx}$.

Exercise 34.6 Check that $\frac{d}{dx}$ does indeed map any polynomial to another polynomial. Also check that it satisfies the requirements of being a linear transformation.

Like vectors, operators can be expressed in various bases. However, there are certain properties that are basis-independent. Most of these are tied to the eigenvalues of the operator.

Definition Given a linear transformation M from a vector space V to V, any vector $v \in V$ that satisfies

$$Mv = \lambda v \ , \tag{34.14}$$

where $\lambda \in F$, is called an **eigenvector** of M with **eigenvalue** λ.

Example

Consider $\mathbb{E}^2$ and the operator

$$M = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} . \tag{34.15}$$

It is simple to check that

$$M \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} , \quad M \begin{pmatrix} 1 \\ -1 \end{pmatrix} = - \begin{pmatrix} 1 \\ -1 \end{pmatrix} . \tag{34.16}$$

This shows that M has two eigenvalues, 1 and -1. The general methods of determining eigenvalues and eigenvectors for a given operator are covered in most linear algebra courses and are not covered here.

Example

We can also determine eigenvalues of $\frac{d}{dx}$ in the space of polynomials. Note that by using the power series expansion that

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots , \tag{34.17}$$

is an element of the vector space. Then from

$$\frac{d}{dx} e^{\alpha x} = \alpha e^{\alpha x} , \tag{34.18}$$

we see that $e^{\alpha x}$ is an eigenvector of this operator with eigenvalue α.

Other basis-independent quantities can be determined if we add some additional structure to the vector space and create an inner-product space.

Definition

A vector space V over the field F is extended to an inner-product space by defining a mapping from $V \times V \to F$, which we denote by $\langle \, , \rangle$, such that for any $v_1 \, , v_2 \in V$

$$\langle v_1 \, , v_2 \rangle = f \tag{34.19}$$

for some $f \in F$. This product satisfies three properties:

1. **Conjugation Symmetry:** $\langle v_1 \, , v_2 \rangle = \langle v_2 \, , v_1 \rangle^*$, where f^* is the complex conjugate of f. Note that for real vector spaces this means that the inner-produce is symmetric in its arguments.
2. **Linearity** in right argument: $\langle v \, , a v_1 + b v_2 \rangle = a \langle v \, , v_1 \rangle + b \langle v \, , v_2 \rangle$.
3. **Positive Semi-Definite:** For any $v \in V$ we have $\langle v \, , v \rangle \geq 0$ with equality only when $v = \mathbf{0}$.

Exercise 34.7 From the definition above show that

$$\langle av_1 + bv_2,\, v\rangle = a^*\langle v_1,\, v\rangle + b^*\langle v_2,\, v\rangle \ . \tag{34.20}$$

Example The usual vector dot product in Euclidean space is the most familiar example of an inner product. In $\mathbb{E}^2$ for example

$$\begin{pmatrix} a \\ b \end{pmatrix} \cdot \begin{pmatrix} c \\ d \end{pmatrix} = ac + bd \ . \tag{34.21}$$

Exercise 34.8 Show that for elements $v_1,\, v_2 \in V$ in the vector space of Fourier series an inner product can be defined by

$$\int_0^{2\pi} d\theta\, v_1\, v_2 \ . \tag{34.22}$$

With the inner-product we can make some choices that make vector spaces feel more familiar. Assuming that V has a basis $\{s_1,\, s_2,\, \ldots\}$ we can use the inner product to create a new basis $\{e_1,\, e_2,\, \ldots\}$ that is **orthonormal**. That is, such that

$$\langle e_i,\, e_j\rangle = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \ . \tag{34.23}$$

This is done by, for instance, defining

$$e_1 = \frac{1}{\sqrt{\langle s_1,\, s_1\rangle}} s_1 \ , \tag{34.24}$$

to ensure that e_1 is **normalized**. Then subtracting of any component of the other basis vectors s_i with $i \neq 1$ that lies in the direction of e_1. For example,

$$s_2' = s_2 - \langle e_1,\, s_2\rangle e_1 \ . \tag{34.25}$$

Then s_2' can be normalized to define e_2. This process can be enacted on the entire basis without producing the zero vector $\mathbf{0}$ because the original basis was linearly independent.

Exercise 34.9 Show that s_2' and e_1 are **orthogonal**, that is

$$\langle e_1,\, s_2'\rangle = 0 \ . \tag{34.26}$$

Using any basis we can also define the **matrix elements** of a linear transformation M (often referred to as an operator) on V. These are defined by

$$M_{ij} = \langle s_i,\, Ms_j\rangle \ . \tag{34.27}$$

The name of matrix element follows from the two dimensional array of numbers this formula defines. Note that it may be a finite matrix or one of infinite size.

With an inner-product in hand it is straightforward to define the concept of a dual vector. This may seem abstract, but is very helpful in keeping track of various aspects of physical systems. We also note that while dual vectors can be defined more generally, we restrict our discussion to those associated with the inner product.

Definition Given a vector space V and a field F, the **dual vector space** $V^\dagger$ (read V-dagger) is the collection of linear transformations from V to F. In this text we define these transformations $v^\dagger \in V^\dagger$ through the inner product by

$$v^\dagger = \langle v, \cdot \rangle \ . \tag{34.28}$$

That is, the action of $v^\dagger$ on some vector $u \in V$ is

$$v^\dagger(u) = \langle v, u \rangle \ . \tag{34.29}$$

One can show that $V^\dagger$ is a vector space. Thus, through the inner produce we have defined a vector space $V^\dagger$ that has a special relationship to V.

Example The dual space to $\mathbb{E}^2$ is the set of row vectors

$$v^\dagger = (a, b) \tag{34.30}$$

where $a, b \in \mathbb{R}$. Note that through usual matrix multiplication these row vectors acting on a column vector (and element of $\mathbb{E}^2$) produce a number.

Exercise 34.10 Show using matrix multiplication in $\mathbb{E}^2$ that

$$v^\dagger(v) = \langle v, v \rangle \ . \tag{34.31}$$

The operators that act on a vector space can also be related to operators that act on the dual space. Suppose that an operator M maps the vector v to the vector u, $Mv = u$. We define $u^\dagger$ in the usual way

$$u^\dagger = \langle u, \cdot \rangle = \langle Mv, \cdot \rangle \equiv v^\dagger M^\dagger \ , \tag{34.32}$$

where in the last definition we note that $M^\dagger$ (the mapping from $V^\dagger \to V^\dagger$) is understood to act to the *left* on $v^\dagger$ where as M acts to the right. The notation here is very suggestive of taking the **Hermitian conjugate** (or taking the complex conjugate and then the transpose) of a matrix. This is

deliberate for the following reason. Consider the matrix elements of M

$$M_{ij} = \langle s_i,\, Ms_j \rangle \ , \tag{34.33}$$

for some basis s_i of V. Then take the complex conjugate to find

$$M^*_{ij} = \langle s_i,\, Ms_j \rangle^* = \langle Ms_j,\, s_i \rangle = (M^\dagger)_{ji} \ . \tag{34.34}$$

As elements of a matrix, exchanging the indices is the same as exchanging row with columns, or taking the transpose. Thus, we see that the matrix elements of $M^\dagger$ are the complex transpose of the matrix elements of M.

Finally, we end with a discussion of products of vector spaces. We can form a **tensor product** of two vector spaces V_1 and V_2 which is denoted $V_1 \otimes V_2$. This space contains vectors that belong to each of the two spaces, but also tensors, which are objects that only live in the product of the spaces. A tensor of rank 2 requires the product of two vector spaces and tensors of arbitrarily large rank can be generated by combining more vector spaces, or more copies of one vector space.

This tensor product is taken to be bilinear, that is, for $u_i \in U$ and $v_i \in V$

$$u \otimes (v_1 + v_2) = u \otimes v_1 + u \otimes v_2 \ , \quad (u_1 + u_2) \otimes v = u_1 \otimes v + u_2 \otimes v \ . \tag{34.35}$$

In this way the basis of V and the basis of U can be combined by $\otimes$ to make a basis for the larger space $U \otimes V$.

Example

Suppose the vector space V is two dimensional with basis

$$\mathcal{B} = \{|s_1\rangle,\, |s_2\rangle\} \ , \tag{34.36}$$

and vector space V' is three dimensional with basis

$$\mathcal{B}' = \{|s'_1\rangle,\, |s'_2\rangle,\, |s'_3\rangle\} \ . \tag{34.37}$$

The basis elements of $V \otimes V'$ are

$$\{|s_1\rangle \otimes |s'_1\rangle,\, |s_1\rangle \otimes |s'_2\rangle,\, |s_3\rangle \otimes |s'_2\rangle,\, |s_1\rangle \otimes |s'_2\rangle,\, |s_2\rangle \otimes |s'_1\rangle,\, |s_2\rangle \otimes |s'_3\rangle\} \ . \tag{34.38}$$

The more concise notation

$$\{|s_1 s'_1\rangle,\, |s_1 s'_2\rangle,\, |s_1 s'_3\rangle,\, |s_2 s'_1\rangle,\, |s_2 s'_2\rangle,\, |s_2 s'_3\rangle\} \ . \tag{34.39}$$

is often used.

If M_U is a linear operator that acts on U and M_V is a linear operator that acts on V then we can create the tensor product $M_U \otimes M_V$ which acts

like

$$(M_U \otimes M_V)(u \otimes v) = (M_U u) \otimes (M_V v) \ , \tag{34.40}$$

for $u \in U$ and $v \in V$. The operator $M_U \otimes M_V$ can also be represented as a matrix in the product space. In order to do this, a choice must be made in how to order the basis elements.

Exercise 34.11

Suppose that in the two dimensional vector space V an operator M has the matrix elements

$$M = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \ , \tag{34.41}$$

in the basis

$$\mathcal{B} = \{|s_1\rangle, |s_2\rangle\} \ . \tag{34.42}$$

That is

$$m_{ij} = \langle s_i, Ms_j \rangle \ . \tag{34.43}$$

Suppose further that in the two dimensional vector space V' there is an operator M' with

$$M' = \begin{pmatrix} m'_{11} & m'_{12} \\ m'_{21} & m'_{22} \end{pmatrix} \ , \tag{34.44}$$

in the basis

$$\mathcal{B}' = \{|s'_1\rangle, |s'_2\rangle\} \ . \tag{34.45}$$

Show that If we order the basis elements of the product space $V \otimes V'$ as

$$\mathcal{B}_1 = \{|s_1 s'_1\rangle, |s_1 s'_2\rangle, |s_2 s'_1\rangle, |s_2 s'_2\rangle\} \ , \tag{34.46}$$

then the matrix elements of $M \otimes M'$ are

$$M \otimes M' = \begin{pmatrix} m_{11}m'_{11} & m_{11}m'_{12} & m_{12}m'_{11} & m_{12}m'_{12} \\ m_{11}m'_{21} & m_{11}m'_{22} & m_{12}m'_{21} & m_{12}m'_{22} \\ m_{21}m'_{11} & m_{21}m'_{12} & m_{22}m'_{11} & m_{22}m'_{12} \\ m_{21}m'_{21} & m_{21}m'_{22} & m_{22}m'_{21} & m_{22}m'_{22} \end{pmatrix} \ . \tag{34.47}$$

Then, show that if the basis elements are ordered as

$$\mathcal{B}_2 = \{|s_1 s'_1\rangle, |s_2 s'_1\rangle, |s_1 s'_2\rangle, |s_2 s'_2\rangle\} \ , \tag{34.48}$$

that

$$M \otimes M' = \begin{pmatrix} m_{11}m'_{11} & m_{12}m'_{11} & m_{11}m'_{12} & m_{12}m'_{12} \\ m_{21}m'_{11} & m_{22}m'_{11} & m_{21}m'_{12} & m_{22}m'_{12} \\ m_{11}m'_{21} & m_{12}m'_{21} & m_{11}m'_{22} & m_{12}m'_{22} \\ m_{21}m'_{21} & m_{22}m'_{21} & m_{21}m'_{22} & m_{22}m'_{22} \end{pmatrix} \ . \tag{34.49}$$

Exercise 34.12 For the the cases covered in the previous exercise, suppose

$$\mathbb{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} , \quad \mathbb{I}' = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} , \tag{34.50}$$

and

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} , \quad \sigma_1' = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} . \tag{34.51}$$

Using basis $\mathcal{B}_1$ find the matrix forms of

$$\mathbb{I} \otimes \mathbb{I}' , \quad \mathbb{I} \otimes \sigma_1' , \quad \sigma_1 \otimes \mathbb{I}' , \quad \sigma_1 \otimes \sigma_1' . \tag{34.52}$$

Then, find the matrix forms of these operators using the basis $\mathcal{B}_2$. Comment on similarities and differences.

SECTION 35

Index Notation

We now introduce a notation that often makes navigating vector spaces easier. Consider the vector equation

$$\vec{F} = m\vec{a} . \tag{35.1}$$

As written it appears to be one equation, but really we know it stands for three

$$\begin{pmatrix} F_x \\ F_y \\ F_z \end{pmatrix} = m \begin{pmatrix} a_x \\ a_y \\ a_z \end{pmatrix} . \tag{35.2}$$

In fact, here we have assumed Cartesian coordinates to describe the vector space $\mathbb{E}^3$, but we could use any other coordinates. In general, there are simply three equations. In index notation this is written as

$$F^i = ma^i , \tag{35.3}$$

where F^i is one of the *components* (the ith one) of the vector $\vec{F}$. Similarly, a^i is the ith component of $\vec{a}$. This is another way to describe all three equations.

We also know that the kinetic energy is

$$E_K = \frac{1}{2} m\vec{v} \cdot \vec{v} , \tag{35.4}$$

which involves the inner-product of the velocity vector with itself. In the

notation used in the previous section we would write this as

$$\vec{v} \cdot \vec{v} = \langle v,\, v \rangle \ , \tag{35.5}$$

and we know that in Cartesian coordinates that it is equal to

$$\vec{v} \cdot \vec{v} = v_x^2 + v_y^2 + v_z^2 \ . \tag{35.6}$$

In index notation we write the inner-product (for a real vector space) as

$$\langle v,\, v \rangle = \sum_{i=1}^{3} \sum_{j=1}^{3} g_{ij} v^i v^j \ . \tag{35.7}$$

Sometimes the object g_{ij} is called the metric on the space as it defines notions of distance. It is a two-index object, or a tensor of rank 2. Notice that we can exchange the order of the sums and the order of the vector components so

$$\begin{aligned} \sum_{i=1}^{3} \sum_{j=1}^{3} g_{ij} v^i v^j &= \sum_{j=1}^{3} \sum_{i=1}^{3} g_{ij} v^j v^i \\ &= \sum_{i=1}^{3} \sum_{j=1}^{3} g_{ji} v^i v^j \ . \end{aligned} \tag{35.8}$$

In the second line we have simply changed the labels we are using to track the two sums. This does not matter because they are both only place holders. By comparing the first and last sums we see that $g_{ij} = g_{ji}$, or that the metric is a **symmetric** tensor.

Definition

A rank two tensor S_{ij} is **symmetric** if every component S_{ij} is equal to S_{ji}. For instance, $S_{12} = S_{21}$. Similarly, an **antisymmetric** tensor A_{ij} is one in which

$$A_{ij} = -A_{ji} \ . \tag{35.9}$$

Note that this implies that any component of this tensor in which the two indices are equal, like A_{11} must be equal to zero because $A_{11} = -A_{11}$ by exchanging the two indices.

Exercise 35.1

By writing out all nine terms in the above sums, convince yourself that each equality in (35.8) is true.

These formulas need to be generalized slightly for complex vector spaces. In this case the norm of a vector v^a (in three complex dimensions) is

$$\langle v,\, v \rangle = |v_1|^2 + |v_2|^2 + |v_3|^2 = \sum_{i=1}^{3} \sum_{j=1}^{3} g_{ij} v^{*i} v^j \ , \tag{35.10}$$

where the metric is still a real array.

Exercise 35.2 Show that the definition (35.10) leads to the usual behavior of complex inner products

$$\langle u,\, v\rangle = \langle v,\, u\rangle^* \ . \tag{35.11}$$

The way we are defining things, sums over repeated indices always have one index up and one down. The upper index indicates a connection to the vector space, as employed above. A lower index is used to indicate a connection to the dual vector space. For instance, the dual vector

$$v^\dagger = \langle v,\, \cdot\rangle = \sum_{i=1}^{3} g_{ij} v^{*i} \equiv v^*_j \ , \tag{35.12}$$

is an object with one **free index** (an index that is not summed over) which is a subscript index.

Let's return to Eq. (35.7) before moving on. Note that it involves two instances of the velocity vector v^i, but that each instance uses a different index. This corresponds to the two different sums being calculated, one with index i and the other with index j. Note that these label names are arbitrary, we could have used a and b or א and ב. These are only used to define the sum, so we say they are **dummy indices**. Dragging these summation symbols $\sum$ around is bothersome. In the rest of this text they are not written explicitly. However, whenever you see a repeated index with one up and one down, then you know a sum is implied. This is sometimes called the **Einstein summation convention**.

We can then write the inner product as

$$\langle v,\, v\rangle = g_{ij} v^{*i} v^j = v^*_j v^j \ . \tag{35.13}$$

Written in this way we can recognize the metric as an object that maps elements of the vector space V to elements of the dual space $V^\dagger$. We sometimes speak loosely and say that we can use the metric to lower the index of v^i because

$$g_{ij} v^i = v_j \ . \tag{35.14}$$

Notice in this equation that the left hand side has one sum over a dummy index and one free index while the right hand side only has a free index. In any equation the number and label of the free indices must be the same. To be very clear, we can express the exact same information by

$$g_{\alpha z} v^\alpha = v_z \ , \tag{35.15}$$

though often using a different alphabet is used to signify something specific. For instance, Greek indices are often used to denote the four dimensions of space and time while Latin indices might be used only for the spatial

components of an object.

We can actually write the same information in many other ways, because we are writing an equation relating components of vectors and matrices. In particular, we have

$$g_{ij}v^{*i}v^j = v^{*i}g_{ij}v^j = v^{*i}v^j g_{ij} = g_{ij}v^j v^{*i} \ . \tag{35.16}$$

Note that in the last equality we are assuming that the components of v^i commute with the components of v^{*i}. Sometimes in quantum mechanical situations this is not the case, and the final equality would not hold. We are also free to relabel any dummy index

$$v^*_i v^i = v^*_j v^j = v^*_\zeta v^\zeta \ . \tag{35.17}$$

Example

In two dimensional Euclidean space $\mathbb{E}^2$ the metric is expressed as a two-by-two matrix. In Cartesian coordinates (x, y) this matrix is

$$g_{ij} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \ . \tag{35.18}$$

In this orthonormal basis the unit vectors are sometimes expressed as

$$\hat{x} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \ , \quad \hat{y} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \ . \tag{35.19}$$

In polar coordinates (r, θ) the metric is

$$g_{ij} = \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix} \ . \tag{35.20}$$

One immediately notices that the components of this matrix carry different units. This is an essential point, and has to do with our wanting to do calculus simply as we move from one space to another. We return to this in a moment, but for now let us proceed simply thinking about orthonormal basis vectors.

We can determine the unit vector of the orthonormal basis using the relations

$$x = r\cos\theta \ , \quad y = r\sin\theta \ . \tag{35.21}$$

A vector from the origin of length $r = \sqrt{x^2 + y^2}$ can be expressed in the two systems

$$\vec{r} = x\hat{x} + y\hat{y} = r\hat{r} \ . \tag{35.22}$$

Using (35.21) we immediately find that

$$\hat{r} = \cos\theta\hat{x} + \sin\theta\hat{y} \tag{35.23}$$

To determine $\hat{\theta}$ we require

$$\hat{r} \cdot \hat{\theta} = 0 \ , \quad \hat{\theta} \cdot \hat{\theta} = 1 \ , \tag{35.24}$$

and that when $\theta = 0$ that $\hat{\theta}$ points in the counterclockwise direction in the plane. So, for instance when $\theta = 0$ we have $\hat{\theta} = \hat{y}$. These requirements lead to

$$\hat{\theta} = -\sin\theta \hat{x} + \cos\theta \hat{y} \ . \tag{35.25}$$

Thus, for some vector v we find

$$v^i = v^r \hat{r} + v^\theta \hat{\theta} = \left(v^r \cos\theta - v^\theta \sin\theta \right) \hat{x} + \left(v^r \sin\theta + v^\theta \cos\theta \right) \hat{y} \ . \tag{35.26}$$

We can express a vector in either coordinate system

$$v^i = v^x \hat{x} + v^y \hat{y} = v^r \hat{r} + v^\theta \hat{\theta} \ , \tag{35.27}$$

and calculate the inner product of two vectors using the different metrics. However, we find a mismatch. Using the Cartesian coordinates we find

$$\delta_{ij} v^i v^j = v^x v^x + v^y v^y = v^r v^r + v^\theta v^\theta \ , \tag{35.28}$$

where we have used the relationships between the two sets of basis vectors to find the last equality. But using the polar coordinates we find

$$\delta_{ij} v^i v^j = v^r v^r + r^2 v^\theta v^\theta \ . \tag{35.29}$$

How can we reconcile this? The metric as it has been given actually corresponds to a much simpler way of determining the components of a vector in the two coordinate systems. If we are changing from one coordinate system, which we can call $x^i = (x, y)$ to a new coordinate system $x'^i = (r, \theta)$ then the components transform as

$$v^i = v'^j \frac{\partial x^i}{\partial x^j} \ . \tag{35.30}$$

This leads to

$$v^x = v^r \frac{\partial x}{\partial r} + v^\theta \frac{\partial x}{\partial \theta} = v^r \cos\theta - v^\theta r \sin\theta \tag{35.31}$$

$$v^y = v^r \frac{\partial y}{\partial r} + v^\theta \frac{\partial y}{\partial \theta} = v^r \sin\theta + v^\theta r \cos\theta \ . \tag{35.32}$$

Note that with this definition if v^x has dimensions of length, then so does v^r but v^θ is dimensionless. Using these these relations it is easy to show that

$$v^x v^x + v^y v^y = v^r v^r + r^2 v^\theta v^\theta \ , \tag{35.33}$$

as the two metrics predict. The important lesson here is that to get the calculus right when we introduce dimensionless coordinates like θ the components of vectors need not have the same dimensions.

Example

In three dimensional Euclidean space $\mathbb{E}^3$ the metric is expressed as a three-by-three matrix. In Cartesian coordinates this matrix is

$$g_{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \ . \tag{35.34}$$

While in spherical polar coordinates (r, θ, ϕ) the metric is

$$g_{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2 \sin^2\theta \end{pmatrix} \ . \tag{35.35}$$

In any set of coordinates the inner product is given by

$$g_{ij} v^i v^j \tag{35.36}$$

Exercise 35.3

Using the metrics given above, find the inner product of a vector with itself in both Cartesian and spherical polar coordinates. You may find it useful recall that the coordinates are related by

$$x = r\cos\phi\sin\theta \ , \quad y = r\sin\phi\sin\theta \ , \quad z = r\cos\theta \ . \tag{35.37}$$

Show that the value of the inner product is the same in each coordinate systems.

So far, index notation may not seem any better than any other notation. Let's consider a more interesting example. Consider the equation that describes how an external torque $\vec{\tau}$ affects the motion of an object with moment of inertia tensor $\overleftrightarrow{I}$:

$$\vec{\tau} = \overleftrightarrow{I}\,\vec{\alpha} \ , \tag{35.38}$$

where $\vec{\alpha}$ is the angular acceleration. Note that our vector-like notation is a little harder to deal with when a tensor is involved. This is because a tensor space is the Cartesian product of vector spaces (or perhaps dual-vector spaces). How this equation is often taught is that $\overleftrightarrow{I}$ is a matrix.

$$\begin{pmatrix} \tau_x \\ \tau_y \\ \tau_z \end{pmatrix} = \begin{pmatrix} I_{xx} & I_{xy} & I_{xz} \\ I_{yx} & I_{yy} & I_{yz} \\ I_{zx} & I_{zy} & I_{zz} \end{pmatrix} \begin{pmatrix} \alpha_x \\ \alpha_y \\ \alpha_z \end{pmatrix} \ . \tag{35.39}$$

This is true but sometimes cumbersome to use.

It is clear from the intent of the equation that $\vec{\tau}$ and $\vec{\alpha}$ are vectors in the space that we are using to describe the rotating object. It must be that $\overleftrightarrow{I}$ acts on $\vec{\alpha}$ to produce $\vec{\tau}$, so it maps from the vector space to the vector space. It is an operator, or a linear transformation on $\mathbb{E}^3$.

In index notation we write this as

$$\tau^i = I^i{}_j \alpha^j \ . \tag{35.40}$$

The notation with one index up and one down on the inertia tensor does not make much of a difference in Euclidean space with Cartesian coordinates, however, in more general cases it is does carry essential information relating the dual space to the vector space. Note that we have use the summation convention and that there is only one free index on either side of the equality. In index notation we can also write this as

$$\tau^i = \alpha^j I^i{}_j \ , \tag{35.41}$$

because this is an equation relating components of arrays, not a vector equation. If, however, we want to recover the correct order for matrix multiplication we must use (35.40). In short, matrix multiplication corresponds to the summed indices being adjacent.

The moment of inertia tensor is an example of how to express an operator on a vector space in index notation. There is, of course, one operator that is always defined. This is simply the identity operator that maps a vector to itself.

Definition

Kronecker delta: This operator is defined by

$$\delta^i_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \ , \tag{35.42}$$

independent of any choice of coordinates.

Exercise 35.4

By explicitly writing out the sums, show that

$$v^j \delta^i_j = v^i \ , \qquad v_j \delta^j_i = v_i \ . \tag{35.43}$$

We can use the Kronecker delta to define the inverse metric g^{ij} by

$$g^{ik} g_{kj} = \delta^i_j \ . \tag{35.44}$$

Exercise 35.5

Find the inverse metric of $\mathbb{E}^3$ in both Cartesian and spherical polar coordinates.

We can use the inverse metric to raise indices

$$g^{ij}v_j = v^i \ . \tag{35.45}$$

In other words, the inverse metric is a mapping from the dual vector space to the vector space. It is the inverse map of the metric, hence the name. It also acts as the inner product on the dual space

$$g^{ij}v_i^* v_j \ . \tag{35.46}$$

Exercise 35.6 Show that

$$g^{ij}v_i^* v_j = g_{ij}v^{*i}v^j \ . \tag{35.47}$$

We also need to specify how taking the Hermitian conjugate (or the transpose) acts on operators using index notation. For vectors and dual vectors we have shown that

$$\left(v^i\right)^\dagger = v_i^* = v^{*j}g_{ji} \ , \qquad (v_i)^\dagger = v^{*i} = v_j^* g^{ji} \ . \tag{35.48}$$

The action of an operator $O^i{}_j$ on a vector v^j is a vector. We also know that taking the Hermitian conjugate of an operator gives an object that maps dual vectors to dual vectors. This implies the form

$$\left(O^i{}_j v^j\right)^\dagger = v_j^* O^{*j}{}_i \ . \tag{35.49}$$

By equating the results of Eq. (35.48) to the vector $u^i = O^i{}_j v^j$ and Eq. (35.49) we find

$$\begin{aligned}
\left(u^i\right)^\dagger =& u_i^* \\
v_m^* \left(O^i{}_m\right)^\dagger =& O^{*k}{}_j v^{*j} g_{ki} \\
=& O^{*k\ell} g_{\ell j} g^{jm} v_m^* g_{ki} \\
=& v_m^* g_{ik} O^{*k\ell} \delta_\ell^m \\
=& v_m^* O_i^{*m} \\
=& v_m^* O^{\dagger m}{}_i
\end{aligned} \tag{35.50}$$

This shows that

$$\left(O^j{}_i\right)^\dagger = O_j^{*i} = O^{\dagger i}{}_j \ . \tag{35.51}$$

In short, we must keep track of where the † is and the order of the indices.

Exercise 35.7 Use index notation to prove that for any two matrices A and B that

$$(AB)^\dagger = B^\dagger A^\dagger . \tag{35.52}$$

There is another object that takes the same value in any coordinate system, and is therefore of special interest.

Definition

Levi-Civita Tensor: This object is an N rank tensor (actually a pseudo-tensor), where N is the dimension of the vector space. In an orthonormal basis its components are either 0, 1, or -1 and in general it is completely antisymmetric. This means that the exchange of any two of its indices changes the value by a minus sign.

Example

In a two dimensional space the Levi-Civita is a rank two tensor given by

$$\varepsilon_{ij} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{35.53}$$

This object is antisymmetric in that $\varepsilon_{ij} = -\varepsilon_{ji}$. This implies that when any two indices take on the same value the component must be zero. Any of its indices can be raised using the metric.

Exercise 35.8

Assuming the Cartesian metric of $\mathbb{E}^2$ show that

$$\varepsilon^{ij} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{35.54}$$

and then that

$$\varepsilon^{ij}\varepsilon_{mn} = \delta^i_m\delta^j_n - \delta^i_n\delta^j_m \ , \quad \varepsilon^{ik}\varepsilon_{jk} = \delta^i_j \ , \quad \varepsilon^{ij}\varepsilon_{ij} = 2 \tag{35.55}$$

Example

In a three dimensional space the Levi-Civita is a rank three tensor ε_{ijk} with nonzero components

$$\begin{aligned} \varepsilon_{123} = \varepsilon_{231} = \varepsilon_{312} = 1 \\ \varepsilon_{132} = \varepsilon_{213} = \varepsilon_{321} = -1 \end{aligned} \tag{35.56}$$

This object is antisymmetric in that

$$\varepsilon_{ijk} = -\varepsilon_{jik} = -\varepsilon_{kji} = -\varepsilon_{ikj} \ . \tag{35.57}$$

One can show that

$$\varepsilon^{ijk}\varepsilon_{mnk} = \delta^i_m\delta^j_n - \delta^i_n\delta^j_m \ , \tag{35.58}$$

and

$$\varepsilon^{imk}\varepsilon_{jmk} = 2\delta^i_j \ . \tag{35.59}$$

It is the Levi-Civita tensor that defines the cross product in $\mathbb{E}^3$:

$$\vec{A} \times \vec{B} = \varepsilon_{ijk}A^iB^j \ . \tag{35.60}$$

Note that the resulting object is actually a dual vector. This does not have much of an effect in Euclidean space with Cartesian coordinates as raising the the index by the metric amounts to multiplication by the identity matrix.

Exercise 35.9

To get some idea of the utility of index notation consider that

$$\vec{A} \times \vec{B} \times \vec{C} = \varepsilon_{ijk} A^i B^j g^{k\ell} \varepsilon_{\ell mn} C^m \ . \tag{35.61}$$

Use the identities above to prove that

$$\vec{A} \times \vec{B} \times \vec{C} = \vec{B}\left(\vec{A} \cdot \vec{C}\right) - \vec{C}\left(\vec{A} \cdot \vec{B}\right) \ . \tag{35.62}$$

Now that cross and dot products have been defined we can translate nearly any vector equation in to index notation. The only remaining quantities to translate involve the ∇ operator. We define this by first expressing the coordinates used in $\mathbb{E}^3$ as the components of a vector x^i. Then, we define

$$\nabla = \frac{\partial}{\partial x^i} \equiv \partial_i \ . \tag{35.63}$$

Note that this operator is also defined as a dual vector. In general, a derivative with respect to a vector is a dual vector. However, we also define

$$\partial^i = g^{ij} \partial_j \ , \tag{35.64}$$

which gives the correct vector version of the operator.

Example

We can now write Maxwell's equations (Heavyside-Lorentz units)

$$\nabla \cdot \vec{E} = \rho \qquad \nabla \cdot \vec{B} = 0 \tag{35.65}$$

$$\nabla \times \vec{E} = -\frac{1}{c}\frac{\partial \vec{B}}{\partial t} \qquad \nabla \times \vec{B} = \frac{1}{c}\vec{J} + \frac{1}{c}\frac{\partial \vec{E}}{\partial t} \tag{35.66}$$

as

$$\partial_i E^i = \rho \qquad \partial_i B^i = 0 \tag{35.67}$$

$$\varepsilon_{ijk} \partial^i E^j = -\frac{1}{c}\partial_t B_k \qquad \varepsilon_{ijk} \partial^i B^j = \frac{1}{c} J_k + \frac{1}{c}\partial_t E_k \tag{35.68}$$

Exercise 35.10

In a popular undergraduate text on electromagnetism, proving the following identity is said to be an exercise "only for masochists."

$$\nabla(\vec{A} \cdot \vec{B}) = \vec{A} \times (\nabla \times \vec{B}) + \vec{B} \times (\nabla \times \vec{A}) + (\vec{A} \cdot \nabla)\vec{B} + (\vec{B} \cdot \nabla)\vec{A} \ . \tag{35.69}$$

Beginning with the right-hand side, use index notation to prove the identity nearly immediately.

We end this section with a discussion of how certain operations that can be defined on operators are indicated in index notation. The simpler of these is the trace of an operator $M^i{}_j$.

Definition

The **Trace** of an operator $M^i{}_j$ is determined by summing over the diagonal elements of the operator

$$\mathrm{Tr}(M) = M^i{}_i \ , \tag{35.70}$$

where the sum is implied as usual. Though we do not prove it here, this quantity is equal to the sum of the eigenvalues of $M^i{}_j$ and as such is independent of basis.

Note that this means that the trace of the product of two operators is

$$\begin{aligned}\mathrm{Tr}(AB) =& A^i{}_j B^j{}_i \\ =& B^j{}_i A^i{}_j \\ =& \mathrm{Tr}(BA) \ .\end{aligned} \tag{35.71}$$

This is straightforward to generalize the product of any number of matrices.

Exercise 35.11

Show that

$$\mathrm{Tr}(ABC) = \mathrm{Tr}(CAB) = \mathrm{Tr}(BCA) \ . \tag{35.72}$$

Definition

The **Determinant** of an N dimensional operator $M^i{}_j$ is obtained from

$$\varepsilon_{i_1 i_2 \ldots i_N} M^{i_1}{}_{j_1} M^{i_2}{}_{j_2} \ldots M^{i_N}{}_{j_N} = \mathrm{Det}(M) \varepsilon_{j_1 j_2 \ldots j_N} \ . \tag{35.73}$$

The determinant turns out to be equal to the product of the eigenvalues of M, so it is also independent of basis.

Example

For a two dimensional matrix the definition of the determinant is easy to check. Consider an arbitrary matrix

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \ . \tag{35.74}$$

The left-hand side of the determinant formula is

$$\begin{aligned}\varepsilon_{ij} M^i{}_k M^j{}_\ell =& \varepsilon_{12} M^1{}_k M^2{}_\ell + \varepsilon_{21} M^2{}_k M^1{}_\ell \\ =& 1\, M^1{}_k M^2{}_\ell - 1\, M^2{}_k M^1{}_\ell\end{aligned} \tag{35.75}$$

This is a two by two array which we can write as

$$\begin{pmatrix} M^1{}_1 M^2{}_1 - M^1{}_1 M^1{}_1 & M^1{}_1 M^2{}_2 - M^2{}_1 M^1{}_2 \\ M^1{}_2 M^2{}_1 - M^2{}_2 M^1{}_1 & M^1{}_2 M^2{}_2 - M^2{}_2 M^1{}_2 \end{pmatrix} = (ad - bc) \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} . \tag{35.76}$$

This is exactly the right-hand side of the determinant definition.

Exercise 35.12 Use the above definition of the determinant to calculate the determinant of a three by three matrix and check that it agrees with the standard formula.

Note that we can use this definition to evaluate the determinant of the identity operator in any vector space. We have

$$\varepsilon_{i_1 i_2 \ldots i_N} \delta^{i_1}{}_{j_1} \delta^{i_2}{}_{j_2} \ldots \delta^{i_N}{}_{j_N} = \mathrm{Det}(I_N) \varepsilon_{j_1 j_2 \ldots j_N} , \tag{35.77}$$

where I_N is the identity operator in an N dimensional vector space, which in index notation is simply a Kronecker delta. But by simply using the Kroncker form the left-hand side immediately becomes

$$\varepsilon_{i_1 i_2 \ldots i_N} \delta^{i_1}{}_{j_1} \delta^{i_2}{}_{j_2} \ldots \delta^{i_N}{}_{j_N} = \varepsilon_{j_1 j_2 \ldots j_N} , \tag{35.78}$$

which proves that $\mathrm{Det}(I_N) = 1$ for all N.

We can also prove another useful property of determinants. Consider the determinant of the product of two matrices

$$\varepsilon_{i_1 i_2 \ldots i_N} A^{i_1}{}_{k_1} B^{k_1}{}_{j_1} A^{i_2}{}_{k_2} B^{k_2}{}_{j_2} \ldots A^{i_N}{}_{k_N} B^{k_N}{}_{j_N} = \mathrm{Det}(AB) \varepsilon_{j_1 j_2 \ldots j_N} . \tag{35.79}$$

Using the definition of the determinant for one matrix we can rewrite the left-hand side as

$$\begin{aligned} &\varepsilon_{i_1 i_2 \ldots i_N} A^{i_1}{}_{k_1} B^{k_1}{}_{j_1} A^{i_2}{}_{k_2} B^{k_2}{}_{j_2} \ldots A^{i_N}{}_{k_N} B^{k_N}{}_{j_N} \\ &\quad = \mathrm{Det}(A) \varepsilon_{k_1 k_2 \ldots k_N} B^{k_1}{}_{j_1} B^{k_2}{}_{j_2} \ldots B^{k_N}{}_{j_N} \\ &\quad = \mathrm{Det}(A) \mathrm{Det}(B) \varepsilon_{j_1 j_2 \ldots j_N} . \end{aligned} \tag{35.80}$$

This proves that

$$\mathrm{Det}(AB) = \mathrm{Det}(A) \mathrm{Det}(B) . \tag{35.81}$$

Exercise 35.13 By using the fact that $\mathrm{Det}(I_N) = 1$ and Eq. (35.81), prove that for any invertible matrix M that

$$\mathrm{Det}\left(M^{-1}\right) = \frac{1}{\mathrm{Det}(M)} . \tag{35.82}$$

SECTION 36

Groups

In this section we introduce the mathematical definition of a group. Before making these definitions let me motivate why. Groups are some of the simplest, but nontrivial, algebraic systems. Their simplicity is also their power. In physical systems, and certainly in particle physics, one often hears how important symmetries are. In most cases these symmetries are identified with specific mathematical groups. By identifying these groups we can learn much about a system, even when its complexity makes precise calculations all but impossible.

The ingredients of a group, sometimes abstractly denoted as G, are a set of objects and an operation that acts on pairs of those objects.

Definition

A **group** is defined to be a set G along with a binary operation $\cdot$ that maps $G \times G \to G$. This operation $\cdot$ is a "multiplication" of two elements of G and must satisfy two properties.

1. The set G is **closed** under the multiplication. In notation, for all a, $b \in G$, it is true that $a \cdot b \in G$.
2. The multiplication is **associative**. That is for all a, b, $c \in G$ it is true that $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.

In addition, the following two conditions regarding the elements of G under the multiplication must hold:

1. The set G must include an **identity** element, often denoted e. Any element multiplied by the identity produces the same element back. That is, for all $a \in G$ we have $a \cdot e = e \cdot a = a$.
2. Every element of G has an **inverse**. An element multiplied by its inverse produces the identity. In notation, for each $a \in G$ there exists an element $a^{-1} \in G$ such that $a \cdot a^{-1} = a^{-1} \cdot a = e$.

This definition is very broad, which is why general results about groups are powerful. However, it is useful to consider a few examples. Let's start with the first and somewhat trivial example.

Example

The Trivial Group: This group has one element, the identity.

$$G = \{e\} \ . \tag{36.1}$$

Note that all the requirements for being a group are satisfied, but in an unsatisfying way. That is, $e \cdot e = e$, $e \cdot (e \cdot e) = (e \cdot e) \cdot e$, $e \in G$, and

$e^{-1} = e$.

A less trivial example, but still quite simple, is a group known as Z_2.

Example **Z_2:** This group has two elements

$$Z_2 = \{e, a \,|\, a \cdot a = e\} \ . \tag{36.2}$$

This reads that the group has two elements with the constraint that $a \cdot a = e$.

Exercise 36.1 Demonstrate that Z_2 satisfies all the conditions of a group. What is a^{-1}?

These examples may still seem quite abstract. In fact, this is something of the point. Manifestations of various groups can be recognized in many situations within the sciences (and of course mathematics) if you know how to spot them. However, in most occurrences it is a particular **representation** of a group that appears. Formal definitions of representations are discussed in Sec. 39 but are more general than we need in this section. Loosely speaking, a representation of a group is formed by mapping the abstract group into some space we are interested in.

Consider two different representations of Z_2.

Example The set $\{1, -1\}$ under the standard multiplication of real numbers is one representation of Z_2. This representation is one-dimensional.

Example The set

$$\left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \right\} \tag{36.3}$$

under matrix multiplication is a representation of Z_2. This representation is 2-dimensional.

Exercise 36.2 Check that the two example sets above have all the characteristics of Z_2.

Example **Z_N:** The cyclic groups are straightforward generalizations of Z_2. It has N elements each of which is a power of some basic element, or **generator**, which we denote as a.

$$Z_N = \left\{ e, a, a^2, \ldots, a^{N-1} \,|\, a^N = e \right\} \tag{36.4}$$

In defining the properties of a group we have been careful *not* to assume that the elements of the group **commute**, that is that $a \cdot b = b \cdot a$. While some elements of a group always commute, the identity commutes with every element and any element commutes with its inverse, in general

elements need not commute with each other. In the very simple examples shown above the elements do commute and such groups are said to be Abelian. Nonabelian groups include elements that do not commute. The group of symmetries of an equilateral triangle is an example.

Example

The Dihedral Group of Order 6: This is sometimes denoted as D_3 or D_6. It has six elements, but is associated to the symmetries of an equilateral triangle. Consider the figure to the right. (The numbering of the vertices is for convenience, they are not intended to be part of the triangle itself.) What actions leave the triangle looking the same? The identity element is simply to do nothing, but what about nontrivial transformations?

The equilateral triangle looks identical under a rotation by $2\pi/3$ radians and by $4\pi/3$ radians. A rotation by 2π radians is identical to no rotation at all, which is the identity element of the group. The triangle also has three reflection axes, one from each vertex that intersects the opposite face perpendicularly. Reflection about each of these axes is an element of the group. And acting with the same reflection twice is equivalent to the identity.

Let's denote the rotation by $2\pi/3$ by a and the reflection about the vertices by b_1, b_2, and b_3, where we imagine labeling the top vertex as 1 and then proceed clockwise around the triangle labeling 2 and 3. This labeling can be thought of as fixed while the triangle moves. We can then write the group as

$$D_6 = \left\{e,\, a,\, a^2,\, b_1,\, b_2,\, b_3\right\} \ . \tag{36.5}$$

We must now choose how to denote the order of actions. We choose to put a second action to the left of the first. For example, acting with a and then b_1 is denoted by $b_1 a$. (Note here that we have dispensed with the "$\cdot$" to show multiplication. This is often left as implied when confusion is unlikely.) This choice of ordering anticipates the action of operators in vector spaces which is central to this text.

Exercise 36.3

Using a triangle with vertices labeled 1, 2, and 3 as outlined above, show that

$$b_2 = b_1 a = a^2 b_1 \ , \tag{36.6}$$

and

$$b_3 = b_1 a^2 = a b_1 \ . \tag{36.7}$$

From these results conclude that

$$b_1 = a b_1 a \ . \tag{36.8}$$

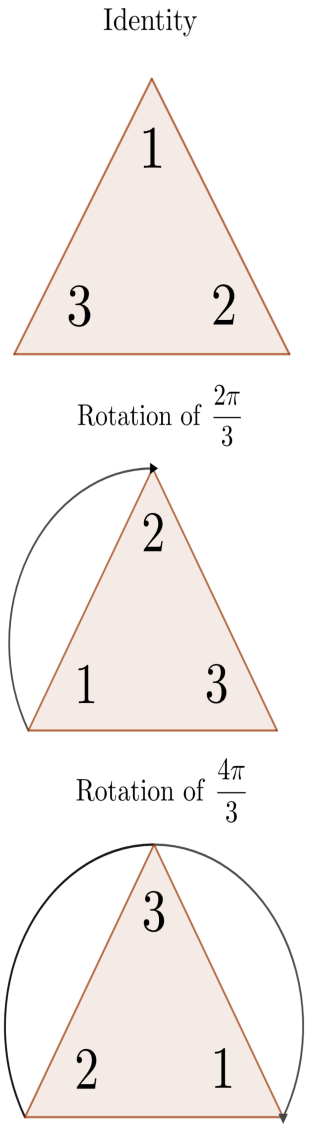
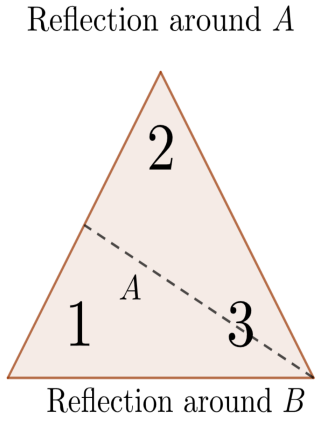
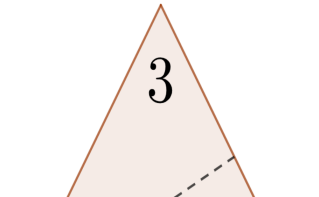


Figure 30. Three of the elements of D_6.





Exercise 36.4 The results of the preceding exercise show that all of the elements of D_6 can be generated by two basic elements a and $b_1 \equiv b$. Show that the group can be expressed as

$$D_6 = \left\{e,\, a,\, a^2,\, b,\, ba,\, ab\right\} \ . \tag{36.9}$$

The final result of the above exercise makes clear that $ab \neq ba$, or that D_6 is not an Abelian group. Something else to notice is that D_6 has several **subgroups**.

Definition A **subgroup** H of a group G is a nonempty subset of G that is closed under the group multiplication of G and for every $h \in H$ it is also true that $h^{-1} \in H$.

Note that the other requirements of being a group as inherited from G or follow from these requirements.

Exercise 36.5 Show that each of

$$\left\{e,\, a,\, a^2\right\} \ , \quad \{e,\, b\} \ , \quad \{e,\, ba\} \ , \quad \{e,\, ab\} \ , \tag{36.10}$$

are subgroups of D_6. Each of these groups is a manifestation of groups we have already introduced. Identify the group that corresponds to each subgroup.

As shown the previous exercise a given group may have several different subgroups. Every group contains the trivial subgroup $\{e\}$, but this is not very interesting. Similarly, by the definition given above the entire group G is technically also a subgroup of G. Again, this is not very interesting. Sometimes the term **proper subgroup** is used to refer to subgroups other than these two uninteresting cases. In this text we will typically simply use the term subgroup to refer to proper subgroups to be a little more concise. Much about groups can be understood by determining their subgroups and the aspects of those subgroups. It is, however, beyond the scope of this text.

We do, however, discuss one topic that pertains to symmetry breaking. A subgroup H of a group G can be used to partition G into sets. These sets turn out to have a group structure of their own. To see this we first define cosets of a subgroup.

Definition The **left coset** and **right coset** of any group element $g \in G$ of a subgroup H of G are given by

$$gH \quad \text{and} \quad Hg \ , \tag{36.11}$$

respectively. Here g is taken to act to the left or right (respectively) of

each element in H, resulting in a set of group elements. This is set is often only a subset, not a subgroup, of G.

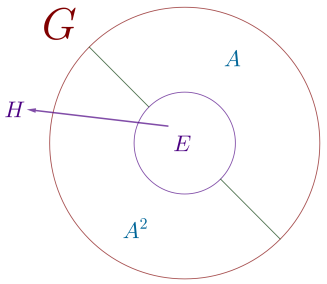


Figure 32. Partition of G by cosets of normal subgroup H.

Example

Consider the subgroup

$$H = \{e,\, b\} \tag{36.12}$$

of D_6. We are interested in finding all the right cosets of H. By explicit calculation we find

$$He = \{e,\, b\}\ , \qquad Hb = \{e,\, b\}\ , \qquad Ha = \{a,\, ba\}\ , \tag{36.13}$$

$$Ha^2 = \left\{a^2,\, ab\right\}\ , \qquad Hba = \{a,\, ba\}\ , \qquad Hab = \left\{a^2,\, ab\right\}\ . \tag{36.14}$$

Note here that the elements of the cosets are not ordered. We can see that there are three right cosets

$$E = \{e,\, b\}\ , \qquad A = \{a,\, ba\}\ , \qquad A^2 = \left\{a^2,\, ab\right\}\ . \tag{36.15}$$

Note that these cosets are disjoint (they have no elements in common) and that A and A^2 are not subgroups. Note also that all of the elements of G belong to exactly one of the cosets. It is in this sense that these cosets are a partition of G.

These cosets have been suggestively labeled because they form a group structure. We take

$$EA \tag{36.16}$$

to mean each element of E multiplied by each element of A. We then find that

$$EA = AE = A\ , \quad EA^2 = A^2E = A^2\ , \quad AA = A^2\ , \quad AA^2 = A^2A = E\ . \tag{36.17}$$

In short, the right cosets of H furnish a representation of Z_3. The fact that the cosets form a group is actually a special property that only occurs because H is a normal subgroup of G.

Definition

A subgroup N of G is called a **normal subgroup** if

$$gN = Ng \tag{36.18}$$

for all $g \in G$. In other words, the left and right cosets of N are equal.

Exercise 36.6

Find the left cosets of

$$H = \{e,\, b\}\ , \tag{36.19}$$

and determine if H is a normal subgroup of D_6.

Exercise 36.7 Determine whether or not

$$\{e,\, ba\} \quad \text{and} \quad \{e,\, ab\} \tag{36.20}$$

are normal subgroups of D_6.

Exercise 36.8 Show that

$$G = \left\{e,\, a,\, a^2\right\} \tag{36.21}$$

is a normal subgroup of D_6. What group structure does its cosets manifest?

Remember that G is simply the group Z_3. This relationship between D_6, its normal subgroup, and the group structure of the resulting cosets is written as

$$D_6/Z_3 \cong Z_2 \ . \tag{36.22}$$

This is read as D_6 mod Z_3 is isomorphic to Z_2. (Two groups that are isomorphic are "the same." It means there is a bijective map between the two groups.) We also call D_6/Z_3 a quotient group.

The final topic of this introduction to groups is the concept of the product of groups. That is, how one might combine the action of two or more groups being present in a given system. We provide the definition for the **direct product** of two groups, but this extends easily to the direct product of any finite number of groups.

Definition The **direct product** (or sometimes Cartesian product) of two groups G_1 and G_2 with binary operations $\cdot_1$ and $\cdot_2$, respectively, is denoted by $G_1 \times G_2$. It is composed of all ordered pairs of the form

$$(g_1,\, g_2) \ , \tag{36.23}$$

where $g_1 \in G_1$ and $g_2 \in G_2$. The group multiplication law is defined by

$$(a_1,\, a_2) \cdot (b_1,\, b_2) = (a_1 \cdot_1 b_1,\, a_2 \cdot_2 b_2) \ . \tag{36.24}$$

It is worth highlighting that the multiplications of the two groups need not be the same or even similar in any way to define this product group structure. For finite groups it is also easy to see (check this yourself if not seen immediately) that the number of elements in a product group is simply the product of the number of elements in each group.

Exercise 36.9 Show that the direct product of two groups does indeed satisfy all the requirements of being a group.

Exercise 36.10 Explicitly construct the product group $Z_2 \times Z_2$. How many elements does this group have? Is it abelian or nonabelian? Is this group equivalent to Z_4? Why or why not?

While this completes our introduction to mathematical groups, please be aware that we have not even begun to cover the richness of this topic. This is merely meant to provide a minimal introduction. Group theory plays an important role in many physical models in addition to mathematics and further study is highly recommended. A classic text that is focused on physics applications is [25].

Section 37

Lie Groups

In this section we discuss a special and incredibly useful class of groups, called Lie (pronounced Lee) groups. If you are not familiar with groups, review Sec. 36. As with other sections, we barely scratch the surface of this rich subject, but this should provide a useful starting point. A very useful reference for Lie group, Lie algebras (Sec. 38) and how they are represented in vector spaces (Sec. 39) is [91]. A classic text focused on particle physics is [92]. Lie groups are infinite dimensional; in fact they can be (and are) discussed as manifolds and topological spaces. (For our purposes a manifold is a space in which calculus can be defined.) Many important Lie groups can be represented by groups of matrices (or numbers) and so we restrict our definition to matrix Lie groups for simplicity.

Definition A **Matrix Lie Group** is a set G of invertible (do you understand why?) matrices (of any dimension) such that any sequence M_n of matrices in G converges either to an element of G or to a matrix that is not invertible. This last point is a technicality that we usually do not need to deal with. Typically the sets we encounter are simply closed (under matrix multiplication) sets of matrices.

Example **U(1):** The unitary group of dimension one. A matrix U is **unitary** if

$$U^\dagger = U^{-1} \,, \tag{37.1}$$

or if its complex transpose (Hermitian conjugate) is equal to its inverse. One-by-one matrices are simply numbers. The Hermitian conjugate of a number is just its complex conjugate. So, this group is the set of numbers

whose complex conjugate is equal to their inverse

$$z^* = \frac{1}{z} \ . \tag{37.2}$$

In other words $z^* z = 1$. These are exactly numbers of the form

$$z = e^{i\theta} \ , \tag{37.3}$$

where $\theta \in [0, 2\pi)$. (Remember that $e^{i2\pi N} = 1$ for any integer N.) The multiplication law for this group is the usual multiplication of numbers. Note that this set has an infinite number of elements, because we can choose any particular θ.

Exercise 37.1 Check that the U(1) group defined in the above example satisfies all the requirements of being a mathematical group given in Sec. 36.

Example **U(N):** Unitary Groups. Unitary matrices of any dimension N also form a Lie group. This group also has an important subgroup **SU(N)**. The S here denotes a "special" matrix, which means its determinant is equal to one.

Exercise 37.2 Check that the U(N) group defined above satisfies all the requirements of being a mathematical group. It may be useful to use the property that for any matrices U_1 and U_2 that

$$(U_1 U_2)^\dagger = U_2^\dagger U_1^\dagger \ . \tag{37.4}$$

Also, check that SU(N) is a subgroup. It may help to use the property that for any matrices U_1 and U_2 that

$$\mathrm{Det}\,(U_1 U_2) = \mathrm{Det}\,(U_1)\,\mathrm{Det}\,(U_2) \ . \tag{37.5}$$

Unitary groups are extremely important in quantum mechanics. Time evolution and the action of most symmetries are represented by unitary operators acting on the vector space of possible states. It is these operators that ensure that probabilities are preserved under the action of the operator.

Example **O(N):** Orthogonal Groups. Orthogonal matrices O satisfy

$$O^T = O^{-1} \ , \tag{37.6}$$

where O^T denotes the transpose of the matrix. Such matrices, of any dimension N, also form a Lie group. This group also has an important

subgroup SO(N). The S again denotes determinant equal to one. The proof that these sets are indeed groups is nearly identical to the exercise above.

There is another way of writing the condition defining orthogonal matrices. We employ index notation to write

$$O^T O = I \quad \Rightarrow \quad O^{T\,i}{}_k O^k{}_j = \delta^i_j \ . \tag{37.7}$$

The index notation relationship for operators under transposition, see Eq. (35.51), allows us to write

$$\begin{aligned} \delta^n_j =& O_k{}^n O^k{}_j \\ =& g_{k\ell} O^\ell{}_m g^{mn} O^k{}_j \ . \end{aligned} \tag{37.8}$$

We then lower the n index on both sides

$$\begin{aligned} g_{ni}\delta^n_j =& g_{k\ell} O^\ell{}_m g^{mn} g_{ni} O^k{}_j \\ g_{ij} =& g_{k\ell} O^\ell{}_m \delta^m_i O^k{}_j \\ g_{ij} =& g_{\ell k} O^\ell{}_i O^k{}_j \ . \end{aligned} \tag{37.9}$$

For the usual orthogonal matrices the metric g_{ij} on the vector space is simple the Euclidian metric that is an identity matrix, just the number one for each diagonal element and zeroes everywhere else. However, if the this metric is generalized to a diagonal matrix with N entries that are the number one and M entries that are negative one, then the matrices that satisfy Eq. (37.9) are elements of the group O(N, M). The metric that governs special relativity is preserved by the **Lorentz group**, O(1, 3), which is discussed in much more detail in Sec. 7.

There are many other Lie groups, but these are the ones that appear most often in particle physics. We mentioned above that unitary groups are essential to quantum mechanics. Orthogonal groups are closely tied to rotations in physical space. For instance, the proper rotation of a two dimensional plane are described by elements of SO(2). These matrices can be expressed as

$$O = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \ . \tag{37.10}$$

Exercise 37.3 Show that the matrix given above is orthogonal.

Exercise 37.4 Consider the action of O an a vector v in the x-y plane

$$v' = \begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \tag{37.11}$$

Describe how v' relates to the original vector v.

Similarly, SO(3) is related to proper rotations in three dimensional space. The groups O(2) and O(3) allow for proper rotations and reflections about the origin $x, y \to -x, -y$. These last are related to the idea of **parity**. These types of transformations are sometimes called **improper rotations**. They are easy to recognize because the determinant of the matrix that represents them is always -1.

It is also worth knowing that some of these Lie groups are closely related or even identical. For instance, we see above that the group SO(2) is determined by one parameter θ. Recall that the group U(1) is also parameterized by a single parameter. Imagine a mapping M :U(1)→SO(2) which is defined by

$$M\left(e^{i\theta}\right) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} . \tag{37.12}$$

It is easy to see that each element of U(1) is mapped to exactly one element of SO(2) and that no element of SO(2) is missed by the mapping. We say that this map is **bijective**. If the map preserves the group multiplication law, that is if for any two elements z_1 and z_2 in U(1) we have

$$z_1 z_2 = M(z_1)M(z_2) , \tag{37.13}$$

then we say the mapping is a **homomorphism**. If a homomorphism between to groups is bijective it is an **isomorphism**. Two groups that are isomorphic (related by an isomorphism) are essentially "the same."

Exercise 37.5 Show that the mapping between U(1) and SO(2) given above is an isomorphism.

It turns out that SU(2) and SO(3) are almost isomorphic. It can be shown that there is a two-to-one mapping from SU(2) into SO(3). While these groups are not exactly the same, many of the their properties are identical, as determined by their Lie algebras.

Section 38

Lie Algebras

The simplest Lie group we discussed was U(1), in which every group element has the form

$$e^{i\theta} , \tag{38.1}$$

for a real number θ. The product of any two elements

$$e^{i\theta_1}e^{i\theta_2} = e^{i(\theta_1+\theta_2)} \ , \tag{38.2}$$

illustrates how the product of group elements leads to a sum of parameters in the exponential. Because of the exponential form, all of the group multiplication structure can be expressed by the sum of terms in the exponent. This is essentially the connection between Lie groups and Lie algebras. All Lie group elements can be expressed in an exponential form. The objects that are exponentiated are elements of the associated Lie algebra.

As we have focused on matrix Lie groups one may wonder what is exponentiated to produce a matrix? The answer is that one must exponentiate a matrix to produce a matrix. This is defined through the power series definition of an exponential.

Definition

The **exponentiation of a square matrix** M is defined by

$$e^M = \sum_{n=0}^{\infty} \frac{M^n}{n!} \ , \tag{38.3}$$

Though we do not prove it, this series converges for square matrices with real or complex entries.

The exponentials of numbers have useful properties and many of these also apply to the exponential of matrices. However, there are some important differences. Many of these are quite simple to prove from the power series definition.

Exercise 38.1

Prove that

$$e^{0_N} = I_N \ , \tag{38.4}$$

where 0_N is the N by N matrix with all entires equal to zero and I_N is the N by N identity matrix.

Exercise 38.2

Prove that for any N by N matrix X

$$\left(e^X\right)^\dagger = e^{X^\dagger} \ . \tag{38.5}$$

Remember that taking the Hermitian conjugate of a product of matrices is equal to the product of Hermitian conjugates in reverse order

$$(M_1 M_2 \cdots M_n)^\dagger = M_n^\dagger M_{n-1}^\dagger \cdots M_1^\dagger \ . \tag{38.6}$$

Exercise 38.3 Prove that for any N by N matrix X and invertible N by N matrix S

$$e^{SXS^{-1}} = Se^X S^{-1} \ . \tag{38.7}$$

Considering that matrices like S are often used to diagonalize other matrices through

$$SXS^{-1} = D \ , \tag{38.8}$$

where D is a diagonal matrix. We see that the same matrix that diagonalizes X also diagonalizes its exponentiation.

In standard courses in linear algebra one learns to diagonalize matrices through **similarity transformations** S, as in Eq. (38.8). These are essentially changes of basis in a vector space. For many operators there exists a basis in which the operator is diagonal, with the diagonal elements given by the eigenvalues of the matrix.

Exercise 38.4 Prove that the exponential of a diagonal matrix is a diagonal matrix. From this result conclude that if a diagonalizable matrix X has eigenvalues v_i, then the matrix

$$e^X \ , \tag{38.9}$$

has eigenvalues e^{v_i} .

We can now prove a useful identity for diagonalizable matrices. It uses several properties of the determinant and trace which can be found in Sec. 35. Given an N by N matrix X with eigenvalues v_i that is diagonalized by the similarity transform S, we find that

$$\begin{aligned} \mathrm{Det}\left(e^X\right) &= \mathrm{Det}\,(S)\,\mathrm{Det}\left(e^X\right)\mathrm{Det}\left(S^{-1}\right) = \mathrm{Det}\left(Se^X S^{-1}\right) \\ &= \prod_{i=1}^{N} e^{v_i} = e^{\sum_i^N v_i} \\ &= e^{\mathrm{Tr}\left(SXS^{-1}\right)} = e^{\mathrm{Tr}\left(XS^{-1}S\right)} \\ &= e^{\mathrm{Tr}(X)} \ . \end{aligned} \tag{38.10}$$

In short, the determinant of the exponential of a matrix is equal to the exponential of the trace of that matrix.

One of the essential characteristics of matrices is that generally the product of two matrices X and Y satisfies

$$XY \neq YX \ . \tag{38.11}$$

In other words, matrix multiplication does not commute. To characterize how much two matrices do not commute we use the **commutator** of two

matrices

$$XY - YX \equiv [X, Y] \ . \tag{38.12}$$

Note that

$$[X, Y] = -[Y, X] \tag{38.13}$$

and that if $[X, Y] = 0$ it means that X and Y commute, or $XY = YX$. Though we do not prove it here, it is true that

$$e^X e^Y = e^Y e^X = e^{X+Y} \qquad \text{if } [X, Y] = 0 \ . \tag{38.14}$$

Or, in general

$$e^X e^Y = e^{X+Y+\text{terms that depend on } [X,Y]} \ . \tag{38.15}$$

Exercise 38.5 Using the power series definition of the matrix exponential show that

$$e^X e^Y = 1 + X + Y + \frac{1}{2}(X+Y)^2 + \frac{1}{2}[X, Y] + \dots \ , \tag{38.16}$$

where the terms that have been neglected involve products of at least three matrices. The general proof of Eq. (38.14) follows this line of argument, but for general terms in the power series.

Exercise 38.6 Using Eq. (38.14) prove that for any square matrix X

$$\left(e^X\right)^{-1} = e^{-X} \ . \tag{38.17}$$

Note that this proves that every exponential of a matrix has an inverse.

Exercise 38.7 Prove that for any square matrices X and Y and complex numbers θ_1 and θ_2 that

$$e^{\theta_1 X} e^{\theta_2 Y} = e^{\theta_1 X + \theta_2 Y} \ , \tag{38.18}$$

if $[X, Y] = 0$.

Definition **Lie Algebra:** Given a matrix Lie group G, the matrix Lie algebra $\mathfrak{g}$ is the set of all matrices X such that

$$e^{i\theta X} \in G \tag{38.19}$$

for any real number θ.

Example The Lie algebra for the group U(1) is simply the number 1. This is because all elements of U(1) are of the form

$$e^{i \times 1 \times \theta} \ , \tag{38.20}$$

for a real number θ.

Example

The Lie algebra for the group U(N) with N greater than one are the matrices X such that

$$U = e^{iX} \ , \tag{38.21}$$

is a unitary matrix. Note that

$$U^\dagger = e^{-iX^\dagger} \ , \quad U^{-1} = e^{-iX} \ . \tag{38.22}$$

Since unitary matrices satisfy $U^\dagger = U^{-1}$ we see that

$$X^\dagger = X \ , \tag{38.23}$$

or that X is Hermitian. Thus the Lie algebra $\mathfrak{u}(N)$ corresponding to the Lie group U(N) is the set of all N by N Hermitian matrices.

Exercise 38.8

Determine the Lie algebra for the group O(N) of orthogonal matrices. That is, matrices O with $O^T = O^{-1}$.

Further restrictions can be placed on the Lie algebras related to SU(N) and SO(N) by making use of the identity

$$\mathrm{Det}\left(e^X\right) = e^{\mathrm{Tr}(X)} \ . \tag{38.24}$$

In short, if the Lie group elements are to have unit determinant then the Lie algebra matrices must be traceless.

Example

A specific example is the algebra $\mathfrak{so}(2)$. The previous exercise shows that the elements of this algebra are two dimensional, antisymmetric, pure imaginary matrices. In other words, they are Hermitian matrices that are purely imaginary. Such matrices are all of the form

$$X = \begin{pmatrix} 0 & -i\theta \\ i\theta & 0 \end{pmatrix} \ , \tag{38.25}$$

where θ is a real number.

We expect elements of the group SO(2) to be of the form

$$e^{iX} \ , \tag{38.26}$$

which we evaluate using the power series definition. This can be evaluated exactly by first noting that

$$(iX)^n = \begin{cases} \theta^n I_2 (-1)^{n/2} & n \text{ even} \\ \theta^{n-1} iX (-1)^{(n-1)/2} & n \text{ odd} \end{cases} \ , \tag{38.27}$$

where I_2 is the two by two identity matrix. This allows us to split the exponential power series into two, one over even $n = 2m$ and one over odd $n = 2m + 1$:

$$\begin{aligned} e^{iX} =& I_2 \sum_{m=0}^{\infty} (-1)^m \frac{\theta^{2m}}{(2m)!} + iX \sum_{m=0}^{\infty} (-1)^m \frac{\theta^{2m+1}}{(2m+1)!} \\ =& I_2 \cos\theta + iX \sin\theta \\ =& \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} . \end{aligned} \tag{38.28}$$

These are the elements of SO(2) as expected.

While we have determined what the sets of objects are that belong to a Lie algebra, we have not yet justified the 'algebra' label. First notice that from the definitions given above is X is in the Lie algebra then so is θX where θ is a real number. We do not prove it, but it is also true that if X and Y are in the Lie algebra then so it $X + Y$. This is straightforward to show when X and Y commute, but it is actually true in general.

Exercise 38.9 Let X and Y be elements of a Lie algebra which commute, that is $[X, Y] = 0$. Prove that $X + Y$ is also in the Lie algebra.

These two properties are enough to show that the Lie algebra is a vector space. The step from vector space to algebra requires a mapping that takes two elements of the space and produces an element of the space. For Lie algebras the commutator is exactly this mapping. While we do not prove it, if X and Y are in a Lie algebra $i\,[X, Y]$ is also in the Lie algebra.

Exercise 38.10 Prove that the commutator is bilinear, that is

$$[\alpha X + \beta Y, Z] = \alpha\,[X, Z] + \beta\,[Y, Z] \tag{38.29}$$
$$[Z, \alpha X + \beta Y] = \alpha\,[Z, X] + \beta\,[Z, Y] \ . \tag{38.30}$$

Exercise 38.11 Prove that the commutator satisfies the Jacobi identity:

$$[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0 \ . \tag{38.31}$$

Exercise 38.12 Show that the commutator satisfies a kind of product rule

$$[XY, Z] = X\,[Y, Z] + [X, Z]\,Y \ . \tag{38.32}$$

Because the Lie algebra of a group is a vector space it is spanned by a set of basis elements. Because every element in the algebra can be expressed as a linear combination of the basis elements, once we know the

commutators of the basis elements we know everything, in principle, about the entire algebra.

Definition For a Lie algebra $\mathfrak{g}$ with basis set

$$\left\{X^1\, X^2, \ldots, X^n\right\} \tag{38.33}$$

we define the **structure constants** f^{abc} of $\mathfrak{g}$ for that basis by

$$\left[X^a,\, X^b\right] = i f^{abc} X^c \ , \tag{38.34}$$

where a summation is implied over the repeated c index. Once the structure constants are known, then the commutator of any two elements of the algebra can be calculated. The basis elements of the algebra can always be chosen such that the structure constants are real and completely antisymmetric in their indices.

Exercise 38.13 Use the Jacobi identity in Eq. (38.31) to show that

$$f^{bcd} f^{ade} + f^{cad} f^{bde} + f^{abd} f^{cde} = 0 \ . \tag{38.35}$$

Example The simplest nontrivial example of the full structure of a Lie algebra is probably $\mathfrak{su}(2)$. It also is quite important for physics applications so we work things out in some detail. From the above discussion we know that this Lie algebra must be composed of traceless, Hermitian, two-by-two matrices.

Exercise 38.14 A general complex two-by-two matrix has four complex entries, and hence is specified by eight real numbers. We would then expect the real vector space of such matrices to be eight dimensional. Show that requiring the $\mathfrak{su}(2)$ matrices to be Hermitian

$$X^\dagger = X, \tag{38.36}$$

and traceless puts five constraints on these eight real numbers, reducing the dimension of the vector space to three real dimensions.

Example $\mathfrak{su}(2)$ **continued:** The exercise above implies that the algebra can be spanned by a basis of three elements. Which we denote as τ^a where a is an index that runs from one to three. It is sometimes convenient to normalize the matrices (rescale them) so that

$$\mathrm{Tr}(\tau^a \tau^b) = \frac{1}{2}\delta^{ab} \ . \tag{38.37}$$

The basis most often used for $\mathfrak{su}(2)$ is

$$\tau^a = \frac{1}{2}\sigma^a \ , \tag{38.38}$$

where the σ^a are the Pauli matrices.

Definition

The **Pauli matrices** are defined to be

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \ , \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \ , \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \ . \tag{38.39}$$

These matrices often show up in discussions of spin-1/2 particles.

Exercise 38.15

Show by explicit calculation that

$$\sigma_i \sigma_j = \mathbb{I}_2 \delta_{ij} + i\varepsilon_{ijk}\sigma_k \ , \tag{38.40}$$

where $\mathbb{I}_2$ is the identity matrix in two dimensions.

Exercise 38.16

Show that

$$\mathrm{Tr}(\tau^a \tau^b) = \frac{1}{2}\delta^{ab} \ , \tag{38.41}$$

for

$$\tau^a = \frac{1}{2}\sigma^a \ . \tag{38.42}$$

Then, show that

$$\left[\tau^a, \tau^b\right] = i\varepsilon^{abc}\tau^c \ . \tag{38.43}$$

This defines the structure constants for $\mathfrak{su}(2)$. Note that they are both real and antisymmetric.

Now that we have determined the Lie algebra we can check to make sure that exponentiation leads to elements of the correct group.

Exercise 38.17

Check that elements of the Lie group SU(2) can be expressed as

$$U(x) = e^{i\tau^a \theta_a(x)} \ , \tag{38.44}$$

where the θ_a are real numbers.

First, show that

$$\theta_a \sigma^a = \begin{pmatrix} \theta_3 & \theta_1 - i\theta_2 \\ \theta_1 + i\theta_2 & -\theta_3 \end{pmatrix} \equiv M \tag{38.45}$$

satisfies

$$M^n = \begin{cases} \vec{\theta}^n \mathbb{I}_2 & n \text{ even} \\ \vec{\theta}^{n-1} M & n \text{ odd} \end{cases} . \tag{38.46}$$

Where we have used the notation

$$\vec{\theta}^2 = \theta_1^2 + \theta_2^2 + \theta_3^2 . \tag{38.47}$$

Use the fact that

$$U(x) = e^{i\tau^a \theta_a(x)} = \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{i}{2}\right)^n M^n \tag{38.48}$$

to show that

$$U(x) = \mathbb{I}_2 \cos\left(\frac{|\vec{\theta}|}{2}\right) + \frac{i}{|\vec{\theta}|} M \sin\left(\frac{|\vec{\theta}|}{2}\right) . \tag{38.49}$$

Finally, check that this matrix is unitary and has unit determinant.

The point of all this is that we are now able to translate products of SU(2) matrices into linear combinations of the Pauli matrices. These are quite simple to manipulate, making them a powerful tool for understanding the Lie group.

Example

Consider $\mathfrak{so}(3)$, whose associated Lie group SO(3) is related to proper rotations in three dimensional space. The elements of this algebra are three-by-three, real valued, antisymmetric matrices. They are also traceless, but any antisymmetric matrix is traceless. (Prove this to yourself.) A real three-by-three antisymmetric matrix is specified by three real numbers, so the basis for this algebra has three elements. We take them to be

$$T^1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} , \quad T^2 = \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix} , \quad T^3 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} . \tag{38.50}$$

Exercise 38.18

Show that

$$\mathrm{Tr}(T^a T^b) = 2\delta^{ab} , \tag{38.51}$$

and that

$$\left[T^a, T^b\right] = i\varepsilon^{abc} T^c . \tag{38.52}$$

Note that these are the same structure constants as we found for $\mathfrak{su}(2)$!

It is because the structure constants for $\mathfrak{su}(2)$ and $\mathfrak{so}(3)$ can be chosen to be identical that they are sometimes loosely said to be the same algebra.

However, the groups they generate are subtly different. This difference appears as we consider representations of these groups in physical systems.

Finally, we note that a subalgebra of some algebra is composed of a subset of the algebra elements that satisfy all the mathematical requirements of being an algebra. This means that there is some subset of the basis elements that are closed under commutation.

Example

The simplest subalgebra is restrict to a single generator. Because the commutator of any matrix with itself vanishes

$$[T, T] = 0 \ , \tag{38.53}$$

restricting to a single generator is a closed subgroup. The group elements are clearly only specified by one parameter

$$g = e^{-iT\theta} \ , \tag{38.54}$$

so the subgroup can be identified as a representation of U(1).

As in the example above, whatever set of generators make up a sub-algebra of the Lie algebra can be exponentiated to produce a Lie group. This group is a sub-group of the Lie group generated by the full algebra. Consequently, when we are interested in finding sub-groups of a Lie group we often consider the simpler process of finding sub-algebras of the Lie algebra.

Section 39

Representations

Much of our discussion so far has been abstract. However, the reason behind all this is to use these tools to model physical systems and processes. These models often use various vector spaces to characterize the dynamics of a physical system. Therefore, we need to know how groups are represented in these vector spaces. As with Lie groups both [91] and [92] provide additional explanation, results, and applications.

The first important point is that the structure of a given group can be represented in more than one way in a given vector space. These different representations of groups lead to physically distinct situations.

Definition

A **representation** R of a group G is a homomorphism, a mapping $R : G \to V$ that preserves the group multiplication structure as in

$$R(g_1)R(g_2) = R(g_1 g_2) \ , \tag{39.1}$$

from the group into operators acting on a vector space V. A mapping that is one-to-one, that is, for every element of the group there is a distinct operator, is called a **faithful representation**.

Example

A trivial example is the **trivial representation**. This representation maps every element of G to the identity operator acting on V. You can check that this does satisfy all the requirements of a representation, but it is not a very interesting one. Sometimes we say that an object (like a quantum field) that is unaffected by the action of a group is a **singlet**. This simply means that it is acted upon only by the trivial representation of a group.

A few simple representations of finite groups were given in Sec. 36. We do not discuss them much further other than to list a few important properties. Before we do this, however, we need to introduce two more ideas.

Definition

Two representations R_1 and R_2 of a group G are said to be **equivalent** if they are related by a similarity transformation

$$R_2 = S^{-1} R_1 S \ . \tag{39.2}$$

As we have discussed, similarity transformations amount to expressing an operator, or matrix, in a new basis. Such transformations may appear to change the representation, but they are not different representations for most purposes.

Exercise 39.1

Given a representation R of a group we can always define another representation, denoted $\overline{R}$ by complex conjugation:

$$\overline{R}(g) = R(g)^* \ . \tag{39.3}$$

Show that this is in fact a representation of the group. This representation is called the **complex conjugate representation** or sometimes just the conjugate representation.

We are often interested in combinations of objects that are **invariant** (do not change) under the action of a group transformation. For example, consider a vector v^a in a vector space V. We want the inner-product of the vectors v^a and u^a to be invariant under symmetry transformation. This inner product can be written as

$$\langle u,\, v \rangle = u^*_a v^a = u^{*a} g_{ab} v^b \ , \tag{39.4}$$

where g_{ab} is the metric that relates vectors to dual vectors. Suppose these

vectors are acted upon by the action of a group, through a representation R

$$u^a \to R^a{}_b u^b \ , \qquad v^a \to R^a{}_b v^b \ . \tag{39.5}$$

Then we have

$$\begin{aligned} u^{*a} g_{ab} v^b \to & R^{*a}{}_c u^{*c} g_{ab} R^b{}_d v^d \\ = & u^{*c} R^{\dagger a}_c g_{ab} R^b{}_d v^d \ . \end{aligned} \tag{39.6}$$

The action of the symmetry preserves the inner product if

$$R^{\dagger a}_c g_{ab} R^b{}_d = g_{cd} \ , \tag{39.7}$$

which is the requirement that R be a unitary matrix. Thus, unitary representations are of particular interest. In real vector spaces orthogonal matrices play a similar role.

Definition

A representation R of a group G on the vector space V is said to be **reducible** if there is a subspace V_1 of V such that for all vectors $v \in V_1$ and $g \in G$

$$R(g)v \in V_1 \ . \tag{39.8}$$

In other words, if there is a collection of vectors within the vector space that is closed under the action of the representation. We typically only care about subspaces that are not all of V or just the zero vector. Any representation that is not reducible is **irreducible**.

What does this definition mean? It means that in regard to this representation of G the vector space splits up into two subspaces $V = V_1 \oplus V_2$. Here the $\oplus$ symbol denotes a **direct sum**. If V has dimension N then the dimensions of V_1 and V_2 must satisfy $N_1 + N_2 = N$. We can always decide to think about a vector space in this way, as the direct sum of smaller subspaces, but the essential point is that the reducible representation never mixes these two spaces, they remain separately closed under R. In fact, R is equivalent to a representation that is block diagonal

$$\begin{array}{c} N \\ N \end{array} \begin{pmatrix} \\ R \\ \\ \end{pmatrix} = \begin{array}{c} \begin{array}{cc} N_1 & N_2 \end{array} \\ \begin{array}{c} N_1 \\ N_2 \end{array} \left(\begin{array}{c|c} R_1 & \mathbf{0} \\ \hline \mathbf{0} & R_2 \end{array} \right) \end{array} , \tag{39.9}$$

which we write as $R = R_1 \oplus R_2$. In other words, the representation breaks up into two smaller representations that each only act on one, distinct subspace. If either R_1 or R_2 is reducible then this process can continue until

R has be broken up into a direct sum or irreducible representations. The irreducible representations of a group are the most interesting. Any other representation can be created by taking direct sums of these fundamental elements.

Example

Consider the finite group Z_2. Recall that this has just two elements, the identity e and an object a with $a^2 = e$. The trivial representation R_0 is given by

$$R_0(e) = 1 \ , \qquad R_0(a) = 1 \ . \tag{39.10}$$

A one dimensional, faithful representation R_1 is given by

$$R_1(e) = 1 \ , \qquad R_1(a) = -1 \ . \tag{39.11}$$

A two dimensional, faithful representation R_2 is given by

$$R_2(e) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \ , \qquad R_2(a) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \ , \tag{39.12}$$

Clearly, the one dimensional representations are irreducible; they are already in block diagonal form with only a single block. However, R_2 is reducible. To justify this claim consider the subspace of two dimensional vectors of the form

$$v = \begin{pmatrix} x \\ x \end{pmatrix} \ . \tag{39.13}$$

Note that $R_2(a)v = v$, (and of course $R_2(e)v = v$) so this subspace is left invariant by R_2, meaning it is reducible.

We can reduce R_2 explicitly by making a similarity transform to R_2' with

$$R_2' = S^{-1} R_2 S \ , \tag{39.14}$$

using the transform

$$S = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \ . \tag{39.15}$$

This is determined by following the standard methods for diagonalizing $R_2(a)$. This leads to

$$R_2'(e) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \ , \qquad R_2'(a) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \ , \tag{39.16}$$

In other words,

$$R_2' = \begin{pmatrix} R_0 & 0 \\ 0 & R_1 \end{pmatrix} \ . \tag{39.17}$$

We have shown that R_2 is equivalent to a representation that is a direct sum of R_0 and R_1. So, we may write $R_2 = R_0 \oplus R_1$.

Of importance to continuous groups, it can be proven that all irreducible representations of Abelian groups are one dimensional.

Exercise 39.2 Show that the representation of Abelian Lie group SO(2) defined by

$$R_2(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \tag{39.18}$$

is reducible by diagonalizing the matrix to find

$$R_2'(\theta) = \begin{pmatrix} e^{-i\theta} & 0 \\ 0 & e^{i\theta} \end{pmatrix} . \tag{39.19}$$

This is a block diagonal form of two representations of U(1), which we remember is isomorphic to SO(2).

A particularly interesting set of representations are those that preserve an inner product on the vector space. If the space is real then these are typically orthogonal transformations and if the space is complex they are unitary transformations. We refer to these collectively as unitary operators.

Definition A representation U of the group G in the vector space V is said to be **unitary** if for every $g \in G$

$$U(g)^\dagger = U\left(g^{-1}\right) . \tag{39.20}$$

Since

$$U\left(g^{-1}\right)U(g) = U\left(g^{-1}g\right) = U(e) = I , \tag{39.21}$$

this simply means that U is a unitary operator in V.

Though we do not prove it, all finite irreducible representations of finite groups are equivalent to unitary representations [25]. We are, of course, very interested in Lie groups, which are infinite. The adjoint representation is an important representation that is always defined for Lie groups.

Definition The **adjoint** representation of a Lie group G with algebra $\mathfrak{g}$ is of particular importance in particle physics. It is a mapping from G to $\mathfrak{g}$, or in other words a representation of the Lie group in the vector space of the Lie algebra. For any $A \in G$ and $X \in \mathfrak{g}$ the mapping is

$$\mathrm{Ad}_A(X) = AXA^{-1} . \tag{39.22}$$

This mapping can be understood by remembering that the product

of elements of G is also an element of G, so

$$Ae^{iX}A^{-1} \ , \tag{39.23}$$

is an element of the group. Consequently, it is associated to some element of the Lie algebra $Y \in \mathfrak{g}$ by

$$Ae^{iX}A^{-1} = e^{iY} \ . \tag{39.24}$$

Note that from Eq. (38.7) we know that

$$Ae^{iX}A^{-1} = e^{iAXA^{-1}} \ , \tag{39.25}$$

which implies that $Y = AXA^{-1}$, or that $AXA^{-1} \in \mathfrak{g}$. This means that Ad_A is a mapping $\mathfrak{g} \to \mathfrak{g}$.

We must check that this mapping is a group homomorphism, that is, that it preserves the group multiplication law. Suppose A and B are two elements of G. Then

$$\begin{aligned} \mathrm{Ad}_B(\mathrm{Ad}_A(X)) =& \mathrm{Ad}_B(AXA^{-1}) = BAXA^{-1}B^{-1} = (BA)\,X\,(BA)^{-1} \\ =& \mathrm{Ad}_{BA}(X) \end{aligned} \tag{39.26}$$

which is clearly the BA adjoint map. This confirms that this is indeed a group homomorphism and therefore that this defines a representation of G on $\mathfrak{g}$.

It is probably no surprise that just as Lie groups have various representations, so do Lie algebras. What is more, these representations are related by the exponential relationship between the group and algebra. In short, if R_1 is a representation of G on the vector space V, then there is a representation r_1 of $\mathfrak{g}$ on V with

$$R_1\left(e^{i\theta X}\right) = e^{i\theta r_1(X)} \ . \tag{39.27}$$

The relation between the two representations can be made clearer by noting that

$$-i\,\frac{d}{d\theta}R_1\left(e^{i\theta X}\right)\bigg|_{\theta=0} = r_1\left(X\right) \ . \tag{39.28}$$

While we do not prove it here, it is true that R_1 is irreducible if and only if r_1 is irreducible. Similarly, if two group representations R_1 and R_2 are equivalent the associated Lie algebra representations r_1 and r_2 are also equivalent. These relationships between representations are useful because they allow us to focus on the, often simpler, case of Lie algebra representations. Once these representations have been determined we can immediately determine the Lie group representations by exponentiation.

The representations of a Lie algebra must satisfy the defining features of the algebra, including

$$\Big[r(X^a),\, r(X^b)\Big] = i\, f^{abc} r(X^c)\ . \tag{39.29}$$

Exercise 39.3

The complex conjugate representation of a given group representation (39.3) also leads to a complex conjugate representation of the corresponding Lie algebra. Show that given a representation r of the Lie algebra there is another representation, often denoted $\overline{r}$ with

$$\overline{r}(X) = -r(X)^*\ . \tag{39.30}$$

This additional representation is called the **complex conjugate representation** of r (or sometimes just the conjugate representation). Though the complex conjugate representation always exists, it may or may not be equivalent to another representation.

Example

The adjoint representation of a Lie group is associated with an adjoint representation of the corresponding Lie algebra. The adjoint representation of X is given by

$$e^{i\theta Y} X e^{-i\theta Y} \tag{39.31}$$

for some group element $e^{i\theta Y}$. We then use Eq. (39.28) to find

$$\begin{aligned} -i\frac{d}{d\theta}\left(e^{i\theta Y} X e^{-i\theta Y}\right)\Big|_{\theta=0} =& -i\left(iY e^{i\theta Y} X e^{-i\theta Y} - e^{i\theta Y} X e^{-i\theta Y} iY\right)\Big|_{\theta=0} \\ =& YX - XY = [Y,\, X]\ . \end{aligned} \tag{39.32}$$

Or, we use the notation

$$\mathrm{ad}_Y(X) = [Y,\, X]\ . \tag{39.33}$$

Note how this mapping preserves the commutator.

$$\mathrm{ad}_Z(\mathrm{ad}_Y(X)) = \mathrm{ad}_Z([Y,\, X]) = [Z,\, [Y,\, X]]\ , \tag{39.34}$$

$$\mathrm{ad}_Y(\mathrm{ad}_Z(X)) = \mathrm{ad}_Y([Z,\, X]) = [Y,\, [Z,\, X]]\ . \tag{39.35}$$

From the Jacobi identity (38.31) we have that

$$[Z,\, [Y, X]] - [Y,\, [Z, X]] = [[Z, Y],\, X]\ . \tag{39.36}$$

Therefore, we find

$$\mathrm{ad}_Z(\mathrm{ad}_Y(X)) - \mathrm{ad}_Y(\mathrm{ad}_Z(X)) = \mathrm{ad}_{[Z,Y]}(X)\ . \tag{39.37}$$

Thus, we see that the adjoint representation of the Lie algebra is tied

to the commutator. Remember that the adjoint representation of the group is on the vector space associated with the algebra and hence the dimension of the vector space (which is equal to the number of generators) is the dimension of the representation. So, we should expand X and Y in this vector space

$$X = x_i T^i \ , \qquad Y = y_j T^j \ , \tag{39.38}$$

where the T^i are the chosen basis of the algebra with $i \in [1, N]$ where N is the dimension of the vector space. The adjoint representation leads to

$$[Y,\, X] = y_j\, x_i [T^j,\, T^i] = y_j\, x_i\, i\, f^{jik} T^k \ . \tag{39.39}$$

The output of this commutator can be viewed in a few ways. We can think of it as a vector in the vector space with the kth coefficient given by $iy_j x_i f^{jik}$. But note that for a fixed k that f^{ijk} is an N by N matrix, exactly the size we are expecting for a representation of the Lie algebra in this space. As obtained in the following exercise, the structure constants themselves provide a representation of the Lie algebra.

Exercise 39.4 Apply the Jacobi Identity given in Eq. (38.31) the basis elements T^i, T^j, and T^k of a Lie algebra to show the structure constants satisfy

$$f^{ij\ell} f^{l\ell m} + f^{jk\ell} f^{i\ell m} + f^{ki\ell} f^{j\ell m} = 0 \ . \tag{39.40}$$

Define matrices

$$\left(T^{(k)}\right)^{ij} = -i f^{ijk} \ , \tag{39.41}$$

where the index in parentheses labels the matrix, and show that Eq. (39.40) implies

$$\left[T^{(i)},\, T^{(j)}\right] = i f^{ijk} T^{(k)} \ . \tag{39.42}$$

Remember that the structure constants are completely antisymmetric. In other words, the structure constants furnish a representation of the Lie algebra, which is the adjoint representation.

The adjoint representation is special because the generators of the algebra can be considered both as states in the vector space

$$\left|T^i\right\rangle \ , \tag{39.43}$$

and as operators acting on that vector space

$$T^j \left|T^i\right\rangle \ . \tag{39.44}$$

Of course, this means that a linear combination of basis vectors can be

written as a single state

$$\alpha \left|T^i\right\rangle + \beta \left|T^j\right\rangle = \left|\alpha T^i + \beta T^j\right\rangle \ . \tag{39.45}$$

We can evaluate Eq. (39.44) by using the fact that the structure constants themselves give the matrix elements of the generators. In short, we can rewrite Eq. (39.41) as

$$\left\langle T^i\right| T^k \left|T^j\right\rangle = -if^{ijk} \ . \tag{39.46}$$

This allows us to evaluate

$$\begin{aligned} T^j \left|T^i\right\rangle =& \mathbb{I}\, T^j \left|T^i\right\rangle \\ =& \left|T^k\right\rangle \left\langle T^k\right| T^j \left|T^i\right\rangle \\ =& -if^{kij} \left|T^k\right\rangle \\ =& if^{jik} \left|T^k\right\rangle \\ =& \left|if^{jik}T^k\right\rangle \\ T^j \left|T^i\right\rangle =& \left|\left[T^j,\, T^i\right]\right\rangle \ . \end{aligned} \tag{39.47}$$

This is the meaning of the connection between the adjoint representation and the commutator. The action of an operator on a state is to produce a state given by the commutator of the two elements of the algebra.

Example

We now spend some significant space determining the finite representations of SU(2). This group is essential to much that follows and provides a simple, but nontrivial, example of the concepts of this section. At the same time, it is often taught in quantum mechanics under the heading of "angular momentum" so it is helpful to use the quantum mechanical notation.

Our first step is to consider representations of $\mathfrak{su}(2)$, with the understanding that once these have been determined the representations of SU(2) immediately follow. As outlined above, we choose the basis of $\mathfrak{su}(2)$ to be $\{\tau^1, \tau^2, \tau^3\}$ where

$$\left[\tau^a,\, \tau^b\right] = i\varepsilon^{abc}\tau^c \ . \tag{39.48}$$

It is this structure that is preserved by *every* representation of $\mathfrak{su}(2)$. That is, by definition of being a representation, any r must satisfy

$$\left[r\left(\tau^a\right),\, r\left(\tau^b\right)\right] = i\varepsilon^{abc} r\left(\tau^c\right) \ . \tag{39.49}$$

Because this is true for all representations, we drop the explicit use of

$r(\tau^a)$ but the results should be understood as applying to any particular representation.

We begin by defining an operator on the space

$$(\tau)^2 = \tau^a \tau^a = \tau^1 \tau^1 + \tau^2 \tau^2 + \tau^3 \tau^3 \ . \tag{39.50}$$

We also define the operators

$$\tau^\pm = \tau^1 \pm i\tau^2 \ . \tag{39.51}$$

Exercise 39.5

Show that

$$\left[(\tau)^2, \tau^a\right] = 0 \ . \tag{39.52}$$

This means that the $(\tau)^2$ operator can be diagonalized by the same basis as any chosen τ^a. Next, show that

$$\left[\tau^3, \tau^\pm\right] = \pm\tau^\pm \ , \tag{39.53}$$

and

$$\left[\tau^+, \tau^-\right] = 2\tau^3 \ . \tag{39.54}$$

Example

Representations of $\mathfrak{su}(2)$ continued: We choose a basis in which $(\tau)^2$ and τ^3 are diagonal. This means we can choose a basis for (at least part of) the vector space V composed of the eigenvectors of $(\tau)^2$ and τ^3. We now employ the notation of quantum mechanics for this vector space. A vector v and its dual $v^\dagger$ are denoted by

$$v = |v\rangle \ , \qquad v^\dagger = \langle v| \ . \tag{39.55}$$

The inner product of two vectors v and u is given by

$$\langle v|u\rangle \ . \tag{39.56}$$

It is convenient to label these vectors by their eigenvalues under $(\tau)^2$ and τ^3. We begin by labeling the vectors as $|\lambda, m\rangle$ with

$$\tau^3|\lambda, m\rangle = m|\lambda, m\rangle \ , \qquad (\tau)^2|\lambda, m\rangle = \lambda|\lambda, m\rangle \ . \tag{39.57}$$

We choose these vectors to be normalized, that is

$$\langle \lambda, m|\lambda, m\rangle = 1 \ . \tag{39.58}$$

Remember that $\tau^\pm|\lambda, m\rangle$ is a vector in V because $\tau^\pm$ is an operator on V. We discover that these vectors are also eigenvectors of τ^3. By using

the commutation relation in Eq. (39.53) we find

$$\begin{aligned}\tau^3\left(\tau^\pm|\lambda,\, m\rangle\right) &= \left(\tau^\pm\tau^3 \pm \tau^\pm\right)|\lambda,\, m\rangle \\ &= \tau^\pm\left(\tau^3 \pm 1\right)|\lambda,\, m\rangle \\ &= \tau^\pm(m\pm 1)|\lambda,\, m\rangle \\ &= (m\pm 1)\left(\tau^\pm|\lambda,\, m\rangle\right)\ . \end{aligned} \tag{39.59}$$

This shows that $\tau^\pm|\lambda,\, m\rangle$ is an eigenvector of τ^3 with eigenvalue $m \pm 1$. In other words, either

$$\tau^\pm|m\rangle \propto |\lambda,\, m\pm 1\rangle\ , \tag{39.60}$$

or

$$\tau^\pm|m\rangle = |0\rangle\ , \tag{39.61}$$

which is the zero vector.

Exercise 39.6 Show that

$$(\tau)^2 - \left(\tau^3\right)^2 = \frac{1}{2}\left[\tau^+\left(\tau^+\right)^\dagger + \left(\tau^+\right)^\dagger\tau^+\right]\ . \tag{39.62}$$

Remember that the τ^a are Hermitian, or that $(\tau^a)^\dagger = \tau^a$. Also show that

$$\tau^\pm\tau^\mp = (\tau)^2 - \left(\tau^3\right)^2 \pm \tau^3\ . \tag{39.63}$$

Example **Representations of $\mathfrak{su}(2)$ continued:** Remember that in an inner product space for any vector $|v\rangle$ we have $\langle v|v\rangle \geq 0$, with equality only if v is the zero vector. This means that a vector of the form $O|v\rangle$, where O is an operator will satisfy

$$\langle v|O^\dagger O|v\rangle \geq 0\ . \tag{39.64}$$

Using Eq. (39.62) this implies that

$$\begin{aligned}\langle\lambda,\, m|\left[(\tau)^2 - \left(\tau^3\right)^2\right]|\lambda,\, m\rangle &\geq 0 \\ \left(\lambda - m^2\right)\langle\lambda,\, m|\lambda,\, m\rangle &\geq 0\ . \end{aligned} \tag{39.65}$$

Because $\langle\lambda,\, m|\lambda,\, m\rangle \geq 0$ this implies that

$$\lambda \geq m^2\ . \tag{39.66}$$

This constraint is very important when combined with action of $\tau^\pm$ on the $|\lambda,\, m\rangle$ states. In Eq. (39.60) we saw that $\tau^\pm$ either raise or lower

the value of m. This is why they are typically referred to as **raising operators** or **lowering operators**. However, Eq. (39.66) implies that the process of raising or lowering cannot continue indefinitely. For a given λ, there must be some state $|\lambda,\, m_{\max}\rangle$ where $(m_{\max}+1)^2 > \lambda$. The only way all the results we have found so far can hold is if

$$\tau^+|\lambda,\, m_{\max}\rangle = |0\rangle \ . \tag{39.67}$$

Any operator acting on the zero vector equals the zero vector so it follows that

$$\begin{aligned} \tau^-\tau^+|\lambda,\, m_{\max}\rangle &= |0\rangle \\ \left[(\tau)^2 - \left(\tau^3\right)^2 - \tau^3\right]|\lambda,\, m_{\max}\rangle &= |0\rangle \\ \left(\lambda - m_{\max}^2 - m_{\max}\right)|\lambda,\, m_{\max}\rangle &= |0\rangle \ , \end{aligned} \tag{39.68}$$

where we used Eq.(39.63) in the second line. This implies that

$$\lambda = m_{\max}\left(m_{\max}+1\right) \ . \tag{39.69}$$

Exercise 39.7 Using a method similar to what is done above, show that

$$\lambda = m_{\min}\left(m_{\min}-1\right) \ . \tag{39.70}$$

Example **Representations of $\mathfrak{su}(2)$ continued:** Equations (39.69) and (39.70) imply that either $m_{\min} = m_{\max}+1$ (which violates the meanings of maximum and minimum) or that

$$m_{\min} = -m_{\max} \ . \tag{39.71}$$

Finally, the process of raising m or lowering m changes the value by one. This means that the number of raising or lowering steps to go from $m_{\max}$ to $m_{\min}$ must be an integer N. Therefore, we find that

$$m_{\max} - m_{\min} = 2m_{\max} = N \ . \tag{39.72}$$

This implies that for every representation of $\mathfrak{su}(2)$ we have the constraint

$$m_{\max} = \frac{N}{2} \ . \tag{39.73}$$

This last relation is what differentiates representations of $\mathfrak{su}(2)$, which are labeled by $m_{\max}$. In quantum mechanics the label ℓ is used instead

of $m_{\max}$ and states are labeled as $|\ell,\, m\rangle$. That is,

$$\tau^3|\ell,\, m\rangle = m|\ell,\, m\rangle \ , \qquad (\tau)^2\,|\ell,\, m\rangle = \ell(\ell+1)|\ell,\, m\rangle \ . \tag{39.74}$$

The representations might be denoted r_ℓ, but are usually simply referred to by ℓ.

We can also determine the action of the raising and lowering operators. In general we must have

$$\tau^+|\ell,\, m\rangle = c_+(\ell,\, m)|\ell,\, m+1\rangle \ , \tag{39.75}$$

with c_+ a potentially complex number. This implies that

$$\langle \ell,\, m+1|\tau^+|\ell,\, m\rangle = c_+(\ell,\, m) \ . \tag{39.76}$$

Taking the Hermitian conjugate of this equation we find

$$\langle \ell,\, m|\tau^-|\ell,\, m+1\rangle = c_+^*(\ell,\, m) \ , \tag{39.77}$$

which immediately leads to

$$\langle \ell,\, m|\tau^-\tau^+|\ell,\, m\rangle = |c_+(\ell,\, m)|^2 \ . \tag{39.78}$$

Exercise 39.8 Use Eqs. (39.63) and (39.78) to show that

$$|c_+(\ell,\, m)|^2 = \ell(\ell+1) - m(m+1) \ . \tag{39.79}$$

Conclude that

$$\tau^+|\ell,\, m\rangle = \sqrt{\ell(\ell+1) - m(m+1)}\,|\ell,\, m+1\rangle \ , \tag{39.80}$$

$$\tau^-|\ell,\, m\rangle = \sqrt{\ell(\ell+1) - m(m-1)}\,|\ell,\, m-1\rangle \ . \tag{39.81}$$

Notice that $\tau^+|\ell,\, \ell\rangle = 0$ and $\tau^-|\ell,\, -\ell\rangle = 0$ as expected.

Example **Representations of $\mathfrak{su}(2)$ continued:** Notice that each representation has a different dimension, as can be seen from the number of eigenvectors of τ^3:

$$\ell = 0 \qquad |0,\, 0\rangle \tag{39.82}$$

$$\ell = \tfrac{1}{2} \qquad |\tfrac{1}{2},\, -\tfrac{1}{2}\rangle \ , \quad |\tfrac{1}{2},\, \tfrac{1}{2}\rangle \tag{39.83}$$

$$\ell = 1 \qquad |1,\, -1\rangle \ , \quad |1,\, 0\rangle \ , \quad |1,\, 1\rangle \tag{39.84}$$

That is, the $\ell = 0$ representation is one-dimensional, the $\ell = \frac{1}{2}$ representation is two-dimensional, and so on. In general the dimension of

the representation is $2\ell + 1$ and some times this is used to refer to a representation, for example

$$\ell = 0 \to \mathbf{1} \ , \quad \ell = \tfrac{1}{2} \to \mathbf{2} \ , \quad \ell = 1 \to \mathbf{3} \ . \tag{39.85}$$

Though we have not proved it, these are all the inequivalent, irreducible representations of $\mathfrak{su}(2)$. They generate the representations of SU(2) by exponentiation.

But how do we get from the above results to actual matrix representations of the Lie algebra? Remember that we have chosen the vectors above to be eigenvectors of τ^3. They can also serve as a basis for vector space that the Lie algebra is defined on. As we discuss in Sec. 33 these eigenvectors are orthogonal, so we can use them to define an orthonormal basis. In such a basis we can find the element of a matrix M_{ij} from

$$M_{ij} = \langle i|M|j\rangle \ , \tag{39.86}$$

where $|i\rangle$ is the ith basis vector. In the case of the **1** representation there is only one basis element. Note that

$$\langle 0,\, 0|\tau^3|0,\, 0\rangle = 0 \ . \tag{39.87}$$

Since

$$\tau^1 = \frac{1}{2}\left(\tau^+ + \tau^-\right) \ , \quad \tau^2 = -\frac{i}{2}\left(\tau^+ - \tau^-\right) \ , \tag{39.88}$$

we also find that

$$\langle 0,\, 0|\tau^1|0,\, 0\rangle = 0 \ , \quad \langle 0,\, 0|\tau^2|0,\, 0\rangle = 0 \ . \tag{39.89}$$

Thus, in the **1** representation the basis of the Lie algebra is simply composed of the zero vector. The exponentiation of zero matrices gives the identity matrix. Thus, every group element is mapped to the identity matrix. This is the trivial representation, sometimes called the singlet representation.

Exercise 39.9 Using the method given above, show that the **2** representation of the algebra has

$$\tau^1 = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \ , \quad \tau^2 = \frac{1}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \ , \quad \tau^3 = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \ . \tag{39.90}$$

The **2** is sometimes called the **fundamental representation** as it is the smallest dimension, faithful representation. In quantum mechanics this is often called the spin-1/2 representation.

Example

The fundamental representation **2** of $\mathfrak{su}(2)$ is **pseudoreal**. We can form the conjugate representation $\overline{\mathbf{2}}$ from the fundamental representation (see Eq. (39.30)) and find

$$\tau^1 = -\frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \ , \quad \tau^2 = \frac{1}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \ , \quad \tau^3 = -\frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \ . \tag{39.91}$$

This representation is equivalent to the **2** if the generators can be related by a similarity transform

$$\tau^a_{\overline{\mathbf{2}}} = U^\dagger \tau^a_{\mathbf{2}} U \ , \tag{39.92}$$

where U is a unitary matrix. (The similarity transformation employs a unitary matrix in this case because both representations are already unitary.) This leads to a similarity transform between the group representations. If

$$R = e^{i\alpha^a \tau^a} \ , \qquad R^* = e^{i\alpha^a \overline{\tau}^a} \ , \tag{39.93}$$

then we have

$$\begin{aligned} R^* =& e^{i\alpha^a \overline{\tau}^a} \\ =& e^{i\alpha^a U^\dagger \tau^a U} \\ =& U^\dagger e^{i\alpha^a \tau^a} U \\ =& U^\dagger R U \ . \end{aligned} \tag{39.94}$$

Exercise 39.10

Show that taking

$$U = \sigma^2 \ , \tag{39.95}$$

satisfies Eq. (39.92). This shows that the **2** and $\overline{\mathbf{2}}$ representations of $\mathfrak{su}(2)$ are equivalent. This not generally the case for fundamental representations of $\mathfrak{su}(N)$ algebras.

Example

The fundamental representation **2** of $\mathfrak{su}(2)$ is **pseudoreal continued:** The previous exercise shows that the conjugate representation is equivalent to the fundamental representation. This means the **2** representation is either **real** or **pseudoreal**. The two cases are distinguished by the combinations of vectors that can be made invariant.

We showed in Eq. (39.6) that the inner product of two vectors is preserved by unitary representations of a group. That is

$$u^{*a} g_{ab} v^b \ , \tag{39.96}$$

is invariant. In a real representaiton

$$u^a S_{ab} v^b \ , \tag{39.97}$$

is also invariant, where $S_{ab} = S_{ba}$ is a symmetric matrix. In a pseudo-real representation

$$u^a A_{ab} v^b \ , \tag{39.98}$$

is invariant, where $A_{ab} = -A_{ba}$ is an antisymmetric matrix.

Exercise 39.11 Show that for vectors u^a and v^a that transform under the fundamental representation of SU(2)

$$u^a \varepsilon_{ab} v^b \ , \tag{39.99}$$

where ε_{ab} is the Levi-Civita tensor in two dimensions, is invariant under fundamental representations of SU(2) transformations. This shows that the **2** representation is pseudo-real.

Exercise 39.12 Using the same method used for the **2** representation, show that the **3** representation of the algebra has

$$\tau^1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \ , \quad \tau^2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{pmatrix} \ , \quad \tau^3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} \ . \tag{39.100}$$

In quantum mechanics this is referred to as the spin-1 representation.

Exercise 39.13 Recall that the Lie algebra for $\mathfrak{so}(3)$ is identical to $\mathfrak{su}(2)$. Take the basis for $\mathfrak{so}(3)$ given in Eq. (38.50) and diagonalize T^3. Transform the Lie algebra into the basis with diagonal T^3 and show that it matches the spin-1 basis of $\mathfrak{su}(2)$. This a manifestation of the fact that every integer ℓ representation of $\mathfrak{su}(2)$ is a representation of $\mathfrak{so}(3)$.

Exercise 39.14 Find the generators of the representation conjugate to the **3** representation. Show that the two representations are equivalent and that **3** is a real representation with S_{ab} simply the metric of three dimensional Euclidean space. (Hint, use the results of the previous exercise.)

Example The **3** representation is also the adjoint representation of the Lie algebra obtained in Eq. (39.33). This can be seen by using the formula in

Eq. (39.41):

$$T^1 = -i\varepsilon^{ij1} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} \tag{39.101}$$

$$T^2 = -i\varepsilon^{ij2} = \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix} \tag{39.102}$$

$$T^3 = -i\varepsilon^{ij3} = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{39.103}$$

Notice that these are exactly the generators from Eq. (38.50). From an above exercise we know these are equivalent to the **3** representation of $\mathfrak{su}(2)$.

The final topic of this section is the combination of representations. For instance, in a quantum mechanical system of two electrons, each of which is in the spin-1/2 (or **2**) representation of $\mathfrak{su}(2)$, the total system is in the $\mathbf{2}\otimes\mathbf{2}$ representation. How does this product of representations relate to the categorization of representations that we just worked through?

First, we need to understand how representations and tensor products work together. For two Lie *groups* G_1 and G_2 with representations R_1 and R_2 the tensor product works as you might expect. There is a representation $R_1 \otimes R_2$ of the product group $G_1 \otimes G_2$ defined by

$$R_1 \otimes R_2\,(g_1 \otimes g_2) = R_1(g_1) \otimes R_2(g_2)\ . \tag{39.104}$$

Consider what this means for the representations of the Lie algebra of this product group. Effectively, the exponentiation turns the tensor product into a sum, the direct sum. In particular, if $\mathfrak{g}_1$ and $\mathfrak{g}_2$ are the Lie algebras associated associated with G_1 and G_2 (respectively), then the Lie algebra of $G_1 \otimes G_2$ is

$$\mathfrak{g}_1 \oplus \mathfrak{g}_2\ . \tag{39.105}$$

We still speak of the tensor product $r_1 \otimes r_2$ of Lie algebra representations, but because the algebra space is a direct sum of the two algebras the correct composition rule, for $X \in \mathfrak{g}_1$ and $Y \in \mathfrak{g}_2$, turns out to be

$$(r_1 \otimes r_2)\,(X \otimes Y) = r_1(X) \otimes I_2 + I_1 \otimes r_2(Y)\ , \tag{39.106}$$

where I_1 and I_2 are the identity elements of their respective vector spaces.

While we have used language that allows for two distinct groups G_1 and G_2 we can also consider two distinct representations, R_1 and R_2, of a single group G on two distinct vector spaces V_1 and V_2. We then have

$R_1 \otimes R_2$ as a representation of G on $V_1 \otimes V_2$ with

$$R_1 \otimes R_2(g) = R_1(g) \otimes R_2(g) \ , \tag{39.107}$$

for all $g \in G$. The associated rule for elements $X \in \mathfrak{g}$ of the Lie algebra of G is

$$r_1 \otimes r_2(X) = r_1(X) \otimes I + I \otimes r_2(X) \ . \tag{39.108}$$

Notice that this form requires that we be careful with products of operators acting on the total $\mathfrak{g} \oplus \mathfrak{g}$ vector space. As an important example, consider $(\tau)^2$.

Example

For any individual Lie algebra generator τ^a we have product operator

$$(\tau \otimes \tau)^a \ , \tag{39.109}$$

the index and parentheses are simply indicating that the index is the same for both operators. Then we define

$$(\tau)^2_T = (\tau \otimes \tau)^a \, (\tau \otimes \tau)^a \ , \tag{39.110}$$

where we are summing over the repeated index.

Exercise 39.15

Show that

$$\begin{aligned}(\tau)^2_T \left(|1\rangle \otimes |2\rangle\right) &= (\tau)^2 |1\rangle \otimes |2\rangle + |1\rangle \otimes (\tau)^2 |2\rangle + 2\tau^a |1\rangle \otimes \tau^a |2\rangle \\ &= (\tau)^2 |1\rangle \otimes |2\rangle + |1\rangle \otimes (\tau)^2 |2\rangle + 2\tau^3 |1\rangle \otimes \tau^3 |2\rangle \\ &\quad + \tau^+ |1\rangle \otimes \tau^- |2\rangle + \tau^- |1\rangle \otimes \tau^+ |2\rangle \ . \end{aligned} \tag{39.111}$$

Example

$\mathbf{2} \otimes \mathbf{2}$: We can now consider the tensor product of the spin-1/2 representations of $\mathfrak{su}(2)$. The basis for each two dimensional $\mathbf{2}$ is

$$|\tfrac{1}{2}, -\tfrac{1}{2}\rangle \ , \quad |\tfrac{1}{2}, \tfrac{1}{2}\rangle \ . \tag{39.112}$$

The total basis is has four elements

$$\begin{aligned} &|\tfrac{1}{2}, -\tfrac{1}{2}\rangle \otimes |\tfrac{1}{2}, -\tfrac{1}{2}\rangle \ , \quad |\tfrac{1}{2}, -\tfrac{1}{2}\rangle \otimes |\tfrac{1}{2}, \tfrac{1}{2}\rangle \ , \\ &|\tfrac{1}{2}, \tfrac{1}{2}\rangle \otimes |\tfrac{1}{2}, -\tfrac{1}{2}\rangle \ , \quad |\tfrac{1}{2}, \tfrac{1}{2}\rangle \otimes |\tfrac{1}{2}, \tfrac{1}{2}\rangle \ . \end{aligned} \tag{39.113}$$

The operators on the product space are

$$\tau^3_T = \tau^3 \otimes \tau^3 \ , \quad \tau^\pm_T = \tau^\pm \otimes \tau^\pm \ , \tag{39.114}$$

along with $(\tau)^2_T$, which is defined in Eq. (39.111) Using Eq. (39.106) we can evaluate the action of these operators on the basis set. For instance,

we find

$$\begin{aligned}\tau_T^3\left(|\tfrac{1}{2},\,\tfrac{1}{2}\rangle\otimes|\tfrac{1}{2},\,\tfrac{1}{2}\rangle\right) =&\tau^3|\tfrac{1}{2},\,\tfrac{1}{2}\rangle\otimes|\tfrac{1}{2},\,\tfrac{1}{2}\rangle+|\tfrac{1}{2},\,\tfrac{1}{2}\rangle\otimes\tau^3|\tfrac{1}{2},\,\tfrac{1}{2}\rangle\\ =&\frac{1}{2}|\tfrac{1}{2},\,\tfrac{1}{2}\rangle\otimes|\tfrac{1}{2},\,\tfrac{1}{2}\rangle+\frac{1}{2}|\tfrac{1}{2},\,\tfrac{1}{2}\rangle\otimes|\tfrac{1}{2},\,\tfrac{1}{2}\rangle\\ =&1\left(|\tfrac{1}{2},\,\tfrac{1}{2}\rangle\otimes|\tfrac{1}{2},\,\tfrac{1}{2}\rangle\right)\ .\end{aligned}\tag{39.115}$$

So, this product state is an eigenvector of the product operator.

Exercise 39.16 Show that the other basis elements are also eigenvectors of τ_T^3 with eigenvalues of 0, 0, and -1. Determine which basis vectors are also eigenvectors of $(\tau_T)^2$ and find the eigenvalues.

Example **$\mathbf{2}\otimes\mathbf{2}$ continued:** The above exercise shows that the basis vectors are eigenvectors of τ_T^3, but those with $m=0$ are not eigenvectors of $(\tau_T)^2$. To determine these we can act with the lowering operator on the $m=1$ state. (We could also use the raising operator on the $m=-1$ state.) We find

$$\begin{aligned}\tau_T^-\left(|\tfrac{1}{2},\,\tfrac{1}{2}\rangle\otimes|\tfrac{1}{2},\,\tfrac{1}{2}\rangle\right) =&\tau^-|\tfrac{1}{2},\,\tfrac{1}{2}\rangle\otimes|\tfrac{1}{2},\,\tfrac{1}{2}\rangle+|\tfrac{1}{2},\,\tfrac{1}{2}\rangle\otimes\tau^-|\tfrac{1}{2},\,\tfrac{1}{2}\rangle\\ =&|\tfrac{1}{2},\,-\tfrac{1}{2}\rangle\otimes|\tfrac{1}{2},\,\tfrac{1}{2}\rangle+|\tfrac{1}{2},\,\tfrac{1}{2}\rangle\otimes|\tfrac{1}{2},\,-\tfrac{1}{2}\rangle\ .\end{aligned}\tag{39.116}$$

Now, the initial state had $\ell=1$ and $m=1$. This means that the lowering operator, from Eq. (39.81), produces the $\ell=1$ and $m=0$ state multiplied by $\sqrt{2}$. So, the normalized basis vector is

$$\frac{1}{\sqrt{2}}\left(|\tfrac{1}{2},\,-\tfrac{1}{2}\rangle\otimes|\tfrac{1}{2},\,\tfrac{1}{2}\rangle+|\tfrac{1}{2},\,\tfrac{1}{2}\rangle\otimes|\tfrac{1}{2},\,-\tfrac{1}{2}\rangle\right)\ .\tag{39.117}$$

Exercise 39.17 Act on the state given in Eq. (39.117) with the $(\tau_T)^2$ operator to confirm it is an eigenvector with eigenvalue 1. Next, Use the lowering operator on the state and show that it produces

$$|\tfrac{1}{2},\,-\tfrac{1}{2}\rangle\otimes|\tfrac{1}{2},\,-\tfrac{1}{2}\rangle\ .\tag{39.118}$$

Example **$\mathbf{2}\otimes\mathbf{2}$ continued:** We have now shown that we can choose a basis in which three of the basis elements are eigenvectors of the product space τ_T^3 and $(\tau_T)^2$ operators. These vectors span an $\ell=1$ representation. The last element of the basis (if we want it to be an orthonormal basis) must be

$$\frac{1}{\sqrt{2}}\left(|\tfrac{1}{2},\,-\tfrac{1}{2}\rangle\otimes|\tfrac{1}{2},\,\tfrac{1}{2}\rangle-|\tfrac{1}{2},\,\tfrac{1}{2}\rangle\otimes|\tfrac{1}{2},\,-\tfrac{1}{2}\rangle\right)\ .\tag{39.119}$$

This is also an eigenstate of both τ_T^3 and $(\tau_T)^2$. This means the complete orthonormal basis is

$$\left\{ \frac{1}{\sqrt{2}} \left(|\tfrac{1}{2}, -\tfrac{1}{2}\rangle \otimes |\tfrac{1}{2}, \tfrac{1}{2}\rangle - |\tfrac{1}{2}, \tfrac{1}{2}\rangle \otimes |\tfrac{1}{2}, -\tfrac{1}{2}\rangle \right), |\tfrac{1}{2}, \tfrac{1}{2}\rangle \otimes |\tfrac{1}{2}, \tfrac{1}{2}\rangle \right.$$
$$\left. \frac{1}{\sqrt{2}} \left(|\tfrac{1}{2}, -\tfrac{1}{2}\rangle \otimes |\tfrac{1}{2}, \tfrac{1}{2}\rangle + |\tfrac{1}{2}, \tfrac{1}{2}\rangle \otimes |\tfrac{1}{2}, -\tfrac{1}{2}\rangle \right), |\tfrac{1}{2}, -\tfrac{1}{2}\rangle \otimes |\tfrac{1}{2}, -\tfrac{1}{2}\rangle \right\} \tag{39.120}$$

Exercise 39.18 Show that the state given in Eq. (39.119) is normalized (its inner product with itself is one) and orthogonal to the other basis vectors. Show that it is also an eigenstate of τ_T^3 and $(\tau_T)^2$ with eigenvalue 0 and that the actions of raising and lowering operators produce the zero vector.

Exercise 39.19 Using the basis given in Eq. (39.120) show that the operators take the form

$$\tau_T^3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \qquad (\tau_T)^2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{39.121}$$

and

$$\tau_T^1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \qquad \tau_T^2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -i & 0 \\ 0 & i & 0 & -i \\ 0 & 0 & i & 0 \end{pmatrix}. \tag{39.122}$$

Example **$\mathbf{2} \otimes \mathbf{2}$ continued:** The results of the previous exercises show that the product space is a representation of $\mathfrak{su}(2)$. However, it is a reducible one. We can make a basis transformation to an equivalent basis in which representation separates into a block diagonal form. The upper-left corner is the trivial representation $\mathbf{1}$ and the bottom-right, three-by-three bock is the $\ell = 1$ or $\mathbf{3}$ representation. This is typically expressed as

$$\mathbf{2} \otimes \mathbf{2} = \mathbf{1} \oplus \mathbf{3} \,. \tag{39.123}$$

Note that the notation is helpful in that $2 \times 2 = 1 + 3$, which is simply a statement about the dimension of the two spaces.

While we do not prove it, this is a specific example of a more general result. The tensor product of an ℓ_1 and ℓ_2 representation, which have dimensions $2\ell_1+1$ and $2\ell_2+1$, respectively, can be put in a block diagonal form. The blocks begin with $\ell = |\ell_1 - \ell_2|$ and increase by one until they reach $\ell_1+\ell_2$. For instance, the tensor product of two $\ell = 1$ representations

is

$$\mathbf{3} \otimes \mathbf{3} = \mathbf{1} \oplus \mathbf{3} \oplus \mathbf{5} \tag{39.124}$$

which includes $\ell = 0$, $\ell = 1$, and $\ell = 2$ blocks.

The ideas described here are used frequently throughout what follows. They are often used in courses on quantum mechanics as well, though they are not always discussed in this language. I hope that this discussion will make such textbook treatments a little less mysterious.

SUBSECTION 39.1

Representations of SU(3)

A slightly more interesting group is SU(3). It provides a calculable demonstration of a more general method for determining the representations of a Lie group. It is also of physical significance as the strong nuclear force is tied to an SU(3) symmetry of nature and the quarks and gluons transform in particular representations of SU(3).

The associated Lie algebra $\mathfrak{su}(3)$ has eight generators. These are often expressed in terms of the Gell-Mann matrices

$$\lambda^1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} , \qquad \lambda^2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} , \qquad \lambda^3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} ,$$
$$\lambda^4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} , \qquad \lambda^5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix} , \qquad \lambda^6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} ,$$
$$\lambda^7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} , \qquad \lambda^8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix} . \tag{39.125}$$

The actual generators of the algebra are

$$T^a = \frac{1}{2}\lambda^a , \tag{39.126}$$

similar to Pauli matrices for SU(2). The prefactors are chosen so that

$$\mathrm{Tr}\left[T^a T^b\right] = \frac{1}{2}\delta^{ab} . \tag{39.127}$$

This normalization is referred to as the **Killing form** of the generators. Note that this normalization is chosen only for this particular representation. For a general representation

$$\mathrm{Tr}\left[r(T^a) r(T^b)\right] = T(r)\delta^{ab} , \tag{39.128}$$

where $T(r)$ is the **index** of the representation. Note that if we sum over the algebra indices a and b we find

$$\mathrm{Tr}\,[r(T^a)r(T^a)] = Tr(r) \times \dim(Ad)\ , \tag{39.129}$$

where $\dim(Ad)$ is the number of basis elements in the Lie algebra (8 for SU(3)) which is also equal to the dimension of the adjoint representation.

Exercise 39.20 Determine the structure constants of $\mathfrak{su}(3)$ for these generators. (Feel free to use your favorite software.) You should find

$$\begin{aligned} &f^{123} = 1\ , \quad f^{147} = \frac{1}{2}\ , \quad f^{165} = \frac{1}{2}\ , \quad f^{246} = \frac{1}{2}\ , \\ &f^{257} = \frac{1}{2}\ , \quad f^{345} = \frac{1}{2}\ , \quad f^{376} = \frac{1}{2}\ , \quad f^{458} = \frac{\sqrt{3}}{2}\ , \quad f^{678} = \frac{\sqrt{3}}{2}\ . \end{aligned} \tag{39.130}$$

All other nonzero structure constants are related to these by the complete antisymmetry of f^{abc}.

Exercise 39.21 Use the structure constants above to show that for SU(3) where we have chosen the Killing form in the defining representation that

$$f^{abc}f^{abc} = 24\ . \tag{39.131}$$

Note that a sum is implied over every repeated index. Note also that the complete antisymmetry of the structure constants implies that many terms in these sums are the same.

Note that the first three generators produce an $\mathfrak{su}(2)$ subalgebra

$$\left[T^a,\, T^b\right] = i\varepsilon^{abc}T^c\ . \tag{39.132}$$

Exercise 39.22 Show that

$$\left\{T^4,\, T^5,\, \frac{1}{2}T^3 + \frac{\sqrt{3}}{2}T^8\right\}\ , \tag{39.133}$$

and

$$\left\{T^6,\, T^7,\, -\frac{1}{2}T^3 + \frac{\sqrt{3}}{2}T^8\right\}\ , \tag{39.134}$$

are each $\mathfrak{su}(2)$ subalgebras.

One convenient aspect of the Gell-Mann matrices is that both λ^3 and λ^8 are diagonal, and consequently they commute (their commutator vanishes). The above exercises show that T^3 and T^8 do not commute with any of T^4, T^5, T^6, and T^7. It is true that T^8 commutes with T^1 and T^2. However, these matrices do not commute with T^3 or with each other. Consequently,

we have at most only two mutually commuting generators.

The algebra spanned by the largest group of commuting generators of a given Lie algebra is called the **Cartan subalgebra**. Because the Cartan subalgebra is composed of commuting generators, in a given representation of the Lie algebra the matrices that correspond to these generators also commute. This means we can choose a basis for the vector space this representation acts on in which each element of the Cartan subalgebra is diagonal. (The Gell-Mann matrices have already done this for the fundamental representation of SU(3).) The basis vectors in this representation are all eigenvectors of each of the generators that make up the Cartan subalgebra.

In other words, suppose $|\psi\rangle$ is one of the basis states of some representation of SU(3). Then we have

$$T^3|\psi\rangle = \mu_1|\psi\rangle \ , \quad T^8|\psi\rangle = \mu_2|\psi\rangle \ , \tag{39.135}$$

where the eigenvalues μ_1 and μ_2 are real numbers because T^3 and T^8 are represented by Hermitian matrices. These eigenvalues are called the **weights** of the state. In the general case when there are n elements of the Cartan subalgebra then the weights can be written as an n-dimensional vector, called the **weight vector** of the state.

Example

The trivial representation of SU(3) simply maps every generator to a matrix of zeros. For any basis element v we then have

$$T^3 v = 0 \ , \quad T^8 v = 0 \ . \tag{39.136}$$

In this case the weight vector for each basis state is (0, 0).

Exercise 39.23

The Gell-Mann matrices correspond to the fundamental representation of SU(3). The basis states are

$$f_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \ , \quad f_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \ , \quad f_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \ . \tag{39.137}$$

Show that the weight vectors for these states are

$$f_1 : \left(\frac{1}{2}, \frac{1}{2\sqrt{3}}\right) \ , \quad f_2 : \left(-\frac{1}{2}, \frac{1}{2\sqrt{3}}\right) \ , \quad f_3 : \left(0, -\frac{1}{\sqrt{3}}\right) \ . \tag{39.138}$$

Exercise 39.24

Recall that the complex conjugate representation of the fundamental representation is given by Eq. (39.30). Show that the weight vectors for this

representation are

$$f_1 : \left(-\frac{1}{2}, -\frac{1}{2\sqrt{3}}\right) , \quad f_2 : \left(\frac{1}{2}, -\frac{1}{2\sqrt{3}}\right) , \quad f_3 : \left(0, \frac{1}{\sqrt{3}}\right) . \tag{39.139}$$

The fact that these are different weight vectors can be used to show that this representation is distinct from the fundamental representation, even though both are three dimensional.

Definition

There is another special operator made from the group generators that commutes with the Cartan subalgebra. This called the **Casimir operator** and is defined by

$$(T)^2 \equiv T^a T^a , \tag{39.140}$$

with the usual implied sum over the gauge index.

Exercise 39.25

Show that for any Lie algebra

$$\left[T^a, T^b\right] = i f^{abc} T^c , \tag{39.141}$$

with completely antisymmetric structure constants f^{abc} that

$$\left[(T)^2, T^a\right] = 0 . \tag{39.142}$$

The previous exercise implies that the Casimir operator commutes with every generator, including those that comprise the Cartan subalgebra. This means that it can be diagonalized along with the Cartan subalgebra generators. Also, due to a result called **Schur's lemma** the Casimir operator must be proportional to the identity operator in any representation.

$$(T)^2 = C(r)\, I . \tag{39.143}$$

The proportionality constant $C(r)$ depends on the representation. By taking the trace of this equation we find

$$\mathrm{Tr}\left[T^a T^a\right] = C(r) \times \dim(r) , \tag{39.144}$$

where $\dim(r)$ is the dimension of the representation in question. If we compare this to the index equation in Eq. (39.129) we find

$$C(r) \times \dim(r) = T(r) \times \dim(Ad) . \tag{39.145}$$

Exercise 39.26

Use Eq. (39.145) for the defining representation given by the Gell-Mann matrices to find

$$C(r) = \frac{4}{3} , \tag{39.146}$$

for this representation.

Exercise 39.27 Note that Eq. (39.145) implies that in the adjoint representation that $T(r)$ is equal to $C(r)$. We can evaluate the index in the adjoint representation by making use of Eq. (39.47)

$$\begin{aligned} C(Ad) \times \dim(Ad) &= \mathrm{Tr}\,[T^a T^a] = \langle T^b | T^a T^a | T^b \rangle \\ &= \Big\langle \Big[T^a,\, T^b\Big] \Big| \Big[T^a,\, T^b\Big] \Big\rangle = f^{abc} f^{abc} \ . \end{aligned} \tag{39.147}$$

Supply all the missing steps in this sequence of equalities. Then, using Eq. (39.131) conclude that

$$C(Ad) = 3 \ . \tag{39.148}$$

Another special representation is the adjoint representation, with $\mathrm{ad}_Y(X) = [Y, X]$ for $X, Y \in \mathfrak{g}$. Remember that the adjoint representation is on the vector space $\mathfrak{g}$ that the Lie algebra acts on. The generators of the algebra are taken to be the basis vectors of this vector space. In other words, we can write the basis states as

$$|T^a\rangle \ . \tag{39.149}$$

Thus, any state in the adjoint representation is taken to be a linear combination of the generators.

The adjoint representation is special in that the generators act as both operators on the vector space and as states of the vector space. From Eq. (39.47) we have

$$T^b\,|T^a\rangle = \Big|\Big[T^b,\, T^a\Big]\Big\rangle \, . \tag{39.150}$$

This means that in the adjoint representation we can find the weight vectors corresponding to each generator belonging to the Cartan subalgebra:

$$T^3 : \ (0,\, 0) \ , \qquad T^8 : \ (0,\, 0) \ . \tag{39.151}$$

In fact, the weights of the adjoint representation are of special importance and so are given the name of **roots**. The structure constants of $\mathfrak{su}(3)$ make clear that the remainder of the generators are not eigenvectors of T^3 and T^8. This simply means that this basis is not the one in which these two are diagonal. The other states $|X_\alpha\rangle$ with root vectors $\alpha = (\alpha_1,\, \alpha_2)$ are defined by

$$T^3 |X_\alpha\rangle = \alpha_1 |X_\alpha\rangle \ , \qquad T^8 |X_\alpha\rangle = \alpha_2 |X_\alpha\rangle \ . \tag{39.152}$$

Note that this means that

$$\Big[T^3,\, X_\alpha\Big] = \alpha_1 X_\alpha \ , \qquad \Big[T^8,\, X_\alpha\Big] = \alpha_2 X_\alpha \ . \tag{39.153}$$

Exercise 39.28 Show that we can take the Hermitian conjugate of the prevous relations to find

$$\left[T^3,\, X_\alpha^\dagger\right] = -\alpha_1 X_\alpha^\dagger \;, \qquad \left[T^8,\, X_\alpha^\dagger\right] = -\alpha_2 X_\alpha^\dagger \;. \tag{39.154}$$

This shows that $X_\alpha^\dagger = X_{-\alpha}$.

While the X_α are states in the adjoint representation, they correspond to operators in any representation. This is useful for the following reason. Consider a state $|\mu\rangle$ with weight vector $\mu = (\mu_1,\, \mu_2)$. We can then act on this state with X_α and determine how the elements of the Cartan subalgebra act on this state. We find

$$\begin{aligned} T^3 X_\alpha|\mu\rangle &= \left(T^3 X_\alpha - X_\alpha T^3 + X_\alpha T^3\right)|\mu\rangle \\ &= \left[T^3,\, X_\alpha\right]|\mu\rangle + X_\alpha T^3|\mu\rangle \\ &= \alpha_1 X_\alpha|\mu\rangle + \mu_1 X_\alpha|\mu\rangle \\ &= (\alpha_1 + \mu_1) X_\alpha|\mu\rangle \;. \end{aligned} \tag{39.155}$$

Exercise 39.29 Show that

$$T^8 X_\alpha|\mu\rangle = (\alpha_2 + \mu_2) X_\alpha|\mu\rangle \;. \tag{39.156}$$

The point of these calculations is to show that the X_α operators act like raising or lowering operators on vectors with weights. This is true in any representation and so leads to restrictions on what representations are allowed. In particular, it is true in the adjoint representation. This allows us to evaluate

$$X_\alpha\,|X_{-\alpha}\rangle = |[X_\alpha,\, X_{-\alpha}]\rangle \;, \tag{39.157}$$

which must have a root of $\alpha - \alpha = (0,\, 0)$. But, having a zero root vector means this state commutes with both elements of the Cartan subalgebra. In other words, $[X_\alpha,\, X_{-\alpha}]$ is a state which must be a linear conbination of T^3 and T^8 because these are the only states with a root vector of $(0,\, 0)$. (If there was another state with this property then the Cartan subalgebra would have to be expanded to include this new operator.) Though we do not prove it here it turns out that

$$[X_\alpha,\, X_{-\alpha}] = \alpha_1 T^3 + \alpha_2 T^8 \;. \tag{39.158}$$

This leads to three special operators related to the root vector α. We have X_α, $X_{-\alpha}$ and $\alpha_1 T^3 + \alpha_2 T^8$. Notice that from Eq. (39.153)

$$\left[\alpha_1 T^3 + \alpha_2 T^8,\, X_{\pm\alpha}\right] = \pm\left(\alpha_1^2 + \alpha_2^2\right) X_{\pm\alpha} \;. \tag{39.159}$$

We define $|\alpha| = \sqrt{\alpha_1^2 + \alpha_2^2}$ and then

$$X_\alpha^\pm \equiv \frac{\sqrt{2}}{|\alpha|} X_{\pm\alpha} \ , \quad X_\alpha^3 \equiv \frac{1}{\alpha^2}\left(\alpha_1 T^3 + \alpha_2 T^8\right) . \tag{39.160}$$

Exercise 39.30 By comparing with the $\mathfrak{su}(2)$ raising and lowering operators in Eq. (39.51) and considering the relations in Eqs. (39.53) and (39.54), show that $X_\alpha^\pm, X_\alpha^3$ constitute an $\mathfrak{su}(2)$ subalgebra.

Okay, what has been accomplished in this. We have assumed a general representation of $\mathfrak{su}(3)$. Using the Cartan subalgebra (maximal set of commuting algebra generators) we have chosen a basis in which this subalgebra is diagonal. This means that for an N dimensional representation there are N basis vectors (or states) which are eigenvectors of the full Cartan subalgebra. These eigenvectors are characterized by a weight vector (the vector of eigenvalues). In the adjoint representation, whose dimension n is equal to the number of generators of the Lie algebra, the weights of the n basis vectors are called roots. What is more the basis vectors with nonzero roots come in pairs, one with a positive root and the other with a negative root of equal magnitude. These, along with a linear combination of elements of the Cartan subalgebra, make up an $\mathfrak{su}(2)$ subalgebra.

This is significant because we already know everything about $\mathfrak{su}(2)$ representations. The fact that all these $\mathfrak{su}(2)$ subalgebra exist puts stringent constraints on what representations are possible in $\mathfrak{su}(3)$. For instance, though we do not prove it here, this $\mathfrak{su}(2)$ implies that if α is a root then no multiple of α, other than $-\alpha$, is a root.

Example The roots of $\mathfrak{su}(3)$ are determined by the commutation relations of the generators. From the structure constants (39.130) of the algebra we can isolate all the commutators with T^3 and T^8 (the basis for the Cartan subalgebra)

$$f^{123} = 1 \ , \quad f^{345} = f^{376} = \frac{1}{2} \ , \quad f^{458} = f^{678} = \frac{\sqrt{3}}{2} \ . \tag{39.161}$$

The second and third equalities imply that

$$\left[T^3,\, T^1\right] = iT^2 \;, \qquad \left[T^3,\, T^2\right] = -\,iT^1 \;, \tag{39.162}$$

$$\left[T^8,\, T^1\right] = 0 \;, \qquad \left[T^8,\, T^2\right] = 0 \;, \tag{39.163}$$

$$\left[T^3,\, T^4\right] = \frac{i}{2}T^5 \;, \qquad \left[T^3,\, T^5\right] = -\,\frac{i}{2}T^4 \;, \tag{39.164}$$

$$\left[T^8,\, T^4\right] = \frac{i\sqrt{3}}{2}T^5 \;, \qquad \left[T^8,\, T^5\right] = -\,\frac{i\sqrt{3}}{2}T^4 \;, \tag{39.165}$$

$$\left[T^3,\, T^7\right] = \frac{i}{2}T^6 \;, \qquad \left[T^3,\, T^6\right] = -\,\frac{i}{2}T^7 \;, \tag{39.166}$$

$$\left[T^8,\, T^7\right] = -\,\frac{i\sqrt{3}}{2}T^6 \;, \qquad \left[T^8,\, T^6\right] = \frac{i\sqrt{3}}{2}T^7 \;. \tag{39.167}$$

Exercise 39.31

Show that

$$\left[T^3,\, \frac{1}{\sqrt{2}}\left(T^1 \pm iT^2\right)\right] = \pm\,\frac{1}{\sqrt{2}}\left(T^1 \pm iT^2\right) \;, \tag{39.168}$$

$$\left[T^8,\, \frac{1}{\sqrt{2}}\left(T^1 \pm iT^2\right)\right] = 0 \;, \tag{39.169}$$

$$\left[T^3,\, \frac{1}{\sqrt{2}}\left(T^4 \pm iT^5\right)\right] = \pm\,\frac{1}{2}\frac{1}{\sqrt{2}}\left(T^4 \pm iT^5\right) \;, \tag{39.170}$$

$$\left[T^8,\, \frac{1}{\sqrt{2}}\left(T^4 \pm iT^5\right)\right] = \pm\,\frac{\sqrt{3}}{2}\frac{1}{\sqrt{2}}\left(T^4 \pm iT^5\right) \;, \tag{39.171}$$

$$\left[T^3,\, \frac{1}{\sqrt{2}}\left(T^6 \pm iT^7\right)\right] = \mp\,\frac{1}{2}\frac{1}{\sqrt{2}}\left(T^6 \pm iT^7\right) \;, \tag{39.172}$$

$$\left[T^8,\, \frac{1}{\sqrt{2}}\left(T^6 \pm iT^7\right)\right] = \pm\,\frac{\sqrt{3}}{2}\frac{1}{\sqrt{2}}\left(T^6 \pm iT^7\right) \;, \tag{39.173}$$

This shows the vectors

$$\frac{1}{\sqrt{2}}\left(T^1 \pm iT^2\right) \;,\quad \frac{1}{\sqrt{2}}\left(T^4 \pm iT^5\right) \;,\quad \frac{1}{\sqrt{2}}\left(T^6 \pm iT^7\right) \tag{39.174}$$

have the root vectors

$$(\pm 1, 0) \;,\quad \pm\left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right) \;,\quad \pm\left(-\frac{1}{2}, \frac{\sqrt{3}}{2}\right) \;. \tag{39.175}$$

Note that T^3 and T^8 both have the root vector (0, 0). This also implies that any linear combination of T^3 and T^8 have root vector (0, 0).

Definition We give two of these roots a special designation as **positive simple roots**:

$$\alpha^{(1)} = \left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right) , \quad \alpha^{(2)} = \left(\frac{1}{2}, -\frac{\sqrt{3}}{2}\right) . \tag{39.176}$$

In general, positive simple roots are roots with the property that all the other roots of the algebra can be expressed as a linear combination of them with integer coefficients. In addition, these coefficients are either all greater than or equal to zero or else all less than or equal to zero.

Example The roots of the $\mathfrak{su}(3)$ are

$$\pm\alpha^{(1)} , \quad \pm\alpha^{(2)} , \quad \pm\left(\alpha^{(1)} + \alpha^{(2)}\right) . \tag{39.177}$$

Now what can we do with these root, weights, and $\mathfrak{su}(2)$s? Consider the action of X^3_α as an operators acting on a general representation. Consider a state $|\mu\rangle$ with a weight vector μ. Acting with X^3_α we have

$$\begin{aligned} X^3_\alpha|\mu\rangle =& \frac{1}{\alpha^2}\left(\alpha_1 T^3 + \alpha_2 T^8\right)|\mu\rangle = \frac{\alpha_1\mu_1 + \alpha_2\mu_2}{\alpha^2}|\mu\rangle \\ \equiv& \frac{\alpha\cdot\mu}{\alpha^2}|\mu\rangle . \end{aligned} \tag{39.178}$$

This gives a geometric-like feel to the roots and weights. The weight vectors in any representation are eigenvectors of each X^3_α. But, because X^3_α acts like the z-component of angular momentum, remember that is is part of a representation of $\mathfrak{su}(2)$, its eigenvalues are integer or half-integer. This means that for any representation and any weight μ and for all roots α it must be that

$$2\frac{\alpha\cdot\mu}{\alpha^2} , \tag{39.179}$$

is an integer. This is a highly constraining requirement.

Exercise 39.32 Show that the constrain in Eq. (39.179) is satisfied for the root vectors of the adjoint representation.

Recall also that

$$X^\pm_\alpha|\mu\rangle \propto |\mu \pm \alpha\rangle . \tag{39.180}$$

This means that the raising and lowering operators relate vectors with weight μ to vectors with weight $\mu \pm \alpha$. And, because representations of $\mathfrak{su}(2)$ have a highest and lowest eigenvalue of X^3_α there must be vectors with a lowest and highest weight in our general representation of $\mathfrak{su}(3)$.

For the notion of highest weight to make any sense we need to be able to compare weights. Given two weights μ_1 and μ_2 we say that $\mu_1 > \mu_2$ if and only if

$$\mu_1 - \mu_2 = \sum \alpha^{(i)} a_i , \tag{39.181}$$

with $a_i \geq 0$. By taking the negative of both sides it is easy to see what it means to have $\mu_2 < \mu_1$, the sum must have no positive coefficients. But, you may notice that if neither of these conditions are met then we may have weights that are not ordered in this way.

Example

For instance $\alpha^{(1)} - \alpha^{(2)}$ does not fit the definition given above, so we cannot order the two simple roots. Conversely,

$$\alpha^{(1)} - (-\alpha^{(1)}) = 2\alpha^{(1)} \tag{39.182}$$

so we have

$$\alpha^{(1)} > -\alpha^{(1)} \ . \tag{39.183}$$

Exercise 39.33

Show that for $\mathfrak{su}(3)$ the highest root (weight of the adjoint representation) is

$$\alpha^{(1)} + \alpha^{(2)} \ , \tag{39.184}$$

and that the lowest root is

$$-\alpha^{(1)} - \alpha^{(2)} \ . \tag{39.185}$$

The action of the raising and lowering operators can move us among the weights. This may bring us toward the highest weight or away from it.

Example

For the adjoint representation we have

$$X_{12}^{\pm} \equiv \frac{1}{\sqrt{2}} \left(T^1 \pm iT^2 \right) \ , \tag{39.186}$$

$$X_{1}^{\pm} \equiv \frac{1}{\sqrt{2}} \left(T^4 \pm iT^5 \right) \ , \tag{39.187}$$

$$X_{2}^{\pm} \equiv \frac{1}{\sqrt{2}} \left(T^6 \mp iT^7 \right) \ . \tag{39.188}$$

The labeling here refers to how the simple roots relate to the roots of these operators. The choice of sign for $X_2^{\pm}$ is such that the $+$ operator moves in the positive direction along $\alpha^{(2)}$. We find that

$$\left[X_{12}^{+}, X_{2}^{-} \right] = \frac{1}{\sqrt{2}} X_1^{+} \ , \quad \left[X_{12}^{-}, X_{2}^{+} \right] = -\frac{1}{\sqrt{2}} X_1^{-} \ , \tag{39.189}$$

and

$$\left[X_{12}^{+}, X_{2}^{+} \right] = \left[X_{12}^{-}, X_{2}^{-} \right] = 0 \ . \tag{39.190}$$

Exercise 39.34

Show that

$$\left[X_{12}^-,\, X_1^+\right] = \frac{1}{\sqrt{2}} X_2^- \;, \qquad \left[X_{12}^+,\, X_1^-\right] = -\frac{1}{\sqrt{2}} X_2^+ \;, \tag{39.191}$$

$$\left[X_{12}^+,\, X_1^+\right] = \left[X_{12}^-,\, X_1^-\right] = 0 \;, \tag{39.192}$$

and

$$\left[X_1^+,\, X_2^+\right] = \frac{1}{\sqrt{2}} X_{12}^+ \;, \qquad \left[X_1^-,\, X_2^-\right] = -\frac{1}{\sqrt{2}} X_{12}^- \;, \tag{39.193}$$

$$\left[X_1^+,\, X_2^-\right] = \left[X_1^-,\, X_2^+\right] = 0 \;. \tag{39.194}$$

These results illustrate the raising operator behavior. We have that

$$X_{12}^+ |X_2^-\rangle \propto |X_1^+\rangle \;, \tag{39.195}$$

and

$$\left(X_{12}^+\right)^2 |X_2^-\rangle \propto X_{12}^+ |X_1^+\rangle = |0\rangle \;. \tag{39.196}$$

In other words, X_{12}^+ acts like a raising operator taking $|X_2^-\rangle$ into $|X_1^+\rangle$. However, in this case trying to raise it further leads to the zero state. Note that this can be seen from the weights themselves

$$-\alpha^{(2)} + \left(\alpha^{(1)} + \alpha^{(2)}\right) = \alpha^{(1)} \;. \tag{39.197}$$

If we again add $\alpha^{(1)} + \alpha^{(2)}$ we do not obtain one of the roots we have identified and, in fact, it would have a higher weight than $\alpha^{(1)} + \alpha^{(2)}$ which we already identified as having the highest weight.

Exercise 39.35

What state is

$$X_{12}^+ X_1^+ X_2^- X_{12}^- X_1^- X_2^+ |X_1^+\rangle \;, \tag{39.198}$$

proportional to?

Exercise 39.36

Evaluate

$$X_2^- X_1^- |X_{12}^+\rangle \;, \tag{39.199}$$

and

$$X_1^- X_2^- |X_{12}^+\rangle \;. \tag{39.200}$$

Show that these two states are linearly independent and both have root vectors of (0, 0). This shows that more than one distinct state may have the same root or weight vector.

We also see that the action of both X_1^+ and X_2^+ on X_{12}^+ vanish. This means that X_{12}^+ has the highest weight in this representation because the actions of all the raising operators produce the zero vector.

In a general representation there is some integer $p \geq 0$ such that for state with weight μ

$$\left(X_\alpha^+\right)^p |\mu\rangle \propto |\mu + p\alpha\rangle \neq |0\rangle \ , \tag{39.201}$$

and

$$\left(X_\alpha^+\right)^{p+1} |\mu\rangle = |0\rangle \ . \tag{39.202}$$

We then have that

$$E_\alpha^3 \left(X_\alpha^+\right)^p |\mu\rangle = \frac{\alpha \cdot (\mu + p\alpha)}{\alpha^2} \left(X_\alpha^+\right)^p |\mu\rangle \ . \tag{39.203}$$

But, because this is the highest weight of an $\mathfrak{su}(2)$ representation we have that

$$\frac{\alpha \cdot (\mu + p\alpha)}{\alpha^2} = j \ , \tag{39.204}$$

for some half-integer j. Recall from our discussion of representations of $\mathfrak{su}(2)$ that the eigenstates of E^3 for j representation have eigenvalues from $-j$ to j in integer steps. The value j is the highest eigenvalue and $-j$ is the lowest.

This means that there is some integer $q \geq 0$ such that

$$\left(X_\alpha^-\right)^q |\mu\rangle \propto |\mu - q\alpha\rangle \neq |0\rangle \ , \tag{39.205}$$

and

$$\left(X_\alpha^-\right)^{q+1} |\mu\rangle = |0\rangle \ . \tag{39.206}$$

The E_α^3 value of this lowest weight is

$$\frac{\alpha \cdot (\mu - q\alpha)}{\alpha^2} = -j \ . \tag{39.207}$$

Exercise 39.37 From Eqs. (39.204) and (39.207) show that

$$q + p = 2j \ , \quad q - p = 2\frac{\alpha \cdot \mu}{\alpha^2} \ . \tag{39.208}$$

These results are especially powerful in the adjoint representation. In this case the weight μ is also a root, so perhaps we should call it β. Then there are two non-negative integers p and q such that

$$q - p = 2\frac{\alpha \cdot \beta}{\alpha^2} \ , \tag{39.209}$$

and there are two other non-negative integers r and s such that

$$s - r = 2\frac{\beta \cdot \alpha}{\beta^2} \ . \tag{39.210}$$

We can multiply these two equations to find

$$4\frac{(\alpha\cdot\beta)^2}{\alpha^2\beta^2} = (q-p)(s-r) \ . \tag{39.211}$$

Recall from the standard definition of a dot product that $\alpha\cdot\beta = |\alpha|\,|\beta|\cos\theta_{\alpha\beta}$. This implies that

$$\cos^2\theta_{\alpha\beta} = \frac{1}{4}(q-p)(s-r) \ . \tag{39.212}$$

So, geometrically, we find only a few possibilities for the angles between two roots, thought of as vectors. The roots are unique, so the possibility of $\theta_{\alpha\beta} = 0$ for $\alpha \neq \beta$ is impossible. We also know that for any root α there is a root $-\alpha$ obtained by taking the Hermitian conjugate of the generator.

Exercise 39.38 Show that for $\beta = -\alpha$ that

$$\theta_{\alpha\beta} = \pi \ . \tag{39.213}$$

Exercise 39.39 Show that, besides $\theta_{\alpha\beta} = 0, \pi$, the only values that produce a $\cos^2\theta_{\alpha\beta}$ which is an integer divided by four are

$$\theta_{\alpha\beta} = \left\{\frac{\pi}{6}, \frac{\pi}{4}, \frac{\pi}{3}, \frac{2\pi}{3}, \frac{3\pi}{4}, \frac{5\pi}{6}\right\} \ . \tag{39.214}$$

Here we do not consider angles larger than π. In other words, we find the the smallest angle between the vectors. Note that these are the *only* angles allowed for the roots of any Lie Algebra. This is closely tied to the categorization of Lie algebras, though we do not pursue that in these notes.

Exercise 39.40 Recall that the roots of $\mathfrak{su}(3)$ are

$$\alpha^{(1)} + \alpha^{(2)} = (1,0) \ , \quad \alpha^{(1)} = \left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right) \ , \quad \alpha^{(2)} = \left(\frac{1}{2}, -\frac{\sqrt{3}}{2}\right) \ , \tag{39.215}$$

find $\theta_{12,1}$, $\theta_{12,2}$, and $\theta_{1,2}$. To what representation of $\mathfrak{su}(2)$ (what j) do actions of the raising and lowering operators correspond to?

While we do not prove it in these notes, the irreducible representations of Lie algebras are specified by unique highest weights. So, if we can determine what the possible highest weights are we can determine what representations are allowed. We work this out for the $\mathfrak{su}(3)$ case, but the generalization of other groups is quite similar.

Suppose in some irreducible representation the highest weight is μ_1. It must be that acting with either of X_1^+ or X_2^+ on the highest weight state must lead to the zero vector. If not there would be another state with

higher weight. This means that the μ_1 state must be at the top of some $\mathfrak{su}(2)$ representation along $\alpha^{(1)}$ at the top of another $\mathfrak{su}(2)$ representation along $\alpha^{(2)}$. In other words

$$\frac{\alpha^{(1)}\cdot\mu_1}{(\alpha^{(1)})^2}=j^{(1)}\ ,\qquad \frac{\alpha^{(2)}\cdot\mu_1}{(\alpha^{(2)})^2}=j^{(2)}\ , \tag{39.216}$$

with the $j^{(i)}$s non-negative, half-integers. Every choice of $j^{(i)}$ specifies a representation of $\mathfrak{su}(3)$, which is sometimes labeled as $(2j^{(1)},2j^{(2)})$. The work now is to determine the characteristics of those representations. A trivial example is taking both $j^{(i)}$s equal to zero. This simply leads to the trivial representation of the algebra.

Example

Consider the **fundamental representation**, denoted **3**, of $\mathfrak{su}(3)$. Recall from Eq. (39.138) that the three weight vectors with weights

$$\mu_1=\left(\frac{1}{2},\frac{1}{2\sqrt{3}}\right)\ ,\quad \mu_2=\left(-\frac{1}{2},\frac{1}{2\sqrt{3}}\right)\ ,\quad \mu_3=\left(0,-\frac{1}{\sqrt{3}}\right) \tag{39.217}$$

are

$$f_1=\begin{pmatrix}1\\0\\0\end{pmatrix}\ ,\quad f_2=\begin{pmatrix}0\\1\\0\end{pmatrix}\ ,\quad f_3=\begin{pmatrix}0\\0\\1\end{pmatrix}\ . \tag{39.218}$$

These correspond to three points in a plane. We can move from one vertex to another by acting with one of the raising or lowering operators. For instance to go from μ_2 to μ_1 we act with X^+_{12} because

$$\mu_2+\alpha^{(1)}+\alpha^{(2)}=\mu_1\ . \tag{39.219}$$

We also find that

$$\mu_2+\alpha^{(2)}=\mu_3\ ,\qquad \mu_3+\alpha^{(1)}=\mu_1\ . \tag{39.220}$$

This implies that μ_1 is the highest weight and μ_2 is the lowest. Thus, we expect that the actions of X^+_{12}, X^+_1, and X^+_2 on the μ_1 state to all vanish. Using the Gell-Mann matrices (39.125) we find

$$X^+_{12}=\frac{1}{\sqrt{2}}\begin{pmatrix}0&1&0\\0&0&0\\0&0&0\end{pmatrix}\ ,\quad X^+_1=\frac{1}{\sqrt{2}}\begin{pmatrix}0&0&1\\0&0&0\\0&0&0\end{pmatrix}\ ,\quad X^+_2=\frac{1}{\sqrt{2}}\begin{pmatrix}0&0&0\\0&0&0\\0&1&0\end{pmatrix}\ , \tag{39.221}$$

$$X^-_{12}=\frac{1}{\sqrt{2}}\begin{pmatrix}0&0&0\\1&0&0\\0&0&0\end{pmatrix}\ ,\quad X^-_1=\frac{1}{\sqrt{2}}\begin{pmatrix}0&0&0\\0&0&0\\1&0&0\end{pmatrix}\ ,\quad X^-_2=\frac{1}{\sqrt{2}}\begin{pmatrix}0&0&0\\0&0&1\\0&0&0\end{pmatrix}\ . \tag{39.222}$$

We find that

$$X_{12}^{+} f_1 = X_1^{+} f_1 = X_2^{+} f_1 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} . \tag{39.223}$$

This confirms that f_1 (with weight μ_1) has the highest weight.

We next calculate

$$2\frac{\alpha^{(1)} \cdot \mu_1}{\left(\alpha^{(1)}\right)^2} = 1 , \qquad 2\frac{\alpha^{(2)} \cdot \mu_1}{\left(\alpha^{(2)}\right)^2} = 0 , \tag{39.224}$$

Thus, we can label this representation as (1,0). This also lets us go the other way. The (1,0) signifies that that this highest weight is at the top of a spin-1/2 representation along $\alpha^{(1)}$ and in the trivial representation of $\alpha^{(2)}$. This means there should be exactly one state moving in the $-\alpha^{(1)}$ direction from the highest weight. Also, as $\alpha^{(1)} + \alpha^{(2)} = j_1 + j_2 = 1/2$ the highest weight is also at the top of a spin-1/2 representation in the $\alpha^{(1)} + \alpha^{(2)}$ direction. This completely accounts for and determines the three states that make up this representation. This weight system is plotted in Fig. 33.

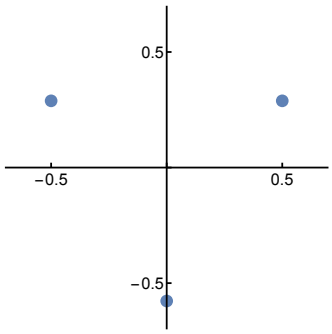


Figure 33. Root system of the fundamental representation of SU(3).

Exercise 39.41 By acting with the lowering operators find all the other weight vectors of the fundamental representation.

Example To see how things work from the other direction, we construct the (0,1) representation. This means that for the highest weight μ_h

$$\frac{\alpha^{(1)} \cdot \mu_h}{\left(\alpha^{(1)}\right)^2} = 0 , \qquad \frac{\alpha^{(2)} \cdot \mu_h}{\left(\alpha^{(2)}\right)^2} = \frac{1}{2} , \tag{39.225}$$

If we take the highest weight to have the form

$$\mu_h = (c_1, c_2) , \tag{39.226}$$

then the above constraint imply that

$$c_1 = \frac{1}{2} , \qquad c_2 = -\frac{1}{2\sqrt{3}} . \tag{39.227}$$

The previous example makes clear an important point for any representation. Suppose we are in the (m, n) representation. Then we have that

$$\frac{\alpha^{(1)} \cdot \mu_h}{\left(\alpha^{(1)}\right)^2} = \frac{m}{2} , \qquad \frac{\alpha^{(2)} \cdot \mu_h}{\left(\alpha^{(2)}\right)^2} = \frac{n}{2} \tag{39.228}$$

for the highest weight μ_h. This means that the highest weight is given by

$$\mu_h = \left(\frac{m+n}{2}, \frac{m-n}{2\sqrt{3}} \right) . \tag{39.229}$$

Exercise 39.42

Being the highest weight means that neither of

$$\mu_h + \alpha^{(1)} , \qquad \mu_h + \alpha^{(2)} , \tag{39.230}$$

are states. Find the minimum norm of these two vectors for the (m, n) representation. This value squared is an upper bound on the value of the Casimir $C(r)$ for this representation. Evaluate this bound for the fundamental representation, its conjugate, and the adjoint representation.

Exercise 39.43

By acting with the lowering operators find all the other weight vectors of the (0,1) representation. The number of weight vectors corresponds to the dimensionality of the representation. You should find that this representation is three dimensional. How do the weights relate to the weights of the (1,0) representation?

The representation just constructed is the conjugate representation of the fundamental, often referred to as $\overline{\mathbf{3}}$. The generators are

$$T^a = -\frac{1}{2}\lambda^{a*} , \tag{39.231}$$

where the λ^a are the Gell-Mann matrices (39.125). Show that the weights you have discovered correspond to the basis vectors

$$f_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} , \quad f_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} , \quad f_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} . \tag{39.232}$$

Compare this weight system (plotted in Fig. 34) with the fundamental representation (shown in Fig. 33).

Notice that Eq. (39.229) can also be written as

$$\mu_h = m \left(\frac{1}{2}, \frac{1}{2\sqrt{3}} \right) + n \left(\frac{1}{2}, -\frac{1}{2\sqrt{3}} \right) . \tag{39.233}$$

This is simple m times the the highest weight of the fundamental representation added to n times the the highest weight of the conjugate of the fundamental representation.

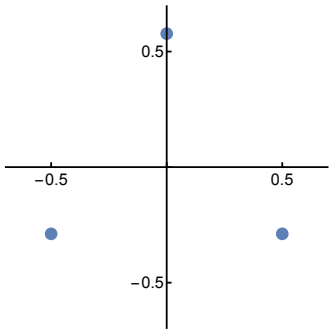


Figure 34. Root system of the conjugate of the fundamental representation of SU(3).

Exercise 39.44 Construct the (1,1) representation. What dimension is this representation? Note that there are two distinct states with weight vector $(0, 0)$.

That the weights of the $\overline{\mathbf{3}}$ or (0,1) representation are the negatives of the **3** or (1,0) representation weights is a specific example of a general result. The conjugate of a (m, n) representation is the (n, m) representation. When $n \neq m$ they have different highest weights and consequently lead to distinct, inequivalent, representations. Note that the adjoint representation, or (1,1) or **8**, always includes roots and their negatives, as seen in Fig. 35. Hence, the conjugate representation of the adjoint representation is again the adjoint representation. In other words, the adjoint representation is a real representation.

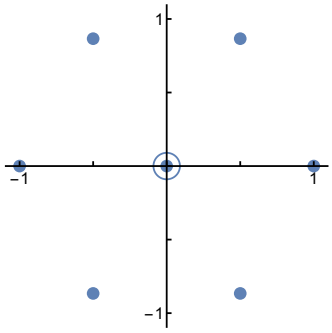


Figure 35. Root system of the adjoint representation of SU(3). The circle around the center denotes a second state.

Exercise 39.45 Construct the weights for the (2,0), or **6**, representation.

Exercise 39.46 Construct the weights for the (3,0), or **10**, representation.

Exercise 39.47 Construct the weights for the (2,1), or **15**, representation. Note that three of the weights correspond to two distinct states.

Exercise 39.48 Check that all of the representations (m, n) considered so far have a dimensionality given by

$$\frac{1}{2}(m+1)(n+1)(m+n+2) \ . \tag{39.234}$$

Using this formula find the dimension of the (2,2) representation.

Finally, we can consider the tensor product of representations. As we saw with $\mathfrak{su}(2)$, these products of representation usually are not irreducible representations themselves, but they can be expressed as the direct sum of irreducible representations. And as before the dimension of the representations can be a useful constraint. However, we also have to be aware that many representation have conjugates that are inequivalent, but have the same dimension.

Example Consider $\mathbf{3} \otimes \overline{\mathbf{3}}$. By simple multiplication we know that the total dimension of this product group is nine. But, from the representations we have catalogued above we can see that there is no dimension nine irreducible representation. However, the highest weights of the two representations are $(1, 0)$ and $(0, 1)$. It turns out that the highest weight in the product group is the sum of these, or $(1, 1)$. This is the **8**. Only the singlet

representation **1** can make up the difference in dimension so we have

$$\mathbf{3} \otimes \overline{\mathbf{3}} = \mathbf{8} \oplus \mathbf{1} \ . \tag{39.235}$$

Another way to argue that this had to be the result is to consider the conjugate representation, which is simply $\overline{\mathbf{3}} \otimes \mathbf{3}$. This equal to the original representation, or in other words this product representation is real. This means that all the terms in the direct sum must be real representations. The only real representations with dimension less than nine are **8** and **1**.

Exercise 39.49 Use logic similar to the previous example to argue that

$$\mathbf{6} \otimes \overline{\mathbf{6}} = \mathbf{27} \oplus \mathbf{8} \oplus \mathbf{1} \ . \tag{39.236}$$

What about $\mathbf{3} \otimes \mathbf{3}$? This is also $(1,0) \otimes (1,0)$ so we know it must contain $(2,0)$ from our understanding of highest weights. So, we have

$$\mathbf{3} \otimes \mathbf{3} = \mathbf{6} \oplus \ldots \ . \tag{39.237}$$

The missing representation on the right-hand side must be dimension three, but is it **3** or $\overline{\mathbf{3}}$?

To answer this question we need to do a bit more work. This also leads us to more general methods for determining what makes up these direct sums. There is a sense in which the fundamental and antifundamental (conjugate) representations are the basic building blocks of the other representations. If we consider some matrix U in SU(3) the a state v in the fundamental representation (**3**) transforms like

$$v^i \to U^i{}_j v^j \ . \tag{39.238}$$

To determine how a state $v^\dagger$ that transforms in the antifundamental representation ($\overline{\mathbf{3}}$) we simply take the Hermitian conjugate of (39.238)

$$v^\dagger_i \to v^\dagger_j U^{\dagger j}{}_i \ . \tag{39.239}$$

It is significant that states in the fundamental representation have an index up, while those in the conjugate representation have a lower index. We can then construct various tensors by simply taking an outer produce of states. For instance, a state in the $\mathbf{3} \otimes \overline{\mathbf{3}}$ representation would be something like

$$v^i u^\dagger_j \ . \tag{39.240}$$

This has the form of a three-by-three matrix. In general, it has nine complex components.

Exercise 39.50 Show that the trace of (39.240)

$$v^i u_i^\dagger \ , \tag{39.241}$$

is invariant under SU(3) transformations. This shows that this particular combination of the matrix components is in the singlet representation **1**.

Exercise 39.51 Show that the trace of

$$v^i u_j^\dagger - \frac{1}{3}\delta_j^i v^k u_k^\dagger \ , \tag{39.242}$$

is zero. How does this object change under SU(3) transformations? By comparing with Eq. (39.22) show that it transforms in the adjoint representation **8**.

Considering we can write

$$v^i u_i^\dagger = \left(v^i u_j^\dagger - \frac{1}{3}\delta_j^i v^k u_k^\dagger\right) + \frac{1}{3}\delta_j^i v^k u_k^\dagger \ , \tag{39.243}$$

we have another illustration of $\mathbf{3} \otimes \overline{\mathbf{3}} = \mathbf{8} \oplus \mathbf{1}$.

We can now use these methods to consider $\mathbf{3} \otimes \mathbf{3}$. This is represented by something like

$$v^i u^j \ . \tag{39.244}$$

This again seems like a three-by-three array. But we argued earlier from highest weight considerations that the **6** must appear in the direct sum. Remember that any matrix or tensor can be expressed as the sum of a symmetric piece and an antisymmetric piece. You can convince yourself that a symmetric three-by-three tensor has six unique components, which matches the **6**.

Exercise 39.52 By using the identities near Eq. (35.58) show that

$$v^i u^j - v^j u^i = \varepsilon^{ijk}\varepsilon_{mnk} v^m u^n \ . \tag{39.245}$$

Exercise 39.53 By using Eq. (35.73) show that under an arbitrary SU(3) transformation U

$$\varepsilon_{mnk} v^m u^n \to \varepsilon_{mn\ell} v^m u^n U^{\dagger\ell}{}_k \ . \tag{39.246}$$

This shows, by comparing with Eq. (39.239), that this object transforms in the $\overline{\mathbf{3}}$ representation

We can then write

$$\begin{aligned} v^i u^j =& \frac{1}{2}\left(v^i u^j + v^j u^i\right) + \frac{1}{2}\left(v^i u^j - v^j u^i\right) \\ =& \frac{1}{2}\left(v^i u^j + v^j u^i\right) + \frac{1}{2}\varepsilon^{ijk}\varepsilon_{mnk} v^m u^n \ . \end{aligned} \tag{39.247}$$

We have already identified the first term with the **6** representation. From the exercise above the antisymmetric terms is in the $\overline{\mathbf{3}}$ representation. We conclude that

$$\mathbf{3} \otimes \mathbf{3} = \mathbf{6} \oplus \overline{\mathbf{3}} \ . \tag{39.248}$$

This trick of rewriting a $\overline{\mathbf{3}}$ index as the antisymmetric combination of two **3** indices means that we can express every tensor product in terms of **3** indices. For instance, consider $\mathbf{3} \otimes \mathbf{6}$. This is a combination of $(1,0)$ and $(2,0)$, so the highest weight is $(3,0)$ or the **10**. Taking v^i to be in the **3** representation and $s^{jk} = s^{kj}$ to be in the **6** the product is

$$v^i s^{jk} = \frac{1}{2}\left(v^i s^{jk} + v^j s^{ik}\right) + \frac{1}{2}\left(v^i s^{jk} - v^j s^{ik}\right) \ . \tag{39.249}$$

The first term is completely symmetric. This is the **10**.

Exercise 39.54 Show that a tensor T^{ijk} that is completely symmetric (the exchange of any two indices gives the same value) has at most 10 distinct components.

The second term in Eq. (39.249) is composed of one **3** index and an antisymmetric pair. This is equivalent to one **3** index and one $\overline{\mathbf{3}}$ index. From Eq. (39.243) we see that this must be the **8**. Therefore, we have

$$\mathbf{3} \otimes \mathbf{6} = \mathbf{10} \oplus \mathbf{8} \ . \tag{39.250}$$

One more example, which illustrates another important concept, is $\overline{\mathbf{3}} \otimes \mathbf{6}$. This is $(0,1) \otimes (2,0)$ so it has highest weight $(2,1)$ which is the **15**. We can write the **6** as $s^{ij} = s^{ji}$ and we can write the $\overline{\mathbf{3}}$ as $v^{mn} = -v^{nm}$. The product is

$$s^{ij} v^{mn} = \frac{1}{2}\left(s^{ij} v^{mn} + s^{im} v^{jn}\right) + \frac{1}{2}\left(s^{ij} v^{mn} - s^{im} v^{jn}\right) \ . \tag{39.251}$$

The first term, the symmetric product, is the highest weight term, the **15**. (You can convince yourself that a rank four tensor in three dimensions that has three symmetric indices and one antisymmetric one has at most 15 distinct components.) The last term has three completely antisymmetric indices. This must be proportional to ε^{ijk} which is a gauge singlet or the **1**. This leaves one remaining **3** index. Thus, we have

$$\overline{\mathbf{3}} \otimes \mathbf{6} = \mathbf{15} \oplus \mathbf{3} \ . \tag{39.252}$$

Finally, we can find more than two terms in the decomposition of the tensor product. Consider $\mathbf{3} \otimes \mathbf{8}$. This is $(1,0) \otimes (1,1)$ so it has highest weight $(2,1)$ which is the $\mathbf{15}$. We also have that

$$\begin{aligned} \mathbf{3} \otimes \mathbf{3} \otimes \overline{\mathbf{3}} =& \mathbf{3} \otimes (\mathbf{8} \oplus \mathbf{1}) \\ =& (\mathbf{3} \otimes \mathbf{8}) \oplus \mathbf{3} \ . \end{aligned} \tag{39.253}$$

But, from the associative property of the tensor produce we can also write this as

$$\begin{aligned} \mathbf{3} \otimes \mathbf{3} \otimes \overline{\mathbf{3}} =& (\mathbf{6} \oplus \overline{\mathbf{3}}) \otimes \overline{\mathbf{3}} \\ =& (\mathbf{6} \otimes \overline{\mathbf{3}}) \oplus (\overline{\mathbf{3}} \otimes \overline{\mathbf{3}}) \\ =& \mathbf{15} \oplus \mathbf{3} \oplus \overline{\mathbf{6}} \oplus \mathbf{3} \ . \end{aligned} \tag{39.254}$$

By comparing Eqs. (39.253) and (39.254) we see that

$$\mathbf{3} \otimes \mathbf{8} = \mathbf{15} \oplus \overline{\mathbf{6}} \oplus \mathbf{3} \ . \tag{39.255}$$

Exercise 39.55 Show that

$$\mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3} = \mathbf{10} \oplus \mathbf{8} \oplus \mathbf{8} \oplus \mathbf{1} \ . \tag{39.256}$$

Exercise 39.56 Show that

$$\mathbf{8} \otimes \mathbf{8} = \mathbf{27} \oplus \mathbf{10} \oplus \overline{\mathbf{10}} \oplus \mathbf{8} \oplus \mathbf{8} \oplus \mathbf{1} \ . \tag{39.257}$$

There is a systematic way to find how all tensor products of representations are represented as direct sums of irreducible representations. However, I do not take the time to outline it here. Those looking for more information can look up Young's Tableaux. The examples here should provide a strong foundation to understand where the procedures using those methods come from.

SUBSECTION 39.2

Representations of Subgroups

It is sometimes the case that one symmetry group of a system is broken, but that a subgroup of the symmetry is preserved. To understand these situations we need to know what subgroups exist and how representations of the larger group transform under the representations of the subgroup.

This type of breaking is often accomplished by a scalar field developing a nonzero vacuum expectation value (VEV). While the full theory may exhibit a large symmetry the VEV may only preserve some smaller symmetry. When this occurs we say the symmetry is spontaneously broken.

As a simple example consider SU(2). This group has three generators. Any of these, by itself, generates a U(1) subgroup. We can check that this

is a subgroup from the Lie Algebra condition

$$[T^a\,, T^a] = 0\ , \tag{39.258}$$

for a single generator. By similar analysis it is simple to see that no two generators of SU(2) produce a subalgebra because no two generators produce a closed set under commutation.

Exercise 39.57 Show that no two generators of SU(2) produce a subalgebra of $\mathfrak{su}(2)$.

This means that there are only two symmetry breaking possibilities. Either all of the symmetry is broken, SU(2)$\to$ 0, or a single U(1) symmetry remains, SU(2)$\to$U(1). This last case is the interesting one. In such a situation we can choose a basis in which the remaining symmetry generator is diagonal. Since we have already taken T^3 to be diagonal in earlier portions of these notes we assume, without loss of generality, that this generator describes the remaining symmetry.

An aspect of a theory that is in the j representation is then given by the diagonal matrix

$$e^{-iT^3} = \begin{pmatrix} e^{-ij} & & \\ & \ddots & \\ & & e^{ij} \end{pmatrix}\ . \tag{39.259}$$

This shows that when the symmetry breaks the j representation breaks into a direct sum of $2j+1$ representations of U(1) each specified by the eigenvalues of T^3. Note that the dimension of the original SU(2) representation determines how many one-dimensional representation of U(1) appear.

To see how such a symmetry breaking might occur consider a scalar field $\phi(x)$. If ϕ is in the trivial representation of SU(2) then if it get a VEV

$$\langle\phi\rangle = v\ , \tag{39.260}$$

the symmetry of the theory is unaffected. Consider instead that ϕ is in the fundamental representation and its VEV is

$$\langle\phi\rangle = \begin{pmatrix} v \\ 0 \end{pmatrix}\ . \tag{39.261}$$

Exercise 39.58 Act with the generators of $\mathfrak{su}(2)$ in the fundamental representation on the VEV of ϕ given above. Show that only T^3 preserves the VEV. This means that only the abelian subgroup generated by T^3 is preserved by this VEV.

Exercise 39.59 Suppose the VEV of ϕ is

$$\langle\phi\rangle = \begin{pmatrix} 0 \\ v \end{pmatrix} \ . \tag{39.262}$$

Show that the VEV is an eigenvector of T^3. What is the eigenvalue? This VEV also preserves the U(1) symmetry generated by T^3, but has charge under this U(1) given by this eigenvalue.

Finally, we consider breaking a nonabelian group down to smaller nonabelian subgroup. As an example, we consider SU(3)→SU(2). We first note that SU(3) contains several SU(2) subgroups, see Exercise 39.22. None of these are disjoint sets, so at most one of them can be a subgroup. (In particular, there is no SU(2)× SU(2) subgroup.) However, there can be an additional U(1) group along with the SU(2).

Exercise 39.60 Suppose a field ϕ is in the fundamental representation of SU(3) and develops a VEV of

$$\langle\phi\rangle = \begin{pmatrix} 0 \\ 0 \\ v \end{pmatrix} \ . \tag{39.263}$$

Show that this VEV preserves an SU(2)×U(1) subgroup generated by $\{T^1, T^2, T^3, T^8\}$. What is the charge of the VEV under the U(1) part.

Section 40

Contour Integration

In this section we provide the briefest of introductions to contour integration of complex functions. This is a very useful tool in many circumstances. We only use it a very little in these notes. It is a part of the very rich subject of complex analysis, which we do no justice to. Further study if highly recommended.

Any complex number z can be expressed in terms of two real numbers. Sometimes this is expressed as

$$z = x + i\,y \ , \tag{40.1}$$

where x and y are real numbers and $i = \sqrt{-1}$. Other times it is more convenient to write

$$z = re^{i\theta} \ , \tag{40.2}$$

where r is the magnitude or modulus of the complex number (kind of like its length in the complex plane) and θ is the phase. Because one complex number is specified by two real numbers, the set of complex numbers can be conceptualized as a two-dimensional region called the complex plane.

It is important to note, however, that complex numbers are not simply two dimensional vectors. They have other algebraic qualities. For instance, we define complex conjugation by

$$z^* = x - i\,y = re^{-i\theta} \ . \tag{40.3}$$

That is, complex conjugation takes $i \to -i$. This means that we can write the real and imaginary parts of a complex number as

$$x = \frac{1}{2}\,(z + z^*) \ , \qquad y = \frac{-i}{2}\,(z - z^*) \ . \tag{40.4}$$

Any function of the complex numbers $f(z,\, z^*)$ can be understood as a mapping from some part of the complex plane to some part of the complex plane. Derivatives can be defined in terms of limits, just like for functions of real numbers. In that case the derivative is only said to exist at some point x_0 if the limit from below x_0 equals the limit from above x_0. Similarly, for a a function of a complex number the derivative with respect to z only exists if the limits along every direction in the complex plan are equal. It can be shown that this is equivalent to f being a function of z only. Roughly speaking, a function that is differentiable at a point is said to be analytic at that point. If a function is not analytic at some point, it is said to be singular.

Functions of complex numbers can be integrated over regions of the complex plane. We are often interested in one dimensional paths C through the complex plane. These can be denoted by

$$\int_C dz\, f(z,\, z^*) \ . \tag{40.5}$$

If the beginning point and ending point of the path are the same (that is, the path is a closed loop) then the integral is written

$$\oint_C f(z,\, z^*) \ , \tag{40.6}$$

and the direction of integration is taken to be counterclockwise. Both of these types of integral are called contour integrals, but the closed contour case is often more useful. This is because of some powerful theorems regarding closed contour integrals. For instance, if $f(z)$ is analytic over some contour C and all the points contained within it then

$$\oint_C f(z) = 0 \ . \tag{40.7}$$

Example Let's consider a very simple illustration of this point. Consider the func-

tion

$$f(z) = z \ . \tag{40.8}$$

This is analytic everywhere in the complex plane. Let's integrate around a rectangular contour of length L and width W centered on the origin, C_1. The integration is over $dz = dx + idy$, but with one of dx or dy zero on each side of the rectangle because only one of the variables is changing. We can write the integral as

$$\begin{aligned}\oint_{C_1} dz\, z =& \int_{-L/2}^{L/2} i\, dy \left(\frac{W}{2} + iy\right) + \int_{W/2}^{-W/2} dx \left(x + i\frac{L}{2}\right) \\ &+ \int_{L/2}^{-L/2} i\, dy \left(-\frac{W}{2} + iy\right) + \int_{-W/2}^{W/2} dx \left(x - i\frac{L}{2}\right) \ ,\end{aligned} \tag{40.9}$$

where the first term corresponds to the right side of the rectangle, starting from the bottom right corner, the next term is the top and so on. These one-dimensional integrals can be combined in pairs

$$\begin{aligned}\oint_{C_1} dz\, z =& i \int_{-L/2}^{L/2} dy \left(\frac{W}{2} + iy + \frac{W}{2} - iy\right) + \int_{-W/2}^{W/2} dx \left(x - i\frac{L}{2} - x - i\frac{L}{2}\right) \\ =& iW \int_{-L/2}^{L/2} dy - iL \int_{-W/2}^{W/2} dx = iWL - iWL = 0 \ .\end{aligned} \tag{40.10}$$

Note that this holds for any (positive) values of L and W.

Let's try another family of contours. This time circles of radius R centered on the origin, C_2. The integration is over

$$dz = e^{i\theta} dr + ire^{i\theta} d\theta \ , \tag{40.11}$$

but as the radius is fixed only the second term contributes. This means that

$$\begin{aligned}\oint_{C_2} dz\, z =& \int_0^{2\pi} iR\, d\theta\, Re^{i\theta} = iR^2 \int_0^{2\pi} d\theta\, e^{2i\theta} \\ =& = \frac{iR^2}{2i} \left(e^{4\pi i} - 1\right) = 0 \ .\end{aligned} \tag{40.12}$$

Here we have used that $e^{2\pi i} = 1$. Again, this holds for any value of R.

Exercise 40.1 Evaluate contour integrals of $f(z, z^*) = z \pm z^*$ over the two families of contours used in the above example. Comment on what you find.

Another integration identity is that if the point z_0 is contained within the region bounded by some closed contour C then

$$f(z_0) = \frac{1}{2\pi i} \oint_C dz \frac{f(z)}{z - z_0} \ , \tag{40.13}$$

where the contour is traversed in the counterclockwise direction. Similarly, derivatives of f are given by

$$\left.\frac{d^n f}{dz^n}\right|_{z=z_0} = \frac{n!}{2\pi i}\oint_C dz \frac{f(z)}{(z-z_0)^{n+1}} \ . \tag{40.14}$$

Just like functions $f(x)$ of the real numbers can be expanded about come point x_0 in a Taylor series

$$\begin{aligned} f(x) =& \sum_{n=0}^{\infty} \frac{(x-x_0)^n}{n!} \left.\frac{d^n f}{dx^n}\right|_{x=x_0} \\ =& f(x_0) + (x-x_0)\left.\frac{df}{dx}\right|_{x=x_0} + \frac{1}{2}(x-x_0)^2 \left.\frac{d^2 f}{dx^2}\right|_{x=x_0} + \dots \ , \end{aligned} \tag{40.15}$$

a function $f(z)$ can be expanded in a Taylor series within any region in which it is analytic. What is more, there is also a series expansion, called a Laurent series, about singular points

$$f(z) = \sum_{n=-\infty}^{\infty} (z-z_0)^n a_n \ . \tag{40.16}$$

The important difference here is the inclusion of negative powers of $z - z_0$. The coefficients of the non-negative powers are given by the same formula as the Taylor series. The a_n with $n < 0$ are given by

$$a_n = \frac{1}{2\pi i}\oint_C dz \frac{f(z)}{(z-z_0)^{n+1}} \ , \tag{40.17}$$

which is well defined for $n < 0$.

A function that has nonzero a_n with $n < 0$ has a singular point, or a pole, at $z = z_0$. If only the pole only comes from the a_{-1} term, then the function is said to have a simple pole at $z = z_0$. The **residue** of a pole is the coefficient a_{-1}. This term is special because

$$a_{-1} = \frac{1}{2\pi i}\oint_C dz\, f(z) \ . \tag{40.18}$$

That is, the integral of a singular function over some contour is related to the residues of the poles within the contour. Or, more precisely the value of the integral of any function around some closed contour is given by

$$\oint_C dz\, f(z) = 2\pi\, i \sum \text{ Residues} \ , \tag{40.19}$$

where the sum is over all the poles enclosed by the contour C. So, if we can extract the residues of the poles of some function we can evaluate a great many contour integrals of that function. If a function has only a

simple pole at z_0, a simple way to find the residue of the function at z_0 is to evaluate the limit

$$\lim_{z\to z_0}(z-z_0)f(z)\ . \tag{40.20}$$

Example

As a simple example consider the function

$$f(z)=\frac{1}{z}\ . \tag{40.21}$$

It is easy to see that this function is analytic everywhere except at the point $z=0$ where it has a simple, isolated pole. If we expand the function about the pole at $z=0$ (which is just the function) we see the residue is equal to one. We can simple evaluate the contour integral round the circular contour of radius R as

$$\oint_C dz\frac{1}{z}=iR\int_0^{2\pi}d\theta\frac{e^{i\theta}}{Re^{i\theta}}=i\int_0^{2\pi}d\theta=2\pi i\ . \tag{40.22}$$

This is indeed $2\pi i$ times sum of the residues.

Exercise 40.2

Find the poles and residues of the function

$$f(z)=\frac{z^2+1}{z(z-2i)}\ . \tag{40.23}$$

There is much that can be done with the results we already have, but there are also many applications that make use of Jordan's Lemma. This has to do with contours that are circular sections C_R of radius R centered at the origin, in the upper half of the complex plane. If we define the integral

$$I_{R>}=\int_{C_R}dz\,f(z)e^{i\alpha z}\ , \tag{40.24}$$

for $\alpha>0$ a real number and if the function $f(z)$ goes to zero at least as fast as $|z|^{-1}$ in the region of the plane spanned by the contour then

$$\lim_{R\to\infty}I_{R>}=0\ . \tag{40.25}$$

Similarly, if the circular section is in the lower half of the complex plane then integrals of the type

$$I_{R<}=\int_{C_R}dz\,f(z)e^{-i\alpha z}\ , \tag{40.26}$$

with $\alpha>0$ vanish as $R\to\infty$. These results allow us to use closed contours that are very large and know that certain parts of the integration (the large circular sections) vanish. Then, if there are poles within the contour region we can be confident that the nonzero integrations come from the noncircular

parts of the contour.

Example

Consider the real integral

$$\int_{-\infty}^{\infty} dx \frac{\sin x}{x} \ , \tag{40.27}$$

We might first consider the related complex integral

$$\oint_C dz \, \frac{\sin z}{z} \ . \tag{40.28}$$

Does this have any poles? It looks like it might have one at $z = 0$. However, we can find the residue there by evaluating

$$\text{Res} = \lim_{z \to 0} z \frac{\sin z}{z} = 0 \ . \tag{40.29}$$

This shows that there is no residue of $z = 0$ and hence all contour integrals vanish. If we consider a contour that traverses the real line from $-R$ to R and then closes in the upper half-plan with a semi-circular arc or radius R the we might write

$$0 = \oint_C dz \, \frac{\sin z}{z} = \int_{-R}^{R} dx \frac{\sin x}{x} + \int_{C_R} dz \frac{\sin z}{z} \ . \tag{40.30}$$

The last term on the right does not satisfy the requirements of Jordan's Lemma, so we do not know what it does as $R \to \infty$.

In order to use Jordan's Lemma we instead consider the complex integral

$$\oint_C dz \, \frac{e^{iz}}{z} \ . \tag{40.31}$$

This has a pole at $z = 0$. We extract the residue by evaluating

$$\text{Res} = \lim_{z \to 0} z \frac{e^{iz}}{z} = 1 \ . \tag{40.32}$$

Now, what contour should we use to isolate the real integral that we want? One option is the contour shown in Fig. 36. In the limit $R \to \infty$ the large circular arc in the upper half-plan goes to zero by Jordan's Lemma. Most of the integral we want then comes from the part of the contour that lies on the real line. However, because the pole at $z = 0$ is also on the real line we must choose a contour that avoids it. In Fig. 36 this is done by making a small circular arc of radius ε that ensures the pole is included within the contour. In order to get to the real integral that we want, we take the limit $\varepsilon \to 0$.

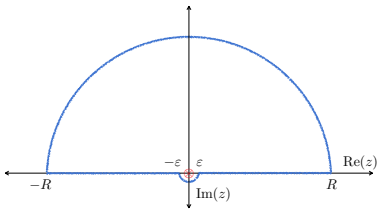


Figure 36. Contour that includes the pole at $z = 0$.

By the integral theorem we know that

$$2\pi i = \oint_C \frac{e^{iz}}{z} = \int_0^\pi iR\,d\theta\, e^{i\theta}\frac{e^{-i\theta}}{R}\exp\left[iRe^{i\theta}\right] + \int_{-R}^{-\varepsilon} dx \frac{\cos x + i\sin x}{x} + \int_\pi^{2\pi} i\varepsilon\,d\theta\, e^{i\theta}\frac{e^{-i\theta}}{\varepsilon}\exp\left[i\varepsilon e^{i\theta}\right] + \int_\varepsilon^R dx\frac{\cos x + i\sin x}{x}\ . \tag{40.33}$$

Note that that limits of integration for the two θ integrals are different. The first begins as $\theta = 0$ and ends at $\theta = \pi$ the second begins at $\theta = \pi$ and ends at $\theta = 2\pi$. Note also that by making the change of variables $x \to -x$ one can show that

$$\int_{-R}^{-\varepsilon} dx\frac{\cos x}{x} = -\int_\varepsilon^R dx\frac{\cos x}{x}\ , \tag{40.34}$$

so the two $\cos x$ parts of the integral cancel. We are left with

$$2\pi = \int_0^\pi d\theta\,\exp\left[iRe^{i\theta}\right] + \int_{-R}^{-\varepsilon} dx\frac{\sin x}{x} + \int_\pi^{2\pi} d\theta\,\exp\left[i\varepsilon e^{i\theta}\right] + \int_\varepsilon^R dx\frac{\sin x}{x}\ , \tag{40.35}$$

for all R and ε. We then take the limit $R \to \infty$ and by Jordan's Lemma the first term goes to zero. We can also take the limit $\varepsilon \to 0$ without introducing any singular behavior because the $\cos x$ terms have already cancelled. This leads us with

$$\begin{aligned} 2\pi =& 0 + \int_{-\infty}^0 dx\frac{\sin x}{x} + \int_\pi^{2\pi} d\theta + \int_0^\infty dx\frac{\sin x}{x} \\ =& 2\pi - \pi + \int_{-\infty}^\infty dx\frac{\sin x}{x}\ . \end{aligned} \tag{40.36}$$

In other words, we have that

$$\int_{-\infty}^\infty dx\frac{\sin x}{x} = \pi\ . \tag{40.37}$$

Exercise 40.3 Evaluate the contour integral for the contour shown in Fig. 37 that does not include the pole at $z = 0$. Show that this contour also leads to

$$\int_{-\infty}^\infty dx\frac{\sin x}{x} = \pi\ . \tag{40.38}$$

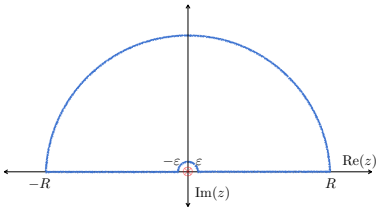


Figure 37. Contour that excludes the pole at $z = 0$.

Exercise 40.4 Use a contour whose large circular section is in the *lower* half of the

complex plan to show that

$$\int_{-\infty}^{\infty} dx \frac{\sin x}{x} = \pi \ . \tag{40.39}$$

Section 41

Green Functions

Green function methods are quite useful in solving difficult differential equations. They are often the place to start when using perturbative methods. In this section we give only the smallest introduction and taste of what they are and how they can be used.

The general idea is simply that often in physics we come across equations of the form

$$\mathcal{D}_x f(x) = S(x) \ , \tag{41.1}$$

where $\mathcal{D}_x$ is some differential operator, $f(x)$ is the trajectory or field configuration we want to know, and $S(x)$ is some kind of source. If you showed this problem to sufficiently confident young students, who had learned some algebra but did not yet know differential equations, and asked them to solve for $f(x)$ they might come up with

$$f(x) = \frac{1}{\mathcal{D}_x} S(x) \ . \tag{41.2}$$

This answer is essentially correct, once we have determined how to make sense of

$$\frac{1}{\mathcal{D}_x} \ . \tag{41.3}$$

It is this inverse of a differential operator that we call a Green function.

Let us try this out on a very simple differential operator: a single derivative

$$\frac{d}{dx} f(x) = S(x) \ . \tag{41.4}$$

To solve this equation for $f(x)$ we simply integrate

$$\int^x dx' \, \frac{d}{dx'} f(x') = \int^x dx' \, S(x') \ . \tag{41.5}$$

The result is

$$f(x) = \int^x dx' \, S(x') + c_0 \ , \tag{41.6}$$

where c_0 is some constant that can be determined from some boundary condition for $f(x)$.

Let's write this same process in a different way that generalizes. We write

$$f(x) = \int dx' \, G(x, x') S(x') \ . \tag{41.7}$$

In this equation the Green function $G(x, x')$ must be a function of both x and the integration variable x'. This formal solution needs to produce Eq. (41.4) so we have

$$\frac{d}{dx} f(x) = \int dx' \, S(x') \frac{d}{dx} G(x, x') = S(x) \ , \tag{41.8}$$

This shows that our Green function must satisfy

$$\frac{d}{dx} G(x, x') = \delta(x - x') \ . \tag{41.9}$$

This implies that the Green function is not continuous at $x = x'$. In fact, by integrating over a small region we find

$$\begin{aligned} \int_{x'-\epsilon}^{x'+\epsilon} dx \frac{d}{dx} G(x, x') &= \int_{x'-\epsilon}^{x'+\epsilon} dx \delta(x - x') \\ G(x' + \epsilon, x') - G(x' - \epsilon, x') &= 1 \ . \end{aligned} \tag{41.10}$$

This holds in the limit of $\epsilon \to 0$ showing that the Green function is discontinuous by exactly one at $x = x'$.

Finally, let's consider boundary conditions. Suppose that $f(x)$ is determined by a Dirichlet boundary condition

$$f(x_0) = c \ . \tag{41.11}$$

If we impose this upon the formal solution in Eq. (41.7) we find

$$f(x_0) = c = \int dx' \, G(x_0, x') S(x') \ . \tag{41.12}$$

Example

Now, let's try this out on a very simple problem. Newton's second law relates the momentum as a function of time $p(t)$ to the acceleration $a(t)$:

$$\frac{dp}{dt} = a(t) \ . \tag{41.13}$$

We would like to determine how the momentum evolves with time $p(t)$ given an initial momentum of $p(0) = p_0$.

The Green function is defined by

$$\frac{d}{dt} G(t, t') = \delta(t - t') \ . \tag{41.14}$$

For all $t \neq t'$ this is simply

$$\frac{d}{dt}G(t,\, t') = 0 \ , \tag{41.15}$$

which has the solution

$$G(t,\, t') = \begin{cases} c_<(t') & t < t' \\ c_>(t') & t > t' \end{cases} \ . \tag{41.16}$$

The discontinuity relation in Eq. (41.10) implies that

$$c_>(t') - c_<(t') = 1 \ . \tag{41.17}$$

This leads to

$$G(t,\, t') = \begin{cases} c_<(t') & t < t' \\ c_<(t') + 1 & t > t' \end{cases} \ . \tag{41.18}$$

The full solution is then

$$\begin{aligned} p(t) &= \int_0^t dt' \, [1 + c_<(t')] \, a(t') + \int_t^\infty dt' \, c_<(t') \, a(t') \\ &= \int_0^t dt' \, a(t') + \int_0^\infty dt' \, c_<(t') \, a(t') \ . \end{aligned} \tag{41.19}$$

Our initial condition is

$$p_0 = p(0) = 0 + \int_0^\infty dt' \, c_<(t') \, a(t') \ , \tag{41.20}$$

so we can write the full solution as

$$p(t) = \int_0^t dt' \, a(t') + p_0 \ . \tag{41.21}$$

Of course, for this simple case we might have obtained this solution by simply integrating the initial equation. But this still provides simple illustration of the methods.

Exercise 41.1 Use the result in Eq. (41.21) to determine the the momentum as a function of time for a constant acceleration

$$a(t) = a_0 \ , \tag{41.22}$$

and for an exponentially falling acceleration

$$a(t) = a_0 e^{-t/\tau} \ . \tag{41.23}$$

Do the results match your intuition for what you expect to occur?

Example

A slightly less trivial example is the simple harmonic system

$$\left(\frac{d^2}{dt^2} + \omega_0^2\right) x(t) = f(t) \ . \tag{41.24}$$

In this case the Green function satisfies

$$\left(\frac{d^2}{dt^2} + \omega_0^2\right) G(t,t') = \delta(t-t') \ . \tag{41.25}$$

Let's use this starting point to investigate what happens near $t = t'$. We integrate in t from $t' - \epsilon$ to $t + \epsilon$ and then take the limit of $\epsilon \to 0$. This produces

$$\begin{aligned} &\lim_{\epsilon\to 0} \int_{t'-\epsilon}^{t'+\epsilon} dt' \left(\frac{d^2}{dt^2} + \omega_0^2\right) G(t,t') = \lim_{\epsilon\to 0} \int_{t'-\epsilon}^{t'+\epsilon} dt'\, \delta(t-t') \\ &\lim_{\epsilon\to 0} \left[\left.\frac{dG}{dt}\right|_{t=t'+\epsilon} - \left.\frac{dG}{dt}\right|_{t=t'-\epsilon} \right] + \omega_0^2 \lim_{\epsilon\to 0} \int_{t'-\epsilon}^{t'+\epsilon} dt' G(t,t') = 1 \ . \end{aligned} \tag{41.26}$$

We see that at least one of the terms on the left-hand side must be nonzero.

If the second term were nonzero then the Green function itself would have a discontinuity at $t = t'$ which would tend to imply its derivative (the first term) would also be nonzero. Another, simpler, possibility is that the Green function is continuous, but there is a discontinuity in its first derivative. In this case we find

$$\lim_{\epsilon\to 0} \left[\left.\frac{dG}{dt}\right|_{t=t'+\epsilon} - \left.\frac{dG}{dt}\right|_{t=t'-\epsilon} \right] = 1 \ . \tag{41.27}$$

That is, we take the Green function to be continuous at $t = t'$ and to have a discontinuity in its first derivative at $t = t'$ as given above.

Away from $t = t'$ we simply solve the homogeneous equation. This produces

$$G(t,\, t') = \begin{cases} A(t')e^{i\omega_0 t} + B(t')e^{-i\omega_0 t} & t < t' \\ C(t')e^{i\omega_0 t} + D(t')e^{-i\omega_0 t} & t > t' \end{cases} \ . \tag{41.28}$$

From continuity at $t = t'$ we find

$$A(t')e^{i\omega_0 t'} + B(t')e^{-i\omega_0 t'} = C(t')e^{i\omega_0 t'} + D(t')e^{-i\omega_0 t'} \ , \tag{41.29}$$

which can be written as

$$\left[A(t') - C(t')\right] e^{i\omega_0 t'} = \left[D(t') - B(t')\right] e^{-i\omega_0 t'} \ . \tag{41.30}$$

The discontinuity relation in Eq. (41.27) produces

$$i\omega_0 \left[C(t') - A(t')\right] e^{i\omega_0 t'} - i\omega_0 \left[D(t') - B(t')\right] e^{-i\omega_0 t'} = 1 \ . \tag{41.31}$$

Combining these two we find

$$2i\omega_0 \left[C(t') - A(t')\right] e^{i\omega_0 t'} = 1 \ , \tag{41.32}$$

or

$$C(t') = A(t') + \frac{e^{-i\omega_0 t'}}{2i\omega_0} \ . \tag{41.33}$$

Similarly, we find

$$D(t') = B(t') - \frac{e^{i\omega_0 t'}}{2i\omega_0} \ . \tag{41.34}$$

Putting this all together we may write the Green function as

$$\begin{aligned} G(t,t') =& A(t')e^{i\omega_0 t} + B(t')e^{-i\omega_0 t} + \frac{\Theta(t-t')}{2i\omega_0}\left(e^{i\omega_0(t-t')} - e^{-i\omega_0(t-t')}\right) \\ =& A(t')e^{i\omega_0 t} + B(t')e^{-i\omega_0 t} + \frac{\Theta(t-t')}{\omega_0}\sin\left[\omega_0(t-t')\right] \ , \end{aligned} \tag{41.35}$$

where the Heavyside function is defined to be

$$\Theta(x) = \begin{cases} 1 & x > 0 \\ 0 & x < 0 \end{cases} \tag{41.36}$$

The function we wish to find is then

$$\begin{aligned} x(t) = & \int_0^\infty dt' \left(A(t')e^{i\omega_0 t} + B(t')e^{-i\omega_0 t}\right) f(t') \\ & + \frac{1}{\omega_0}\int_0^t dt' \ \sin\left[\omega_0(t-t')\right] f(t') \ . \end{aligned} \tag{41.37}$$

Note that

$$x_0 = x(0) = \int_0^\infty dt' \left[A(t') + B(t')\right] f(t') \ , \tag{41.38}$$

and

$$v_0 = \left.\frac{dx}{dt}\right|_{t=0} = i\omega_0 \int_0^\infty dt' \left[A(t') - B(t')\right] f(t') \ . \tag{41.39}$$

Make sure you understand why the second term does not contribute. These two results imply that

$$\int_0^\infty dt' \, A(t') f(t') = \frac{x_0 - iv_0/\omega_0}{2} \qquad \int_0^\infty dt' \, B(t') f(t') = \frac{x_0 + iv_0/\omega_0}{2} \ . \tag{41.40}$$

Finally, we arrive at

$$\begin{aligned} x(t) =& \frac{e^{i\omega_0 t}}{2}\left(x_0 + \frac{v_0}{i\omega_0}\right) + \frac{e^{-i\omega_0 t}}{2}\left(x_0 - \frac{v_0}{i\omega_0}\right) \\ &+ \frac{1}{\omega_0}\int_0^t dt' \, \sin\left[\omega_0(t-t')\right] f(t') \\ =& x_0 \cos(\omega_0 t) + \frac{v_0}{\omega_0}\sin(\omega_0 t) + \frac{1}{\omega_0}\int_0^t dt' \, \sin\left[\omega_0(t-t')\right] f(t') \ . \end{aligned} \tag{41.41}$$

Note that without a driving function $f(t) = 0$ we simply recover the usual harmonic motion.

What is more useful is that we can also include the effects of any type of driving. Let's consider the case of sinusoidal driving at some frequency Ω

$$f(t) = \kappa \sin(\Omega t) \ . \tag{41.42}$$

The driving term in $x(t)$ is then

$$\frac{\kappa}{\omega_0}\int_0^t dt' \, \sin\left[\omega_0(t-t')\right]\sin(\Omega t') = \frac{\kappa}{\omega_0}\frac{\Omega\sin(\omega_0 t) - \omega_0 \sin(\Omega t)}{\Omega^2 - \omega_0^2} \ . \tag{41.43}$$

Note that when $\Omega \to \omega_0$ this term grows linearly with time. That is, driving this system at its characteristic frequency ω_0 leads to a very large response. This is an example of a resonance.

The previous example also makes connection with the form of many Green functions that appear in quantum field theory. Suppose we define the function $x(t)$ in terms of its Fourier transform $\widetilde{x}(\omega)$ by

$$x(t) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} e^{-i\omega t}\widetilde{x}(\omega) \ . \tag{41.44}$$

We can similarly write the source term $f(t)$ in terms of its Fourier transform $\widetilde{f}(\omega)$. The differential equation in (41.24) becomes

$$\left(-\omega^2 + \omega_0^2\right)\widetilde{x}(\omega) = \widetilde{f}(\omega) \ . \tag{41.45}$$

It is then algebraically simple to write

$$\widetilde{x}(\omega) = \frac{\widetilde{f}(\omega)}{\omega_0^2 - \omega^2} \ . \tag{41.46}$$

Similarly, the Green function $\widetilde{G}(\omega)$ can be defined by

$$G(t,t') = \int \frac{d\omega}{2\pi} e^{-i(t-t')\omega}\widetilde{G}(\omega) \ . \tag{41.47}$$

Because the delta function can be written as

$$\delta(t-t') = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\, e^{-i\omega(t-t')} \ , \tag{41.48}$$

we find that the equation that defines the Green function (41.25) becomes the frequency space equation

$$\left(-\omega^2+\omega_0^2\right)\widetilde{G}(\omega) = 1 \ . \tag{41.49}$$

In other words, the frequency space Green function is

$$\widetilde{G}(\omega) = \frac{1}{\omega_0^2-\omega^2} = \frac{1}{2\omega_0}\left(\frac{1}{\omega_0-\omega}+\frac{1}{\omega_0+\omega}\right) \ . \tag{41.50}$$

This shows that simply dividing by the transformed differential operator in Eq. (41.46) does, in fact, lead to the correct Green function.

We can then find the Green function in time be evaluating

$$G(t,t') = \int \frac{d\omega}{2\pi} e^{-i(t-t')\omega}\frac{1}{\omega_0^2-\omega^2} \ . \tag{41.51}$$

Unfortunately, this integral has a problem. There are singularities at $\omega = \pm\omega_0$. Otherwise, it looks like an integral that we might evaluate using contour integration. Consider for instance

$$\int d\omega\, e^{-i(t-t')\omega}\frac{1}{\omega_0-\omega} \to \oint_C dz\, e^{-i(t-t')z}\frac{1}{\omega_0-z} \ . \tag{41.52}$$

This function has the right form to use Jordan's Lemma, allowing us to drop circular sections of the contour as $R\to\infty$. Note however that we need to use a contour in the lower half plane when $t>t'$—see (40.26)—and a contour in the upper half plane when $t<t'$, see (40.24).

To evaluate this integral we might try a contour that avoids the pole $z=\omega_0$ by simply going around it in a small semi-circle of radius ε. However, this is tricky to evaluate because the semi-circle is not centered at the origin. Instead, we modify the function slightly. For instance,

$$\oint_C dz\, e^{-i(t-t')z}\frac{1}{\omega_0+i\varepsilon-z} \ , \tag{41.53}$$

has a pole at $z=\omega_0+i\varepsilon$, where we assume $0<\varepsilon\ll 1$. The residue of this pole is

$$\mathrm{Res}(\omega_0+i\varepsilon) = e^{-i\omega_0(t-t')}e^{-\varepsilon(t'-t)} \ . \tag{41.54}$$

This implies that

$$\int_{-\infty}^{\infty} d\omega\, e^{-i\omega(t-t')}\frac{1}{\omega_0+i\varepsilon-\omega} = 2\pi i\Theta(t'-t)e^{-i\omega_0(t-t')}e^{-\varepsilon(t'-t)} \ , \tag{41.55}$$

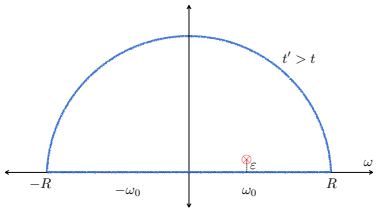


Figure 38. Contour in the upper half plane for pole at $z=\omega+i\varepsilon$. This yields a nonzero result.

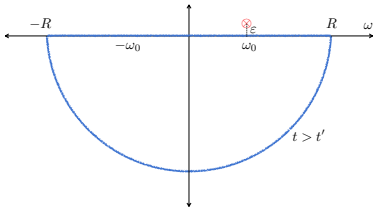


Figure 39. Contour in the lower half plane for pole at $z=\omega+i\varepsilon$. This integral is zero because there is no pole contained within the contour.

there the $\Theta(t'-t)$ appears because the pole in only included within the contour when we use the upper half plane contour, which is when $t' > t$. Similarly, we find

$$\int_{-\infty}^{\infty} d\omega\, e^{-i\omega(t-t')} \frac{1}{\omega_0 - i\varepsilon - \omega} = 2\pi i \Theta(t-t') e^{-i\omega_0(t-t')} e^{-\varepsilon(t-t')} \ , \quad (41.56)$$

because the pole is on the lower half plane in this case.

Exercise 41.2 Show that for the pole at $\omega = -\omega_0$ that

$$\int_{-\infty}^{\infty} d\omega\, e^{-i\omega(t-t')} \frac{1}{\omega_0 + i\varepsilon + \omega} = 2\pi i \Theta(t-t') e^{i\omega_0(t-t')} e^{-\varepsilon(t-t')} \quad (41.57)$$

$$\int_{-\infty}^{\infty} d\omega\, e^{-i\omega(t-t')} \frac{1}{\omega_0 - i\varepsilon + \omega} = 2\pi i \Theta(t'-t) e^{i\omega_0(t-t')} e^{-\varepsilon(t'-t)} \ . \quad (41.58)$$

Recall that these two terms came from one denominator

$$\frac{1}{\omega_0^2 - \omega^2} \ . \quad (41.59)$$

This means we should either take $+i\varepsilon$ for both poles or take $-i\varepsilon$ for both. In effect, we define two Green functions

$$G_{\pm}(t,t') = \int \frac{d\omega}{2\pi} e^{-i(t-t')\omega} \frac{1}{(\omega_0 \pm i\varepsilon)^2 - \omega^2} \ . \quad (41.60)$$

Our calculations have shown that

$$\begin{aligned} G_+(t,t') =& \frac{i}{2\omega_0} \left[\Theta(t'-t) e^{-i\omega_0(t-t')} + \Theta(t-t') e^{i\omega_0(t-t')} \right] \quad (41.61) \\ =& \frac{i}{2\omega_0} \cos\left[\omega_0(t-t')\right] + \frac{1}{2\omega_0} \sin\left[\omega_0(t-t')\right] \left[\Theta(t'-t) - \Theta(t-t')\right] \\ G_-(t,t') =& \frac{i}{2\omega_0} \left[\Theta(t'-t) e^{i\omega_0(t-t')} + \Theta(t-t') e^{-i\omega_0(t-t')} \right] \quad (41.62) \\ =& \frac{i}{2\omega_0} \cos\left[\omega_0(t-t')\right] - \frac{1}{2\omega_0} \sin\left[\omega_0(t-t')\right] \left[\Theta(t'-t) - \Theta(t-t')\right] \end{aligned}$$

where we have taken $\epsilon \to 0$.

Exercise 41.3 Show the steps the go form the first equality to the second in the above descriptions of $G_{\pm}(t,t')$.

We can then find the position as a function of time, given a driving function $f(t)$ as

$$x_{\pm}(t) = \int_0^{\infty} dt'\, G_{\pm}(t,t') f(t') \ . \quad (41.63)$$

This leads to

$$x_\pm(t) = \frac{i}{2\omega_0} e^{\mp i\omega_0 t} \int_t^\infty dt'\, e^{\pm i\omega_0 t'} f(t') + \frac{i}{2\omega_0} e^{\pm i\omega_0 t} \int_0^t dt'\, e^{\mp i\omega_0 t'} f(t') \ , \tag{41.64}$$

and

$$\dot{x}_\pm(t) = \pm\frac{1}{2} e^{\mp i\omega_0 t} \int_t^\infty dt'\, e^{\pm i\omega_0 t'} f(t') \mp \frac{1}{2} e^{\pm i\omega_0 t} \int_0^t dt'\, e^{\mp i\omega_0 t'} f(t') \ . \tag{41.65}$$

This means the initial values are

$$x_\pm(0) = \frac{i}{2\omega_0} \int_0^\infty dt'\, e^{\pm i\omega_0 t'} f(t') \ , \qquad \dot{x}_\pm(0) = \pm\frac{1}{2} \int_0^\infty dt'\, e^{\pm i\omega_0 t'} f(t') \ . \tag{41.66}$$

This clarifies the meaning of the different sign possibilities. They correspond to different boundary conditions, or different initial values. In this case they denote an object that is moving in the positive or negative direction when $t = 0$.

Exercise 41.4 Show that, using the definitions $x_\pm(0) \equiv x_0$ and $\dot{x}_\pm(0) \equiv v_0$ that

$$x_\pm(t) = x_0 \cos(\omega_0 t) + \frac{v_0}{\omega_0} \sin(\omega_0 t) \mp \frac{1}{\omega_0} \int_0^t dt' \sin\left[\omega_0 (t - t')\right] f(t') \ . \tag{41.67}$$

As the exercise above shows, the Fourier transform method produces the same results as in the direct calculation case. Often one or the other is better suited for a specific problem. For this simple system we can calculate in both ways and show that they match exactly.

Example An important example is the Helmholtz operator $\nabla^2 + k_0^2$. We would like to find the Green function that satisfies the equation

$$\left(\nabla^2 + k_0^2\right) G(\vec{x}, \vec{x}') = \delta^3(\vec{x} - \vec{x}') \ . \tag{41.68}$$

Our first step is to consider the Fourier transform of $G(\vec{x}, \vec{x}')$

$$G(\vec{x}, \vec{x}') = \int \frac{d^3k}{(2\pi)^3} G(\vec{k}) e^{i(\vec{x} - \vec{x}')\cdot\vec{k}} \ . \tag{41.69}$$

When this is inserted into the Helmholtz equation we find

$$(-k^2 + k_0^2) G(\vec{k}) = 1 \ , \tag{41.70}$$

where $\vec{k} \cdot \vec{k} = k^2$. In other words, we see that

$$G(\vec{k}) = \frac{-1}{k^2 - k_0^2} \ . \tag{41.71}$$

We then find the Green function is

$$G(\vec{x},\,\vec{x}') = -\int \frac{d^3k}{(2\pi)^3}\frac{e^{i(\vec{x}-\vec{x}')\cdot\vec{k}}}{k^2-k_0^2}\ . \tag{41.72}$$

To evaluate this integral we first orient the vector $\vec{k}$ so that $\vec{x}-\vec{x}'$ lies along the polar axis of spherical polar coordinates. Then the result

$$(\vec{x}-\vec{x}')\cdot\vec{k} = |\vec{x}-\vec{x}'|\,k\cos\theta\ , \tag{41.73}$$

uses the same θ as the measure of the integral

$$d^3k = dk\,d(\cos\theta)\,d\phi\,k^2\ . \tag{41.74}$$

This means that we have

$$G(\vec{x},\,\vec{x}') = \frac{-1}{(2\pi)^3}\int_0^\infty dk\,\frac{k^2}{k^2-k_0^2}\int_{-1}^1 d(\cos\theta)e^{i|\vec{x}-\vec{x}'|k\cos\theta}\int_0^{2\pi}d\phi \tag{41.75}$$

The integral over ϕ simply produces a factor of 2π. The integral over $\cos\theta$ can be evaluated exactly to produce

$$\begin{aligned}
G(\vec{x},\,\vec{x}') &= \frac{-1}{(2\pi)^2}\int_0^\infty dk\,\frac{k^2}{k^2-k_0^2}\frac{1}{i|\vec{x}-\vec{x}'|k}\left(e^{i|\vec{x}-\vec{x}'|k}-e^{-i|\vec{x}-\vec{x}'|k}\right)\\
&= \frac{-1}{(2\pi)^2 i|\vec{x}-\vec{x}'|}\left[\int_0^\infty dk\,\frac{k^2}{k^2-k_0^2}\frac{1}{k}e^{i|\vec{x}-\vec{x}'|k}-\int_0^\infty dk\,\frac{k^2}{k^2+k_0^2}\frac{1}{k}e^{-i|\vec{x}-\vec{x}'|k}\right]\\
&= \frac{-1}{(2\pi)^2 i|\vec{x}-\vec{x}'|}\left[\int_0^\infty dk\,\frac{k}{k^2-k_0^2}e^{i|\vec{x}-\vec{x}'|k}+\int_{-\infty}^0 dk\,\frac{k}{k^2-k_0^2}e^{i|\vec{x}-\vec{x}'|k}\right]\\
&= \frac{-1}{(2\pi)^2 i|\vec{x}-\vec{x}'|}\int_{-\infty}^\infty dk\,\frac{k}{k^2-k_0^2}e^{i|\vec{x}-\vec{x}'|k}\ ,
\end{aligned} \tag{41.76}$$

where in the second to last line we have taken $k\to -k$ in the second integral.

The final integral that we have found satisfies the requirements of Jordan's lemma (40.24), where we close the contour in the upper half of the complex plane. This means that we can use the residue theorem to evaluate the integral. One complication, however, is the poles are at $k=\pm k_0$, which is on the contour. To deal with this we allow k_0 to have a small imaginary part, $k=\pm k_0 \pm i\varepsilon$. We note that

$$\frac{k}{k^2-(k_0\pm i\varepsilon)^2} = \frac{k}{2k}\left(\frac{1}{k+k_0\pm i\varepsilon}+\frac{1}{k-k_0\mp i\varepsilon}\right)\ . \tag{41.77}$$

From this we see that there are two possibilities for having a pole in the

upper half plane. When we take $k = -k_0 + i\varepsilon$ the residue of the pole is

$$\frac{1}{2} e^{-i|\vec{x}-\vec{x}'|k_0} \ , \tag{41.78}$$

where we have safely taken the $\epsilon \to 0$ limit. When we instead take $k = k_0 + i\varepsilon$ the residue is

$$\frac{1}{2} e^{i|\vec{x}-\vec{x}'|k_0} \ . \tag{41.79}$$

Using the residue theorem we then have two results

$$\begin{aligned} G_{\pm}(\vec{x}, \vec{x}') &= \frac{-1}{(2\pi)^2 i|\vec{x}-\vec{x}'|} 2\pi\, i \frac{1}{2} e^{\pm i|\vec{x}-\vec{x}'|k_0} \\ &= -\frac{1}{4\pi} \frac{1}{|\vec{x}-\vec{x}'|} e^{\pm i|\vec{x}-\vec{x}'|k_0} \ . \end{aligned} \tag{41.80}$$

These two results correspond to two different boundary conditions. The function $G_+(\vec{x}, \vec{x}')$ is associated with outgoing spherical waves while $G_-(\vec{x}, \vec{x}')$ is associated with ingoing spherical waves.

Exercise 41.5 Follow a similar method to the example above to find the Green function for the differential operator

$$\nabla^2 - m^2 \ . \tag{41.81}$$

Show that there is only a single Green function

$$G(\vec{x}, \vec{x}') = -\frac{1}{4\pi} \frac{1}{|\vec{x}-\vec{x}'|} e^{-|\vec{x}-\vec{x}'|m} \ . \tag{41.82}$$

Note that in the $m \to 0$ limit this implies the famous result

$$\nabla^2 \frac{1}{|\vec{x}-\vec{x}'|} = -4\pi\delta^3(\vec{x}-\vec{x}') \ . \tag{41.83}$$

A generalization of the Laplacian that leads to wave solutions is the D'Alembertian

$$\partial_t^2 - \nabla^2 \ . \tag{41.84}$$

What we are interested is is the Green function that depends on both space and time $G(x, x')$ with

$$\left(\partial_t^2 - \nabla^2 + m^2\right) G(x, x') = \delta^4(x - x') \ , \tag{41.85}$$

where x stands in for the full t, x, y, z. It is again useful to consider the

Fourier transform

$$G(x,x') = \int \frac{d^4p}{(2\pi)^4} G(p) e^{-ip_t t + i\vec{x}\cdot\vec{p}} \ , \tag{41.86}$$

which leads to

$$\left(-p_t^2 + p^2 + m^2\right) G(p) = 1 \ , \tag{41.87}$$

or

$$G(p) = \frac{1}{-p_t^2 + p^2 + m^2} \ . \tag{41.88}$$

To find the Green function in position space we evaluate

$$\begin{aligned} G(x,x') &= \int \frac{d^3p}{(2\pi)^3} e^{i(\vec{x}-\vec{x}')\cdot\vec{p}} \int \frac{dp_t}{2\pi} \frac{e^{-ip_t(t-t')}}{p^2 + m^2 - p_t^2} \qquad (41.89) \\ &= \int \frac{d^3p}{(2\pi)^3} e^{i(\vec{x}-\vec{x}')\cdot\vec{p}} \int \frac{dp_t}{2\pi} \frac{e^{-ip_t(t-t')}}{2\sqrt{p^2+m^2}} \left(\frac{1}{\sqrt{p^2+m^2} - p_t} + \frac{1}{\sqrt{p^2+m^2} + p_t} \right) \end{aligned}$$

The inner integral looks like the one in Eq (41.51) and can be evaluated in exactly the same way. We again shift the poles off the real axis and find

$$\begin{aligned} G_\pm(x,x') = \int \frac{d^3p}{(2\pi)^3} e^{i(\vec{x}-\vec{x}')\cdot\vec{p}} \frac{i}{2\sqrt{p^2+m^2}} \qquad (41.90) \\ \left[\theta(t'-t) e^{\pm i\sqrt{p^2+m^2}(t'-t)} + \theta(t-t') e^{\pm i\sqrt{p^2+m^2}(t-t')} \right] \ . \end{aligned}$$

Notice, that the integral would vanish if not for the singularities, the poles in the momentum. Therefore, it is the values

$$p_t^2 = p^2 + m^2 \ , \tag{41.91}$$

that dominate this integral.

Section 42

Statistics and Uncertainty

In this section we provide a very brief introduction of some basic statistical ideas. As with most of these reviews there is a great deal which is not covered. Further information can be found in [93] and a very nice treatment focused on collider experiments is found in [7].

Much of this text describes detailed mathematical models. In order to gain confidence that these models have something to do with the actual structure of nature we must compare these to experimental data. In order to understand how such data may or may not support a given model we must understand something of statistics. As with many other sections in

this text, this is the briefest of introductions and should be supplemented by further study.

A central issue of comparing experimental data with a physical model is uncertainty. While models are typically exact in their predictions the realities of measurement are not so simple. Any measurement can only provide limits on a "true" value of something. For instance, if you measure your weight by standing on a scale the scale gives you back a number. We often take this number to be our true weight. Most scales used for this purpose, however, are only accurate to some fraction of a kilogram.[3] So, while a true weight might be

$$77.085093831484846058282775276072537649857336\ldots \text{ kg}\,. \tag{42.1}$$

the scale might simply display a weight of 77.1 kg. A more useful recording of this might be 77.1 ± 0.05 kg. This indicates that the scale is confident that the true value is between 77.15 kg and 77.05 kg, but cannot be any more precise than that. This ±0.05 kg is admission that the measurement is uncertain. Note that uncertainty is not the same as saying we know nothing, rather it is a way of being honest about the limits of what we know.

Of course, there could be other issues that confound this measurement. The scale might not be calibrated correctly, making it read 5 kgs too high or 3 kg too low. This is part of getting the machine to work right. But even when working correctly, there is an intrinsic limit to how precise a measurement can be. This is an example of a **systematic uncertainty**. It is an uncertainty related to the actual measurement device. These are difficult to determine and part of the hard work of doing real experiments.

Another source of uncertainty can come from asking different types of questions. Instead of "What is my weight?" one could ask "Is this a fair coin?" What does it mean to be a fair coin? Although classical physics is perfectly deterministic (the outcomes are predicted exactly by the initial conditions) the variations in the initial conditions that lead to different outcomes in a coin flip are so small, smaller than a flipper can control, the outcomes are random for all practical purposes. It is perhaps worth emphasizing that this is qualitatively different from the uncertainties inherent in quantum mechanics. Within standard interpretations, the probabilistic nature of quantum measurements are simply a part of nature, the different outcomes simply are not fully determined by the initial state.

In the case of a classical coin, the system is assumed to be fully determined by the initial state, but we do not have the ability to reliably replicate that state. Therefore, a fair coin is one for which the probabil-

[3]Some readers may complain that kilograms are a measure of mass, not weight. I am using kilograms because they are somewhat more familiar than Newtons as far as the reading of home scales are concerned.

ity of coming up heads is equal to the probability of coming up tails, and both of these probabilities are 50%. Of course, there is a third option, that the coin lands on edge and this must truly be a possibility for a three-dimensional coin. But, in practice, this probability is so small that it is neglected.

In short, the fair coin is an idealized model. And it is reasonable to compare that model to any particular coin. How might we do this? We can flip the coin once, but that single piece of data is insufficient to address the question with any confidence. If the coin comes up heads then I cannot tell if it is fair or always comes up heads. What we are really after is a statement about many coin flips. While the above definition of what it means to be fair is in terms of a probability, that either heads or tails is equally likely for one flip, checking whether or not a coin is fair requires many flips.

In Fig. 40 we show what this coin flip data might look like. After 10 flips (top panel) one might record 4 heads and 6 tails, so the percentages are 40% and 60%, respectively. Is this evidence that the coin is fair? After 100 flips (middle panel) there are 55 heads and 45 tails, or percentages of 55% and 45%. Performing 1000 flips (bottom panel) produces 495 heads and 505 tails, or percentages of 49.5% and 50.5%. None of these percentages are equal to the perfect fair coin, so what can we conclude?

The formal idea is that if we could perform an infinite number of coin flips the percentages would converge to exactly 50% for each outcome. No finite number of flips can ever prove that a coin is fair, but after some finite number we can ask how probable our data are if the true model is the fair coin. This is done by first calculating the absolute probability of the outcomes. This can be done by way of the **binomial distribution**. If the true probability of an event occurring is P (which means that probability that event does not occur is $1-P$) then the probability $p_{k,N}$ that the event occurs k times in N trials is

$$p_{k,N} = P^k(1-P)^{N-k}\frac{N!}{k!(N-k)!} \ . \tag{42.2}$$

Examples of distributions for finding a certain number of heads flips for 10, 100, and 1000 coin flips are given in Fig. 41.

Notice that the absolute probability for most likely scenario gets smaller as the number of trials goes up. This is simply a consequence of their being more possible outcomes, each of which has some nonzero probability. So the absolute probability of a given outcome is not so useful in checking how likely our results are if we are assuming a fair coin model. What is more useful is asking how far away our results are from the mean (or average) of the distribution.

As is obvious from the figure, the mean of the binomial distribution is

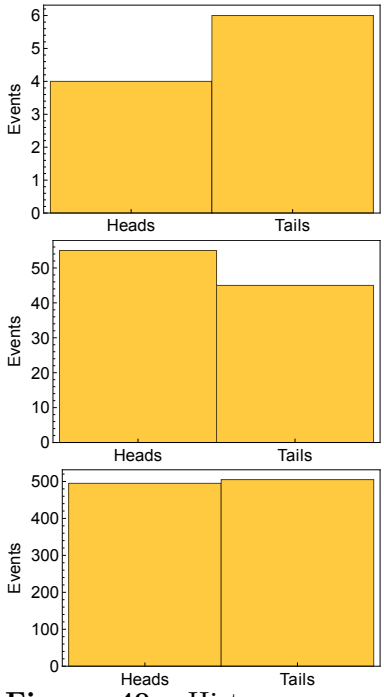


Figure 40. Histograms of simulated coin flip data for 10 (top), 100 (middle) and 1000 (bottom) total flips.

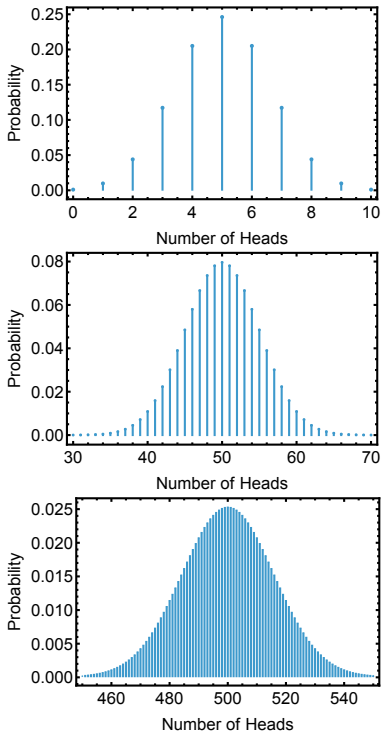


Figure 41. Probability distributions that a certain number of heads flips occur for 10 (top), 100 (middle) and 1000 (bottom) total flips.

simply

$$\overline{k}_N = NP \ . \tag{42.3}$$

So, for 10 throws the average number of heads is 5. But how can we quantify what it means to be close or far from this average? This is done by way of the **standard deviation**, which is a measure of how peaked the distribution is about the average and typically denoted by σ. Sometimes the **variance** (which is simply σ^2) is used in a similar way, but we only discuss the standard deviation in this section. For the binomial distribution the standard deviation is

$$\sigma_N = \sqrt{NP(1-P)} \ . \tag{42.4}$$

So, for a given experimental data set, we can compute how many standard deviations our results are from the mean of the distribution. That is, we assume a model, determine the probability distribution of the results, calculate the mean and standard deviation, and then interpret the experimental results.

For example, in the 10 flip case shown in Fig. 40 the standard deviation is $\sigma_{10} = \sqrt{10}/2 \approx 1.6$. The distance of the data (4 heads) from the mean (5 heads) was $0.63\sigma_{10}$, less than one standard deviation away. Similarly, we find that

$$\sigma_{100} = \sqrt{100}/2 = 5 \ , \qquad \sigma_{1000} = \sqrt{1000}/2 \approx 15.8 \ . \tag{42.5}$$

This means that the 100 flip data of 55 heads is exactly one standard deviation away from the mean and the 1000 flip result of 495 heads is only 0.32 standard deviations from the mean. Because the increasing the number of events tends to produce data that is closer (in terms of standard deviations) to the mean we begin to be more confident that the fair coin model is correct.

Suppose we were to compare to an unfair coin model, in which heads has a 60% probability on each flip. In this case the predicted mean value is $\overline{k} = 0.6N$ and the standard deviation is

$$\sigma_N = \sqrt{N \times 0.24} \ . \tag{42.6}$$

Compared to this model the 10 flip data is 1.3 standard deviations away from the mean, the 100 flip data is 1.02 standard deviations away and the 1000 flip data is 6.8 standard deviations away. This shows that increasing the data eventually indicates that this particular unfair coin model is less likely to be the correct one. Our data are, of course, logically possible given such a coin, but this way of considering the data gives us confidence that this unfair coin model is less likely.

Exercise 42.1 Using the data illustrated in Fig. 40 determine the largest deviation from a fair coin model that is still consistent with the 1000 flip data. First use limit of 2 standard deviations to define consistent, then consider 3 standard deviations and finally 5 standard deviations.

Unlike coin flips, we are often interested in physical systems for which there is a continuous probability distribution $f(x)$, where x is the continuous variable that the probability depends on. For instance, suppose we are measuring the angle of deflection of an alpha particle by a gold nucleus. This value could be anything from 0 to 2π radians. But the total probability of scattering through all angles would need to be one so

$$\int_0^{2\pi} d\theta\, f(\theta) = 1 \ . \tag{42.7}$$

In general, our function $f(x)$ must integrate to one over whatever domain of x is allowed. We can then use this same function to determine the probability that the a specific outcome is between two values of x

$$P_{[a,b]} = \int_a^b dx\, f(x) \ . \tag{42.8}$$

Going back to our scattering example, we could ask what the probability is that an alpha particle's deflection angle is between $\pi/6$ radians and $\pi/3$ radians. This would be given by

$$P_{[\pi/6,\pi/3]} = \int_{\pi/6}^{\pi/3} d\theta\, f(\theta) \ . \tag{42.9}$$

Note that this implies that the probability of measuring an exact value, like $\theta = \pi/4$, is zero because the integration region is zero. This is not actually an issue because we can never measure something exactly, we can only bound it. Any bounding region will lead a finite probability.

The coarse grained nature of physical measurements also motivates how continuous probability distributions often appear in physical data. Typically, the range of the continuous parameter is divided up into several subregions, which we call bins. Often these bins are of equal length, but they do not have to be. When the data are recorded the number of events that belong to a particular bin are recorded and displayed as a histogram. The form of the histogram is taken to be an approximation of the true continuous probability distribution.

For example, in Fig. 42 we record two histograms of 100 simulated 100 coin flip runs. That is, we simulated the flipping of 100 coins and recording the number of heads event, and this complete process was done a total to 100 times. In the top histogram the histogram bin width was chosen to be 5 while in the bottom plot the bin width is 2. Both of these histograms

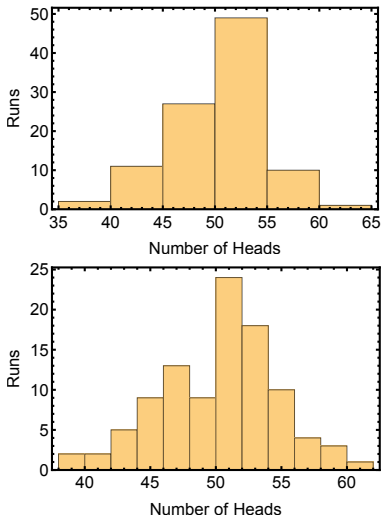


Figure 42. Histograms the number of heads produced in 100 separate 100 coin flip runs. The top (bottom) plot uses a bin width of 5 (2).

can be compared to the middle distribution in Fig. 41. Both have a rough agreement, but neither looks like an exact fit. We can see that the different bin widths affect how the distribution appears.

For comparison, in Fig. 43 we present similar histograms, but in this case 1000 runs of 100 flips each were made. We see that the greater number of runs has allowed both of the distributions to look more like the ideal distribution in Fig. 41. In addition, we find that the smaller bin widths allow the histogram to exhibit more of the shape of the ideal distribution.

In general, a histogram includes N_{total} measurements. In Fig. 42 $N_{\text{total}} = 100$, while in Fig. 43 $N_{\text{total}} = 1000$. These measurements are spread over some number of bins. The number of events in the ith bin is denoted N_i. Of course, the total in all the bins must be the total number of measurements

$$\sum_i N_i = N_{\text{total}} \ , \tag{42.10}$$

which can be rewritten as

$$\sum_i \frac{N_i}{N_{\text{total}}} = 1 \ , \tag{42.11}$$

which reminds us of the integration of the probability distribution, as in (42.7). This form also shows that the probabilities associated with each bin are

$$P_i = \frac{N_i}{N_{\text{total}}} \ . \tag{42.12}$$

The integral, however, includes a measure like dx. This should be associated with bin width, which we might denote Δx. In other words we have

$$\sum_i \Delta x \frac{N_i}{N_{\text{total}} \Delta x} = 1 \ , \tag{42.13}$$

which implies that the value of the probability distribution at some point x_0 is approximately

$$f(x_0) \approx \frac{N_0}{N_{\text{total}} \Delta x} = \frac{P_0}{\Delta x} \ , \tag{42.14}$$

where N_0 is the number of measurements in the bin that contains x_0. Note that this matches the our understanding of $f(x)$ as a probability density, the amount of probability per x. This approximation becomes an equality in the formal limit of $N_{\text{total}} \to \infty$ and $\Delta x \to 0$, while $N_{\text{total}} \Delta x$ remains finite.

Of course, in practice we can never take $N_{\text{total}} \to \infty$. Given a finite N_{total} we might wonder how close our histogram is to the true probability distribution. As long as we can assume that each contribution to the histogram is independent, we can use the binomial distribution to determine the likelihood of some number k events contributing to a specific histogram

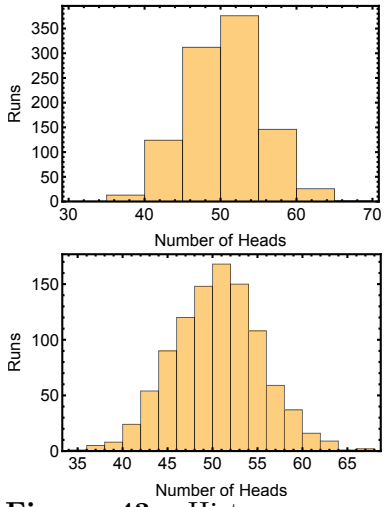


Figure 43. Histograms the number of heads produced in 1000 separate 100 coin flip runs. The top (bottom) plot uses a bin width of 5 (2).

bin. First, we determine the probability to measure a value within the ith bin

$$P_i = \int_{x_i}^{x_i+\Delta x} dx f(x) \ . \tag{42.15}$$

Here x_i is the value of the continuous parameter x at the left edge of the ith bin. Then, the probability of finding $N_i = k$ is

$$p_{i,k} = P_i^k (1-P_i)^{N_{\rm total}-k} \frac{N_{\rm total}!}{k!\,(N_{\rm total}-k)!} \ . \tag{42.16}$$

While this is the correct answer, factorials are often cumbersome to use. A useful approximation to use is the **Poisson distribution** which is the limit of the binomial in the large $N_{\rm total}$ limit. In this limit

$$p_{i,k} \to \frac{(N_{\rm total}P_i)^k}{k!} e^{-N_{\rm total}P_i} \ . \tag{42.17}$$

The mean of the Poisson distribution is

$$\overline{k}_{N_{\rm total}} = N_{\rm total}P_i = N_{\rm total}\frac{N_i}{N_{\rm total}} = N_i \ , \tag{42.18}$$

and the standard deviation is

$$\sigma_{N_{\rm total}} = \sqrt{N_{\rm total}P_i} = \sqrt{N_i} \ . \tag{42.19}$$

This means that, accepting the standard experience that results typically fall within one standard deviation of the mean, that we expect the number of events in a bin will be within $\sqrt{N_i}$ of N_i. Of course, this is not going to be the case for all bins. There will be larger fluctuations from the average expectation with some regularity. Nevertheless, this is a useful heuristic.

We end with some idea of how likely being some number of standard deviations (or some number of σ) away from the predicted average is. We have asserted that being multiple σ away is less likely, but it would be useful to make precise what "less likely" really means. In doing so we use neither that binomial distribution nor the Poisson distribution, but instead the **normal distribution** (or Gaussian distribution). It is true that in the limit of a large number of trials both the binomial and Poisson distributions approach the normal distribution, so it is useful to use it for the general case.

The normal distribution

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\overline{x})^2}{2\sigma^2}\right] \tag{42.20}$$

is completely defined by its mean $\overline{x}$ and standard deviation σ. The proba-

bility of being within $a\sigma$ of the mean is simply

$$P_{a\sigma} = \int_{\overline{x}-a\sigma}^{\overline{x}+a\sigma} dx \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\overline{x})^2}{2\sigma^2}\right] . \tag{42.21}$$

Exercise 42.2 Show by changing variables that

$$P_{a\sigma} = \frac{2}{\sqrt{\pi}} \int_0^{a/\sqrt{2}} dz\, e^{-z^2} . \tag{42.22}$$

The exercise above shows that the probability of being within $a\sigma$ of the mean is equal to $\mathrm{erf}(a/\sqrt{2})$ where the **error function** is defined by

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^{x} dz\, e^{-z^2} . \tag{42.23}$$

Then the probability of being outside of the $a\sigma$ region around the mean is $1 - P_{a\sigma}$. However, what we are typically interested in is not the total probability of being either $a\sigma$ above or $a\sigma$ below. Instead, if there is a measurement that is higher than we expected we want to know only the probability of it being $a\sigma$ high. If it is low we want to know the probability of it being $a\sigma$ low. Therefore, the probability $P_{a\sigma\mathrm{Dev}}$ of seeing a $a\sigma$ deviation is

$$P_{a\sigma\mathrm{Dev}} = \frac{1}{2}\left[1 - \mathrm{erf}\left(\frac{a}{\sqrt{2}}\right)\right] . \tag{42.24}$$

It is these values that we typically assign to observed deviations in histograms or other data. While the error function is not simple to evaluate exactly, it can be done numerically. With this formula we find that 1σ deviations occur 15.9% of the time while 2σ deviation occur 2.3% of the time. Note that this implies that if one looks at hundreds of systems that two sigma deviations are expected to occur several times. A 3σ deviation is less likely, only about 0.13% of the time, while a 4σ deviation occurs in only about 0.0032% of cases. In particle physics the standard threshold used to claim discovery is a 5σ deviation which should only occur by random chance in something like 0.00003% of circumstances.

References

[1] A. Pais, *Inward Bound: Of Matter and Forces in the Physical World.* Oxford University Press, 1986.

[2] L. Hoddeson, L. Brown, M. Riordan and M. Dresden, *The Rise of the Standard Model: A History of Particle Physics from 1964 to 1979.* Cambridge University Press, 1997.

[3] T. Lancaster and S. J. Blundell, *Quantum Field Theory for the Gifted Amateur*. Oxford University Press, 2014.

[4] J. F. Donoghue and L. Sorbo, *A Prelude to Quantum Field Theory*. Princeton University Press, 3, 2022.

[5] A. Lahiri and P. B. Pal, *A First Book of Quantum Field Theory*. Alpha Science International Ltd., 2nd ed., 2005.

[6] P. B. Pal, *An Introductory Course of Particle Physics*. CRC Press, 7, 2014, 10.1201/b17199.

[7] A. J. Larkoski, *Elementary Particle Physics: An Intuitive Introduction*. Cambridge University Press, 6, 2019.

[8] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, *JHEP* **07** (2014) 079 [1405.0301].

[9] L. H. Ryder, *Quantum Field Theory*. Cambridge University Press, 6, 1996, 10.1017/CBO9780511813900.

[10] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory*. Addison-Wesley, Reading, USA, 1995.

[11] S. Weinberg, *The Quantum theory of fields. Vol. 1: Foundations*. Cambridge University Press, 6, 2005, 10.1017/CBO9781139644167.

[12] S. Weinberg, *The quantum theory of fields. Vol. 2: Modern applications*. Cambridge University Press, 8, 2013, 10.1017/CBO9781139644174.

[13] G. B. Folland, *Quantum field theory: A tourist guide for mathematicians*, vol. 149 of *Mathematical Surveys and Monographs*. American Mathematical Society, 2008.

[14] M. D. Schwartz, *Quantum Field Theory and the Standard Model*. Cambridge University Press, 3, 2014.

[15] S. Coleman, *Lectures of Sidney Coleman on Quantum Field Theory*. WSP, Hackensack, 12, 2018, 10.1142/9371.

[16] F. Halzen and A. D. Martin, *Quarks and Leptons: An Introductory Course in Modern Particle Physics*. John Wiley and Sons Inc., 1984.

[17] D. H. Perkins, *Introduction to high energy physics*. Cambridge University Press, 4th ed., 2000.

[18] H. Georgi, *Weak Interactions and Modern Particle Theory*. Dover, 2009.

[19] J. F. Donoghue, E. Golowich and B. R. Holstein, *Dynamics of the Standard Model: Second edition*. Cambridge University Press, 11, 2022, 10.1017/9781009291033.

[20] Particle Data Group collaboration, *Review of Particle Physics*, *Int. J. Mod. Phys. A* **41** (2026) 2630011.

[21] S. Weinberg, *Foundations of Modern Physics*. Cambridge University Press, 4, 2021, 10.1017/9781108894845.

[22] S. Coleman, *Sidney Coleman's Lectures on Relativity*. Cambridge University Press, 2022.

[23] "Lhc data analysis." https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.lhc_data_analysis.

[24] "Facts and figures about the lhc." https://home.cern/resources/faqs/facts-and-figures-about-lhc.

[25] W.-K. Tung, *Group Theory in Physics*. World Scientific, 8, 1985, 10.1142/0097.

[26] E. Schrödinger, *Quantisierung als Eigenwertproblem*, *Annalen Phys.* **386** (1926) 109.

[27] O. Klein, *Quantentheorie und fünfdimensionale Relativitätstheorie*, *Z. Phys.* **37** (1926) 895.

[28] W. Gordon, *Der Comptoneffekt nach der Schrödingerschen Theorie*, *Z. Phys.* **40** (1926) 117.

[29] P. A. M. Dirac, *The quantum theory of the electron*, *Proc. Roy. Soc. Lond. A* **117** (1928) 610.

[30] H. Weyl, *Electron and Gravitation. 1.*, *Z. Phys.* **56** (1929) 330.

[31] W. Pauli, *Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren*, *Z. Phys.* **31** (1925) 765.

[32] F. Gelis, *Quantum Field Theory*. Cambridge University Press, 7, 2019.

[33] B. Lucini, M. Teper and U. Wenger, *Glueballs and k-strings in SU(N) gauge theories: Calculations with improved operators*, *JHEP* **06** (2004) 012 [hep-lat/0404008].

[34] M. Teper, *Large N and confining flux tubes as strings - a view from the lattice*, *Acta Phys. Polon. B* **40** (2009) 3249 [0912.3339].

[35] P. Bicudo, N. Cardoso and M. Cardoso, *Pure gauge QCD flux tubes and their widths at finite temperature*, *Nucl. Phys. B* **940** (2019) 88 [1702.03454].

[36] A. Athenodorou and M. Teper, *The glueball spectrum of SU(3) gauge theory in 3 + 1 dimensions*, *JHEP* **11** (2020) 172 [2007.06422].

[37] J. Bulava, B. Hörz, F. Knechtli, V. Koch, G. Moir, C. Morningstar et al., *String breaking by light and strange quarks in QCD*, *Phys. Lett. B* **793** (2019) 493 [1902.04006].

[38] P. W. Anderson, *Plasmons, Gauge Invariance, and Mass*, *Phys. Rev.* **130** (1963) 439.

[39] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, *Phys. Rev. Lett.* **13** (1964) 321.

[40] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, *Phys. Rev. Lett.* **13** (1964) 508.

[41] G. S. Guralnik, C. R. Hagen and T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, *Phys. Rev. Lett.* **13** (1964) 585.

[42] Y. Nambu, *Quasiparticles and Gauge Invariance in the Theory of Superconductivity*, *Phys. Rev.* **117** (1960) 648.

[43] J. Goldstone, *Field Theories with Superconductor Solutions*, *Nuovo Cim.* **19** (1961) 154.

[44] R. de Bruyn Ouboter and A. N. Omelyanchouk, *On massive photons inside a superconductor as follows from london and ginzburg-landau theory*, *Low Temperature Physics* **43** (2017) 889.

[45] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, *Phys. Rev. Lett.* **10** (1963) 531.

[46] M. Kobayashi and T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, *Prog. Theor. Phys.* **49** (1973) 652.

[47] A. D. Sakharov, *Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe*, *Pisma Zh. Eksp. Teor. Fiz.* **5** (1967) 32.

[48] M. B. Gavela, P. Hernandez, J. Orloff, O. Pene and C. Quimbay, *Standard model CP violation and baryon asymmetry. Part 2: Finite temperature*, *Nucl. Phys. B* **430** (1994) 382 [hep-ph/9406289].

[49] K. Kajantie, M. Laine, K. Rummukainen and M. E. Shaposhnikov, *The Electroweak phase transition: A Nonperturbative analysis*, *Nucl. Phys. B* **466** (1996) 189 [hep-lat/9510020].

[50] S. M. Carroll, *Spacetime and Geometry: An Introduction to General Relativity*. Cambridge University Press, 2019.

[51] R. d'Inverno and J. Vickers, *Introducing Einstein's Relativity: A Deeper Understanding*. Oxford University Press, 2nd ed., 2022.

[52] S. Weinberg, *Infrared photons and gravitons*, *Phys. Rev.* **140** (1965) B516.

[53] KATRIN collaboration, *Direct neutrino-mass measurement based on 259 days of KATRIN data*, *Science* **388** (2025) adq9592 [2406.13516].

[54] DES collaboration, *Dark Energy Survey Year 3 results: Cosmological constraints from galaxy clustering and weak lensing*, *Phys. Rev. D* **105** (2022) 023520 [2105.13549].

[55] B. Pontecorvo, *Neutrino Experiments and the Problem of Conservation of Leptonic Charge*, *Sov. Phys. JETP* **26** (1968) 984.

[56] Z. Maki, M. Nakagawa and S. Sakata, *Remarks on the unified model of elementary particles*, *Prog. Theor. Phys.* **28** (1962) 870.

[57] I. Esteban, M. C. Gonzalez-Garcia, M. Maltoni, I. Martinez-Soler, J. P. Pinheiro and T. Schwetz, *NuFit-6.0: updated global analysis of three-flavor neutrino oscillations*, *JHEP* **12** (2024) 216 [2410.05380].

[58] E. Majorana, *Teoria simmetrica dell'elettrone e del positrone*, *Nuovo Cim.* **14** (1937) 171.

[59] T. Yanagida, *Horizontal gauge symmetry and masses of neutrinos*, *Conf. Proc. C* **7902131** (1979) 95.

[60] M. Gell-Mann, P. Ramond and R. Slansky, *Complex Spinors and Unified Theories*, *Conf. Proc. C* **790927** (1979) 315 [1306.4669].

[61] S. L. Glashow, *The Future of Elementary Particle Physics*, *NATO Sci. Ser. B* **61** (1980) 687.

[62] R. N. Mohapatra and G. Senjanovic, *Neutrino Mass and Spontaneous Parity Nonconservation*, *Phys. Rev. Lett.* **44** (1980) 912.

[63] A. Hook, *TASI Lectures on the Strong CP Problem and Axions*, *PoS* **TASI2018** (2019) 004 [1812.02669].

[64] F. Wilczek, *Problem of Strong P and T Invariance in the Presence of Instantons*, *Phys. Rev. Lett.* **40** (1978) 279.

[65] S. Weinberg, *A New Light Boson?*, *Phys. Rev. Lett.* **40** (1978) 223.

[66] R. D. Peccei and H. R. Quinn, *CP Conservation in the Presence of Instantons*, *Phys. Rev. Lett.* **38** (1977) 1440.

[67] J. C. Pati and A. Salam, *Lepton Number as the Fourth Color*, *Phys. Rev. D* **10** (1974) 275.

[68] H. Georgi and S. L. Glashow, *Unity of All Elementary Particle Forces*, *Phys. Rev. Lett.* **32** (1974) 438.

[69] R. N. Mohapatra, *Unification and Supersymmetry: The Frontiers of Quark-Lepton Physics*. Springer, New York, 3rd ed., 2002, 10.1007/b98865.

[70] S. Raby, *Introduction to the Standard Model and Beyond*. Cambridge University Press, 8, 2021, 10.1017/9781108644129.

[71] S. P. Martin, *A Supersymmetry primer*, *Adv. Ser. Direct. High Energy Phys.* **18** (1998) 1 [hep-ph/9709356].

[72] Super-Kamiokande collaboration, *Search for proton decay via $p \to e^+\pi^0$ and $p \to \mu^+\pi^0$ with an enlarged fiducial volume in Super-Kamiokande I-IV*, *Phys. Rev. D* **102** (2020) 112011 [2010.16098].

[73] E. Gildener, *Gauge Symmetry Hierarchies*, *Phys. Rev. D* **14** (1976) 1667.

[74] S. Weinberg, *Gauge Hierarchies*, *Phys. Lett. B* **82** (1979) 387.

[75] L. Susskind, *Dynamics of Spontaneous Symmetry Breaking in the Weinberg-Salam Theory*, *Phys. Rev. D* **20** (1979) 2619.

[76] S. R. Coleman and E. J. Weinberg, *Radiative Corrections as the Origin of Spontaneous Symmetry Breaking*, *Phys. Rev. D* **7** (1973) 1888.

[77] B. Bellazzini, C. Csáki and J. Serra, *Composite Higgses*, *Eur. Phys. J. C* **74** (2014) 2766 [1401.2457].

[78] N. Arkani-Hamed, S. Dimopoulos and G. R. Dvali, *The Hierarchy problem and new dimensions at a millimeter*, *Phys. Lett. B* **429** (1998) 263 [hep-ph/9803315].

[79] I. Antoniadis, N. Arkani-Hamed, S. Dimopoulos and G. R. Dvali, *New dimensions at a millimeter to a Fermi and superstrings at a TeV*, *Phys. Lett. B* **436** (1998) 257 [hep-ph/9804398].

[80] L. Randall and R. Sundrum, *A Large mass hierarchy from a small extra dimension*, *Phys. Rev. Lett.* **83** (1999) 3370 [hep-ph/9905221].

[81] L. Randall and R. Sundrum, *An Alternative to compactification*, *Phys. Rev. Lett.* **83** (1999) 4690 [hep-th/9906064].

[82] N. Craig, *Naturalness: past, present, and future*, *Eur. Phys. J. C* **83** (2023) 825 [2205.05708].

[83] J. S. Townsend, *A Modern Approach to Quantum Mechanics*. University Science Books, 2nd ed., 2012.

[84] D. J. Griffiths and D. F. Schroeter, *Introduction to Quantum Mechanics*. Cambridge University Press, 3rd ed., 2018.

[85] A. Berera and L. Del Debbio, *Quantum Mechanics*. Cambridge University Press, 2021.

[86] D. H. McIntyre, *Quantum Mechanics: A Paradigms Approach*. Cambridge University Press, 2022.

[87] D. Tong, *Quantum Mechanics: Lectures on Theoretical Physics*, Lectures on Theoretical Physics. Cambridge University Press, 2025.

[88] E. Merzbacher, *Lectures on Quantum Mechanics*. John Wiley and Sons Inc., 3rd ed., 1998.

[89] S. Weinberg, *Lectures on Quantum Mechanics*. Cambridge University Press, 2nd ed., 2015.

[90] S. P. Martin, *Quantum Mechanics: A Comprehensive Graduate Guide*, Graduate Texts in Physics. Springer Cham, 2026.

[91] B. C. Hall, *Lie Groups, Lie Algebras, and Representations*, vol. 222 of *Graduate Texts in Mathematics*. Springer, Cham, 2015, 10.1007/978-3-319-13467-3.

[92] H. Georgi, *Lie Algebras In Particle Physics : from Isospin To Unified Theories*. Taylor & Francis, Boca Raton, 2000, 10.1201/9780429499210.

[93] J. R. Taylor, *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*. MIT Press, 3rd ed., 2022.