

A Simple Introduction to Particle Physics

Part I - Foundations and the Standard Model

Matthew B. Robinson,¹ Karen R. Bland,²
Gerald B. Cleaver,³ and Jay R. Dittmann⁴

*Department of Physics, One Bear Place # 97316
Baylor University
Waco, TX 76798-7316*

Abstract

This is the first of a series of papers in which we present a brief introduction to the relevant mathematical and physical ideas that form the foundation of Particle Physics, including Group Theory, Relativistic Quantum Mechanics, Quantum Field Theory and Interactions, Abelian and Non-Abelian Gauge Theory, and the $SU(3) \otimes SU(2) \otimes U(1)$ Gauge Theory that describes our universe apart from gravity. Our approach, at first, is an algebraic exposition of Gauge Theory and how the physics of our universe comes out of Gauge Theory.

With an algebraic understanding of Gauge Theory and the relevant physics of the Standard Model from this paper, in a subsequent paper we will “back up” and reformulate Gauge Theory from a geometric foundation, showing how it connects to the algebraic picture initially built in these notes.

Finally, we will introduce the basic ideas of String Theory, showing both the geometric and algebraic correspondence with Gauge Theory as outlined in the first two parts.

These notes are not intended to be a comprehensive introduction to any of the ideas contained in them. Their purpose is to introduce the “forest” rather than the “trees”. The primary emphasis is on the algebraic/geometric/mathematical underpinnings rather than the calculational/phenomenological details. Among the glaring omissions are CPT theorems, evaluations of Feynman Diagrams, Renormalization, and Anomalies. The topics were chosen according to the authors’ preferences and agenda.

These notes are intended for a student who has completed the standard undergraduate physics and mathematics courses. The material in the first part is intended as a review and is therefore cursory. Furthermore, these notes should not and will not in any way take the place of the related courses, but rather provide a primer for detailed courses in QFT, Gauge Theory, String Theory, etc., which will fill in the many gaps left by this paper.

¹m_robinson@baylor.edu

²karen.bland@baylor.edu

³gerald.cleaver@baylor.edu

⁴jay_dittmann@baylor.edu

Contents

1	Part I — Preliminary Concepts	5
1.1	Review of Classical Physics	5
1.1.1	Hamilton's Principle	5
1.1.2	Noether's Theorem	6
1.1.3	Conservation of Energy	7
1.1.4	Lorentz Transformations	7
1.1.5	A More Detailed Look at Lorentz Transformations	9
1.1.6	Classical Fields	10
1.1.7	Classical Electrodynamics	11
1.1.8	Classical Electrodynamics Lagrangian	12
1.1.9	Gauge Transformations	13
1.2	References and Further Reading	14
2	Part II — Algebraic Foundations	15
2.1	Introduction to Group Theory	15
2.1.1	What is a Group?	15
2.1.2	Finite Discrete Groups and Their Organization	17
2.1.3	Group Actions	18
2.1.4	Representations	20
2.1.5	Reducibility and Irreducibility — A Preview	23
2.1.6	Algebraic Definitions	24
2.1.7	Reducibility Revisited	29
2.2	Introduction to Lie Groups	33
2.2.1	Classification of Lie Groups	34

2.2.2	Generators	38
2.2.3	Lie Algebras	40
2.2.4	The Adjoint Representation	42
2.2.5	$SO(2)$	43
2.2.6	$SO(3)$	43
2.2.7	$SU(2)$	45
2.2.8	$SU(2)$ and Physical States	45
2.2.9	$SU(2)$ for $j = \frac{1}{2}$	48
2.2.10	$SU(2)$ for $j = 1$	50
2.2.11	$SU(2)$ for Arbitrary j	50
2.2.12	Root Space	51
2.2.13	Adjoint Representation of $SU(2)$	57
2.2.14	$SU(2)$ for Arbitrary j ... Again	59
2.2.15	$SU(3)$	61
2.2.16	What is the Point of All of This?	64
2.3	References and Further Reading	65
3	Part III — Quantum Field Theory	66
3.1	A Primer to Quantization	66
3.1.1	Quantum Fields	66
3.1.2	Spin-0 Fields	67
3.1.3	Why $SU(2)$ for Spin?	68
3.1.4	Spin $\frac{1}{2}$ Particles	69
3.1.5	The Lorentz Group	71
3.1.6	The Dirac Sea Interpretation of Antiparticles	73
3.1.7	The QFT Interpretation of Antiparticles	74

3.1.8	Lagrangians for Scalars and Dirac Particles	75
3.1.9	Conserved Currents	76
3.1.10	The Dirac Equation with an Electromagnetic Field	76
3.1.11	Gauging the Symmetry	77
3.2	Quantization	81
3.2.1	Review of What Quantization Means	81
3.2.2	Canonical Quantization of Scalar Fields	82
3.2.3	The Spin-Statistics Theorem	86
3.2.4	Left-Handed and Right-Handed Fields	87
3.2.5	Canonical Quantization of Fermions	89
3.2.6	Insufficiencies of Canonical Quantization	90
3.2.7	Path Integrals and Path Integral Quantization	91
3.2.8	Interpretation of the Path Integral	93
3.2.9	Expectation Values	94
3.2.10	Path Integrals with Fields	95
3.2.11	Interacting Scalar Fields and Feynman Diagrams	98
3.2.12	Interacting Fermion Fields	102
3.3	Final Ingredients	103
3.3.1	Spontaneous Symmetry Breaking	104
3.3.2	Breaking Local Symmetries	106
3.3.3	Non-Abelian Gauge Theory	107
3.3.4	Representations of Gauge Groups	109
3.3.5	Symmetry Breaking Revisited	110
3.3.6	Simple Examples of Symmetry Breaking	112
3.3.7	A More Complicated Example of Symmetry Breaking	114

3.4	Particle Physics	115
3.4.1	Introduction to the Standard Model	115
3.4.2	The Gauge and Higgs Sector	116
3.4.3	The Lepton Sector	120
3.4.4	The Quark Sector	124
3.5	References and Further Reading	126
4	The Standard Model — A Summary	127
4.1	How Does All of This Relate to Real Life?	127
4.2	The Fundamental Forces	128
4.3	Categorizing Particles	129
4.4	Elementary Particles	130
4.4.1	Elementary Fermions	130
4.4.2	Elementary Bosons	132
4.5	Composite Particles	133
4.6	Visualizing It All	134
5	A Look Ahead	135

1 Part I — Preliminary Concepts

1.1 Review of Classical Physics

1.1.1 Hamilton's Principle

Nearly all physics begins with what is called a **Lagrangian** for a particle, which is initially defined as the kinetic energy minus the potential energy,

$$L \equiv T - V$$

where $T = T(q, \dot{q})$ and $V = V(q)$. Then, the **Action** is defined as the integral of the Lagrangian from an initial time to a final time,

$$S \equiv \int_{t_i}^{t_f} dt L(q, \dot{q})$$

It is important to realize that S is a “functional” of the particle’s world-line in (q, \dot{q}) space, not a function. This means that it depends on the entire path (q, \dot{q}) , rather than a given point on the path. The only fixed points on the path are $q(t_i)$, $q(t_f)$, $\dot{q}(t_i)$, and $\dot{q}(t_f)$. The rest of the path is generally unconstrained, and the value of S depends on the entire path.

Hamilton's Principle says that nature extremizes the path a particle will take in going from $q(t_i)$ at time t_i to position $q(t_f)$ at time t_f . In other words, the path that extremizes the action will be the path the particle will travel.

But, because S is a functional, depending on the entire path in (q, \dot{q}) space rather than a point, it cannot be extremized in the “Calculus I” sense of merely setting the derivative equal to 0. Instead, we must find the path for which the action is “stationary”. This means that the first-order term in the Taylor Expansion around that path will vanish, or $\delta S = 0$ at that path.

To find this, consider some arbitrary path (q, \dot{q}) . If it is a path that minimizes the action, then we will have

$$\begin{aligned} 0 &= \delta S = \delta \int_{t_i}^{t_f} dt L(q, \dot{q}) = \int_{t_i}^{t_f} dt L(q + \delta q, \dot{q} + \delta \dot{q}) - S \\ &= \int_{t_i}^{t_f} dt L(q, \dot{q}) + \int_{t_i}^{t_f} dt \left(\delta q \frac{\partial L}{\partial q} + \delta \dot{q} \frac{\partial L}{\partial \dot{q}} \right) - S \\ &= \int_{t_i}^{t_f} dt \left(\delta q \frac{\partial L}{\partial q} + \frac{\partial L}{\partial \dot{q}} \frac{d}{dt} \delta q \right) \end{aligned}$$

Integrating the second term by parts, and taking the variation of δq to be at 0 at t_i and t_f ,

$$\delta S = \int_{t_i}^{t_f} dt \left(\delta q \frac{\partial L}{\partial q} - \delta q \frac{d}{dt} \frac{\partial L}{\partial \dot{q}} \right) = \int_{t_i}^{t_f} dt \delta q \left(\frac{\partial L}{\partial q} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}} \right) = 0$$

The only way to guarantee this for an arbitrary variation δq from the path (q, \dot{q}) is to demand

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}} - \frac{\partial L}{\partial q} = 0$$

This equation is called the **Euler-Lagrange** equation, and it produces the equations of motion of the particle.

The generalization to multiple coordinates q_i ($i = 1, \dots, n$) is straightforward:

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} = 0 \quad (1.1)$$

1.1.2 Noether's Theorem

Given a Lagrangian $L = L(q, \dot{q})$, consider making an infinitesimal transformation

$$q \rightarrow q + \epsilon \delta q$$

where ϵ is some infinitesimal constant. This transformation will give

$$L(q, \dot{q}) \rightarrow L(q + \epsilon \delta q, \dot{q} + \epsilon \delta \dot{q}) = L(q, \dot{q}) + \epsilon \delta q \frac{\partial L}{\partial q} + \epsilon \delta \dot{q} \frac{\partial L}{\partial \dot{q}}$$

If the Euler-Lagrange equations of motion are satisfied, so that $\frac{\partial L}{\partial q} = \frac{d}{dt} \frac{\partial L}{\partial \dot{q}}$, then under $q \rightarrow q + \epsilon \delta q$,

$$L \rightarrow L + \epsilon \delta q \frac{\partial L}{\partial q} + \epsilon \delta \dot{q} \frac{\partial L}{\partial \dot{q}} = L + \epsilon \delta q \frac{d}{dt} \frac{\partial L}{\partial \dot{q}} + \epsilon \frac{\partial L}{\partial \dot{q}} \frac{d}{dt} \delta q = L + \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}} \epsilon \delta q \right)$$

So, under $q \rightarrow q + \epsilon \delta q$, we have $\delta L = \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}} \epsilon \delta q \right)$. We define the **Noether Current**, j , as

$$j \equiv \frac{\partial L}{\partial \dot{q}} \delta q$$

Now, if we can find some transformation δq that leaves the action invariant, or in other words such that $\delta S = 0$, then $\frac{dj}{dt} = 0$, and therefore the current j is a constant in time. In other words, j is *conserved*.

As a familiar example, consider a projectile, described by the Lagrangian

$$L = \frac{1}{2} m (\dot{x}^2 + \dot{y}^2) - mgy \quad (1.2)$$

This will be unchanged under the transformation $x \rightarrow x + \epsilon$, where ϵ is any constant (here, $\delta q = 1$ in the above notation), because $x \rightarrow x + \epsilon \Rightarrow \dot{x} \rightarrow \dot{x}$. So, $j = \frac{\partial L}{\partial \dot{q}} \delta q = m\dot{x}$ is conserved. We recognize $m\dot{x}$ as the momentum in the x -direction, which we expect to be conserved by conservation of momentum.

So in summary, **Noether's Theorem** merely says that whenever there is a continuous symmetry in the action, there is a corresponding conserved quantity.

1.1.3 Conservation of Energy

Consider the quantity

$$\frac{dL}{dt} = \frac{d}{dt}L(q, \dot{q}) = \frac{\partial L}{\partial q} \frac{dq}{dt} + \frac{\partial L}{\partial \dot{q}} \frac{d\dot{q}}{dt} + \frac{\partial L}{\partial t}$$

Because L does not depend explicitly on time, $\frac{\partial L}{\partial t} = 0$, and therefore

$$\frac{dL}{dt} = \frac{\partial L}{\partial q} \dot{q} + \frac{\partial L}{\partial \dot{q}} \ddot{q} = \left(\frac{d}{dt} \frac{\partial L}{\partial \dot{q}} \right) \dot{q} + \frac{\partial L}{\partial \dot{q}} \ddot{q} = \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}} \dot{q} \right)$$

where we have used the Euler-Lagrange equation to get the second equality. So, we have $\frac{dL}{dt} = \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}} \dot{q} \right)$, or

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}} \dot{q} - L \right) = 0 \quad (1.3)$$

For a general non-relativistic system, $L = T - V$, so $\frac{\partial L}{\partial \dot{q}} = \frac{\partial T}{\partial \dot{q}}$ because V is a function of q only, and normally

$$T \propto \dot{q}^2 \quad \Rightarrow \quad \frac{\partial L}{\partial \dot{q}} \dot{q} = 2T$$

So, $\frac{\partial L}{\partial \dot{q}} \dot{q} - L = 2T - (T - V) = T + V = E$, the total energy of the system, which is conserved according to (1.3). We identify $T + V \equiv H$ as the **Hamiltonian**, or total energy function, of the system.

Furthermore, we define $\frac{\partial L}{\partial \dot{q}} \equiv p$ to be the momentum of the system. Then, the relationship between the Lagrangian and the Hamiltonian is the Legendre transformation

$$p\dot{q} - L = H$$

1.1.4 Lorentz Transformations

Consider some event that occurs at spatial position $(x, y, z)^T$, at time t . (The superscript T denotes the transpose, so this is a column vector.) We arrange this event in a column 4-vector as $(ct, x, y, z)^T$, where c is the speed of light (the units of c give each element the same units). A more useful notation is to refer to this vector as $a^\mu = (ct, x, y, z)^T$, where $\mu = 0, 1, 2, 3$. This 4-vector, with the μ index raised, is called a “vector”, or a “contravariant vector”. Then, we define the row vector $a_\mu = (-ct, x, y, z)$. This is called a “covector”, or a “covariant vector”. In general, the sign of the 0th component (the component in the first position) changes when going from vector to covector.

There is something very deep going on here regarding the geometrical picture between vectors and covectors, but we will not discuss it until the next paper in this series.

The dot product between two such vectors (a covector and vector) is then defined as the product with one index raised and the other lowered. Whenever indices are contracted in such a way, it is understood that they are to be summed over.¹

$$a \cdot b = a^\mu b_\mu = a^0 b_0 + a^1 b_1 + a^2 b_2 + a^3 b_3 = -a^0 b^0 + a^1 b^1 + a^2 b^2 + a^3 b^3$$

Or, plugging in the spacetime notation from above, where

$$a^\mu = (ct_1, x_1, y_1, z_1)^T \quad \text{and} \quad b^\mu = (ct_2, x_2, y_2, z_2)^T$$

we have

$$a \cdot b = a_\mu b^\mu = -c^2 t_1 t_2 + x_1 x_2 + y_1 y_2 + z_1 z_2$$

We can also discuss the differential version of this. If $s^\mu = (ct, x, y, z)$, then $ds^2 = -c^2 dt^2 + dx^2 + dy^2 + dz^2$.

In his theory of Special Relativity, Einstein postulated that all inertial reference frames are equivalent, and that the speed of light is the same in all frames. To put this in more mathematical terms, if observers in different inertial frames 1 and 2 each see an event, they will see, respectively,

$$\begin{aligned} ds_1^2 &= -c^2 dt_1^2 + dx_1^2 + dy_1^2 + dz_1^2 \\ ds_2^2 &= -c^2 dt_2^2 + dx_2^2 + dy_2^2 + dz_2^2 \end{aligned}$$

We then demand that $ds_1^2 = ds_2^2$. To do this, we must find a modification of the standard Galilean transformations that will leave ds^2 unchanged. The derivation for the correct transformations can be found in any introductory or modern physics text, so we merely quote the result. If we assume that frame 2 is moving only in the z -direction with respect to frame 1 (and that their x , y , and z axes are aligned), then we find that the transformations are

$$\begin{aligned} t_2 &= \gamma(ct_1 - \beta z_1) \\ z_2 &= \gamma(z_1 - \beta ct_1) \end{aligned} \tag{1.4}$$

where $\beta = \frac{v}{c}$ and $\gamma = \frac{1}{\sqrt{1-\beta^2}}$. These transformations, which preserve ds^2 when transforming one frame to another, are called **Lorentz Transformations**.

Discussions of the implications of these transformations, including time dilation, length contraction, and the relationship between energy and mass can be found in most introductory texts. You are encouraged to review the material if you are not familiar with it.

¹because we are summing over components, we can write $a^\mu b_\mu$ or $a_\mu b^\mu$ — they mean the same thing

1.1.5 A More Detailed Look at Lorentz Transformations

As we have seen, we have a quantity $ds^2 = -c^2 dt^2 + dx^2 + dy^2 + dz^2$, which does not change under transformations (1.4). Thinking of physical ideas this way, in terms of “what doesn’t change when something else changes”, will prove to be an extraordinarily powerful approach. In order to understand Special Relativity in such a way, we begin with a simpler example.

Consider a spatial rotation around, say, the z -axis (or, equivalently, mixing the x and y coordinates). Such a transformation is called an **Euler Transformation**, and takes the form

$$\begin{aligned} t' &= t \\ x' &= x \cos \theta + y \sin \theta \\ y' &= -x \sin \theta + y \cos \theta \\ z' &= z \end{aligned} \tag{1.5}$$

where θ is the angle of rotation, called the **Euler Angle**. We can simultaneously express a Lorentz transformation as a sort of “rotation” that mixes a spatial dimension and a time dimension, as follows (these transformations are equivalent to (1.4):

$$\begin{aligned} t' &= t \cosh \theta - x \sinh \theta \\ x' &= -t \sinh \theta + x \cosh \theta \\ y' &= y \\ z' &= z \end{aligned} \tag{1.6}$$

where θ is defined by the relationship $\beta = \tanh \theta$.

We denote a transformation mixing two spatial dimensions simply a **Rotation**, whereas a transformation mixing a spatial dimension and a time dimension is a **Boost**. Any two frames whose origins coincide at $t = t' = 0$ can be transformed into each other through some combination of rotations and boosts.

To rephrase this in more precise language, given a 4-vector x^μ , it will be related to the equivalent 4-vector in another frame, x'^μ , by some matrix L , according to $x'^\mu = L^\mu_\nu x^\nu$ (where the summation convention discussed earlier is in effect for the repeated index).

We also introduce what is called the **Metric** matrix,

$$\eta_{\mu\nu} = \eta^{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

In general, $\eta^{\mu\nu} \equiv (\eta_{\mu\nu})^{-1}$.

Using the metric, the dot product of any 4-vector $x^\mu = (ct, x, y, z)^T$ can be easily written as $x^2 = x^\mu x_\mu = \eta_{\mu\nu} x^\mu x^\nu = -c^2 t^2 + x^2 + y^2 + z^2$. In general, a Lorentz transformation can be defined as a matrix L^μ_ν (including boosts and rotations) that leaves $\eta_{\mu\nu} x^\mu x^\nu$ unchanged.

For example, a scalar, or an object with no uncontracted indices, like ϕ or $x^\mu x_\mu$, is simply invariant under Lorentz transformations ($\phi \rightarrow \phi$, $x^\mu x_\mu \rightarrow x^\mu x_\mu$).

A vector, or an object with only one uncontracted index, like x^μ or $a^\mu b^\nu_\mu$, transforms according to $x'^\mu = L^\mu_\nu x^\nu$, or $(a^\mu b^\nu_\mu)' = L^\nu_\alpha (a^\mu b^\alpha_\mu)$.

Now, consider the dot product $x^2 = x^\mu x_\mu = \eta_{\mu\nu} x^\mu x^\nu$. If x^2 is invariant, then $x'^2 = x^2 \Rightarrow \eta_{\mu\nu} x'^\mu x'^\nu = \eta_{\mu\nu} L^\mu_\alpha L^\nu_\beta x^\alpha x^\beta$ demands that $\eta_{\mu\nu} L^\mu_\alpha L^\nu_\beta = \eta_{\alpha\beta}$. So, the constraint for Lorentz transformations is that they are the set of all matrices such that

$$\eta_{\mu\nu} L^\mu_\alpha L^\nu_\beta = \eta_{\alpha\beta}$$

We take this to be the *defining* constraint for a Lorentz transformation.

1.1.6 Classical Fields

When deriving the Euler-Lagrange equations, we started with an action S which was an integral over time only ($S \equiv \int dt L$). If we are eventually interested in a relativistically acceptable theory, this is obviously no good because it treats time and space differently (the action is an integral over time but not over space).

So, let's consider an action defined not in terms of the Lagrangian, but of the "Lagrangian per unit volume", or the **Lagrangian Density** \mathcal{L} . The Lagrangian will naturally be the integral of \mathcal{L} over all space, $L = \int d^n x \mathcal{L}$. The integral is in n -dimensions, so $d^n x$ means $dx^1 dx^2 dx^3 \dots dx^n$.

Now, the action will be $S = \int dt L = \int dt d^n x \mathcal{L}$. In the normal 1+3 dimensional Minkowski spacetime we live in, this will be $S = \int dt d^3 x \mathcal{L} = \int d^4 x \mathcal{L}$.

Before, L depended not on t , but on the path $q(t)$, $\dot{q}(t)$. In a similar sense, \mathcal{L} will not depend on \bar{x} and t , but on what we will refer to as **Fields**, $\phi(\bar{x}, t) = \phi(x^\mu)$, which exist in spacetime.

Following a nearly identical argument as the one leading to (1.1), we get the relativistic field generalization

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi_i)} \right) - \frac{\partial \mathcal{L}}{\partial \phi_i} = 0$$

for multiple fields ϕ_i ($i = 1, \dots, n$).

Noether's Theorem says that, for $\phi \rightarrow \phi + \epsilon \delta \phi$, we have a current $j^\mu \equiv \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \delta \phi$, and if $\phi \rightarrow \phi + \epsilon \delta \phi$ leaves $\delta \mathcal{L} = 0$, then $\partial_\mu j^\mu = 0 \Rightarrow -\frac{\partial j^0}{\partial t} + \bar{\nabla} \cdot \bar{j} = 0$, where j^0 is the **Charge Density**, and \bar{j} is the **Current Density**. The total charge will naturally be $Q \equiv \int_{all\ space} d^3x j^0$.

Finally, we also have a **Hamiltonian Density** and momentum

$$\mathcal{H} \equiv \frac{\partial \mathcal{L}}{\partial \dot{\phi}_\mu} \dot{\phi}_\mu - \mathcal{L} \quad (1.7)$$

$$\Pi^\mu \equiv \frac{\partial \mathcal{L}}{\partial \dot{\phi}_\mu} \quad (1.8)$$

One final comment for this section. For the remainder of these notes, we will ultimately be seeking a relativistic field theory, and therefore we will never make use of Lagrangians. We will always use Lagrangian densities. We will always use the notation \mathcal{L} instead of L , but we will refer to the Lagrangian densities simply as Lagrangians. We drop the word "densities" for brevity, and because there will never be ambiguity.

1.1.7 Classical Electrodynamics

We choose our units so that $c = \mu_0 = \epsilon_0 = 1$. So, the magnitude of the force between two charges q_1 and q_2 is $F = \frac{q_1 q_2}{4\pi r^2}$. In these units, Maxwell's equations are

$$\bar{\nabla} \cdot \bar{E} = \rho \quad (1.9)$$

$$\bar{\nabla} \times \bar{B} - \frac{\partial \bar{E}}{\partial t} = \bar{J} \quad (1.10)$$

$$\bar{\nabla} \cdot \bar{B} = 0 \quad (1.11)$$

$$\bar{\nabla} \times \bar{E} + \frac{\partial \bar{B}}{\partial t} = 0 \quad (1.12)$$

If we define the **Potential** 4-vector $A^\mu = (\phi, \bar{A})$, then we can define $\bar{B} = \bar{\nabla} \times \bar{A}$ and $\bar{E} = -\bar{\nabla} \phi - \frac{\partial \bar{A}}{\partial t}$. Writing \bar{B} and \bar{E} this way will automatically solve the homogenous Maxwell equations, (1.11) and (1.12).

Then, we define the totally antisymmetric **Electromagnetic Field Strength Tensor** $F^{\mu\nu}$ as

$$F^{\mu\nu} \equiv \partial^\mu A^\nu - \partial^\nu A^\mu = \begin{pmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & -B_z & B_y \\ E_y & B_z & 0 & -B_x \\ E_z & -B_y & B_x & 0 \end{pmatrix}$$

We define the 4-vector current as $J^\mu = (\rho, \bar{J})$. It is straightforward, though tedious, to

show that

$$\begin{aligned}\partial^\lambda F^{\mu\nu} + \partial^\nu F^{\lambda\mu} + \partial^\mu F^{\nu\lambda} = 0 &\Rightarrow \bar{\nabla} \cdot \bar{B} = 0 \quad \text{and} \quad \bar{\nabla} \times \bar{E} + \frac{\partial \bar{B}}{\partial t} = 0 \\ \partial_\mu F^{\mu\nu} = J^\nu &\Rightarrow \bar{\nabla} \cdot \bar{E} = \rho \quad \text{and} \quad \bar{\nabla} \times \bar{B} - \frac{\partial \bar{E}}{\partial t} = \bar{J}\end{aligned}$$

1.1.8 Classical Electrodynamics Lagrangian

Bringing together the ideas of the previous sections, we now want to construct a Lagrangian density \mathcal{L} which will, via Hamilton's Principle, produce Maxwell's equations.

First, we know that \mathcal{L} must be a scalar (no uncontracted indices). From our intuition with "Physics I" type Lagrangians, we know that kinetic terms are quadratic in the derivatives of the fundamental coordinates (i.e. $\frac{1}{2}m\dot{x}^2 = \frac{1}{2}m(\frac{dx}{dt}) \cdot (\frac{dx}{dt})$). The natural choice is to take A^μ as the fundamental field. It turns out that the correct choice is

$$\mathcal{L}_{EM} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - J^\mu A_\mu \quad (1.13)$$

(note that the F^2 term is quadratic in $\partial^\mu A^\nu$). So,

$$S = \int d^4x \left[-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - J^\mu A_\mu \right] \quad (1.14)$$

Taking the variation of (1.14) with respect to A^μ ,

$$\begin{aligned}\delta S &= \int d^4x \left[-\frac{1}{4}F_{\mu\nu}\delta F^{\mu\nu} - \frac{1}{4}\delta F_{\mu\nu}F^{\mu\nu} - J^\mu \delta A_\mu \right] \\ &= \int d^4x \left[-\frac{1}{2}F_{\mu\nu}\delta F^{\mu\nu} - J^\mu \delta A_\mu \right] \\ &= \int d^4x \left[-\frac{1}{2}F_{\mu\nu}(\partial^\mu \delta A^\nu - \partial^\nu \delta A^\mu) - J^\mu \delta A_\mu \right] \\ &= \int d^4x \left[-F_{\mu\nu}\partial^\mu \delta A^\nu - J^\mu \delta A_\mu \right]\end{aligned}$$

Integrating the first term by parts, and choosing boundary conditions so that δA vanishes at the boundaries,

$$\begin{aligned}&= \int d^4x \left[\partial_\mu F^{\mu\nu} \delta A_\nu - J^\nu \delta A_\nu \right] \\ &= \int d^4x \left[\partial_\mu F^{\mu\nu} - J^\nu \right] \delta A_\nu\end{aligned}$$

So, to have $\delta S = 0$, we must have $\partial_\mu F^{\mu\nu} = J^\nu$, and if this is written out one component at a time, it will give exactly the inhomogenous Maxwell equations (1.9) and (1.10). And

as we already pointed out, the homogenous Maxwell equations become identities when written in terms of A^μ .

As a brief note, the way we have chosen to write equation (1.13), in terms of a “potential” A_μ , and the somewhat mysterious antisymmetric “field strength” $F_{\mu\nu}$, is indicative of an extremely deep and very general mathematical structure that goes well beyond classical electrodynamics. We will see this structure unfold as we proceed through these notes. We just want to mention now that this is not merely a clever way of writing electric and magnetic fields, but a specific example of a general theory.

1.1.9 Gauge Transformations

Gauge Transformations are usually discussed toward the end of an undergraduate course on E&M. Students are typically told that they are extremely important, but the reason why is not obvious. We will briefly introduce them here, and while their significance may still not be transparent, we will return to them several times throughout these notes.

Given some specific potential A^μ , we can find the field strength action as in (1.14). However, A^μ does not uniquely specify the action. We can take any arbitrary function $\chi(x^\mu)$, and the action will be invariant under the transformation

$$A^\mu \rightarrow A'^\mu = A^\mu + \partial^\mu \chi \quad (1.15)$$

or

$$A^\mu \rightarrow A'^\mu = \left(\phi - \frac{\partial \chi}{\partial t}, \bar{A} + \bar{\nabla} \chi \right)$$

Under this transformation, we have

$$\begin{aligned} F'^{\mu\nu} &= \partial^\mu A'^\nu - \partial^\nu A'^\mu = \partial^\mu (A^\nu + \partial^\nu \chi) - \partial^\nu (A^\mu + \partial^\mu \chi) \\ &= \partial^\mu A^\nu - \partial^\nu A^\mu + \partial^\mu \partial^\nu \chi - \partial^\nu \partial^\mu \chi \\ &= F^{\mu\nu} \end{aligned} \quad (1.16)$$

So, $F'^{\mu\nu} = F^{\mu\nu}$.

Furthermore, $J^\mu A_\mu \rightarrow J^\mu A_\mu + J^\mu \partial_\mu \chi$. Integrating the second term by parts with the usual boundary conditions,

$$\int d^4x J^\mu \partial_\mu \chi = - \int d^4x (\partial_\mu J^\mu) \chi$$

But, according to Maxwell’s equations, $\partial_\mu J^\mu = \partial_\mu \partial_\nu F^{\mu\nu} \equiv 0$ because $F^{\mu\nu}$ is totally antisymmetric. So, both $F^{\mu\nu}$ and $J^\mu \partial_\mu \chi$ are invariant under (1.15), and therefore the action of S is invariant under (1.15).

While the importance of gauge transformations may not be obvious at this point, it will become perhaps the most important idea in particle physics. As a note before moving on, recall previously when we mentioned the idea of “what doesn’t change when something else changes” when talking about Lorentz transformations. A gauge transformation is exactly this (in a different context): the fundamental fields are changed by χ , but the equations which govern the physics are unchanged.

In the next section, we provide the mathematical tools to understand why this idea is so important.

1.2 References and Further Reading

The material in this section can be found in nearly any introductory text on Classical Mechanics, Classical Electrodynamics, and Relativity. The primary sources for these notes are [3], [12], and [13].

For further reading, we recommend [6], [18], [19], [22], [33], and [34].

2 Part II — Algebraic Foundations

2.1 Introduction to Group Theory

There are several symbols in this section which may not be familiar. We therefore provide a summary of them for reference.

$\mathbb{N} = \{0, 1, 2, 3, \dots\}$	\Rightarrow is read “implies”	\doteq is “represented by”
$\mathbb{Z} = \{0, \pm 1, \pm 2, \pm 3, \dots\}$	iff is read “if and only if”	\subset is “subset of”
\mathbb{Q} = Rational Numbers	\forall is read “for every”	\equiv is “defined as”
\mathbb{R} = Real Numbers	\exists is read “there exists”	
\mathbb{C} = Complex Numbers	\in is read “in”	
$\mathbb{Z}_n = \mathbb{Z} \mod n$	\ni is read “such that”	

Now that we have reviewed the primary ideas of classical physics, we are almost ready to start talking about particle physics. However, there is a bit of mathematical “machinery” we will need first. Namely, **Group Theory**.

Group theory is, in short, the mathematics of symmetry. We are going to begin talking about what will seem to be extremely abstract ideas, but eventually we will explain how those ideas relate to physics. As a preface of what is to come, the most foundational idea here is, as we said before, “what doesn’t change when something else changes”. A group is a precise and well-defined way of specifying the thing or things that change.

2.1.1 What is a Group?

To begin with, we define the notion of a **Group**. This definition may seem cryptic, but it will be explained in the paragraphs that follow.

A group, denoted (G, \star) , is a set of objects, denoted G , and some operation on those objects, denoted \star , subject to the following:

1. $\forall g_1, g_2 \in G, g_1 \star g_2 \in G$ also. (closure)
2. $\forall g_1, g_2, g_3 \in G$, it must be true that $(g_1 \star g_2) \star g_3 = g_1 \star (g_2 \star g_3)$. (associativity)
3. $\exists g \in G$, denoted e , $\ni \forall g_i \in G, e \star g_i = g_i \star e = g_i$. (identity)
4. $\forall g \in G, \exists h \in G \ni h \star g = g \star h = e$, (so $h = g^{-1}$). (inverse)

Now we explain what this means. By “objects” we literally mean anything. We could be talking about \mathbb{Z} or \mathbb{R} , or we could be talking about a set of Easter eggs all painted different colors.

The meaning of “some operation”, which we are calling \star , can literally be anything you can do to those objects. A formal definition of what \star means could be given, but it will be easier to understand with examples.

Note: The definition of a group doesn’t demand that $g_i \star g_j = g_j \star g_i$. This is a very important point, but we will discuss it in more detail later. We mention it now so it is not new later.

Example 1: $(G, \star) = (\mathbb{Z}, +)$

Consider the set G to be \mathbb{Z} , and the operation to be $\star = +$, or simply addition.

We first check closure. If you take any two elements of \mathbb{Z} and add them together, is the result in \mathbb{Z} ? In other words, if $a, b \in \mathbb{Z}$, is $a + b \in \mathbb{Z}$? Obviously the answer is yes; the sum of two integers is an integer, so closure is met.

Now we check associativity. If $a, b, c \in \mathbb{Z}$, it is trivially true that $a + (b + c) = (a + b) + c$. So, associativity is met.

Now we check identity. Is there an element $e \in \mathbb{Z}$ such that when you add e to any other integer, you get that same integer? Clearly the integer 0 satisfies this. So, identity is met.

Finally, is there an inverse? For any integer $a \in \mathbb{Z}$, will there be another integer $b \in \mathbb{Z}$ such that $a + b = e = 0$? Again, this is obvious, $a^{-1} = -a$ in this case. So, inverse is met.

So, $(G, \star) = (\mathbb{Z}, +)$ is a group.

Example 2: $(G, \star) = (\mathbb{R}, +)$

Obviously, any two real numbers added together is also a real number.

Associativity will hold (of course).

The identity is again 0.

And finally, once again, $-a$ will be the inverse of any $a \in \mathbb{R}$.

Example 3: $(G, \star) = (\mathbb{R}, \cdot)$ (multiplication)

Closer is met; two real numbers multiplied together give a real number.

Associativity obviously holds.

Identity also holds. Any real number $a \in \mathbb{R}$, when multiplied by 1 is a .

Inverse, on the other hand, is trickier. For any real number, is there another real number you can multiply by it to get 1? The instinctive choice is $a^{-1} = \frac{1}{a}$. But, this doesn’t quite work because of $a = 0$. This is the *only* exception, but because there’s an exception, (\mathbb{R}, \cdot) is not a group.

Note: If we take the set $\mathbb{R} - \{0\}$ instead of \mathbb{R} , then $(\mathbb{R} - \{0\}, \cdot)$ is a group.

Example 4: $(G, \star) = (\{1\}, \cdot)$

This is the set with only the element 1, and the operation is normal multiplication. This is indeed a group, but it is extremely uninteresting, and is called the **Trivial Group**.

Example 5: $(G, \star) = (\mathbb{Z}_3, +)$

This is the set of integers mod 3, containing only the elements 0, 1, and 2 (3 mod 3 is 0, 4 mod 3 is 1, 5 mod 3 is 2, etc.)

You can check yourself that this is a group.

2.1.2 Finite Discrete Groups and Their Organization

From the examples above, several things should be apparent about groups. One is that there can be any number of objects in a group. We have a special name for the number of objects in the group's set. *The **Order** of a group is the number of elements in it.*

The order of $(\mathbb{Z}, +)$ is infinite (there are an infinite number of integers), as is the order of $(\mathbb{R}, +)$ and $(\mathbb{R} - \{0\}, \cdot)$. But, the order of $(\{1\}, \cdot)$ is 1, and the order of $(\mathbb{Z}_3, +)$ is 3.

If the order of a group is finite, the group is said to be **Finite**. Otherwise it is **Infinite**.

It is also clear that the elements of groups may be **Discrete**, or they may be **Continuous**. For example, $(\mathbb{Z}, +)$, $(\{1\}, \cdot)$, and $(\mathbb{Z}_3, +)$ are all discrete, while $(\mathbb{R}, +)$ and $(\mathbb{R} - \{0\}, \cdot)$ are both continuous.

Now that we understand what a discrete finite group is, we can talk about how to organize one. Namely, we use what is called a **Multiplication Table**. A multiplication table is a way of organizing the elements of a group as follows:

(G, \star)	e	g_1	g_2	\cdots
e	$e \star e$	$e \star g_1$	$e \star g_2$	\cdots
g_1	$g_1 \star e$	$g_1 \star g_1$	$g_1 \star g_2$	\cdots
g_2	$g_2 \star e$	$g_2 \star g_1$	$g_2 \star g_2$	\cdots
\vdots	\vdots	\vdots	\vdots	\ddots

We state the following property of multiplication tables without proof. *A multiplication table must contain every element of the group exactly one time in every row and every column.* A few minutes thought should convince you that this is necessary to ensure that the definition of a group is satisfied.

As an example, we will draw a multiplication table for the group of order 2. We won't look at specific numbers, but rather call the elements g_1 and g_2 . We begin as follows:

(G, \star)	e	g_1
e	?	?
g_1	?	?

Three of these are easy to fill in from the identity:

(G, \star)	e	g_1
e	e	g_1
g_1	g_1	?

And because we know that every element must appear exactly once, the final question mark must be e . So, there is only one possible group of order 2.

We will consider a few more examples, but we stress at this point that the temptation to plug in numbers should be avoided. Groups are abstract things, and you should try to think of them in terms of the abstract properties, not in terms of actual numbers.

We can proceed with the multiplication table for the group of order 3. You will find that, once again, there is only one option. (Doing this is instructive and it would be helpful to work this out yourself.)

(G, \star)	e	g_1	g_2
e	e	g_1	g_2
g_1	g_1	g_2	e
g_2	g_2	e	g_1

You are encouraged to work out the possibilities for groups of order 4. (*Hint: there are 4 possibilities.*)

2.1.3 Group Actions

So far we have only considered elements of groups and how they relate to each other. The point has been that a particular group represents nothing more than a structure. There are a set of things, and they relate to each other in a particular way. Now, however, we want to consider an extremely simple version of how this relates to nature.

Example 6

Consider three Easter eggs, all painted different colors (red, orange, and yellow), which we denote R, O, and Y. Now, assume they have been put into a row in the order (ROY). If we want to keep them lined up, not take any eggs away, and not add any eggs, what can we do to them? We can do any of the following:

1. Let e be doing nothing to the set, so $e(ROY) = (ROY)$.
2. Let g_1 be a cyclic permutation of the three, $g_1(ROY) = (OYR)$
3. Let g_2 be a cyclic permutation in the other direction, $g_2(ROY) = (YRO)$
4. Let g_3 be swapping the first and second, $g_3(ROY) = (ORY)$
5. Let g_4 be swapping the first and third, $g_4(ROY) = (YOR)$
6. Let g_5 be swapping the second and third, $g_5(ROY) = (RYO)$

You will find that these 6 elements are closed, there is an identity, and each has an inverse.² So, we can draw a multiplication table (you are strongly encouraged to write at least part of this out on your own):

(G, \star)	e	g_1	g_2	g_3	g_4	g_5
e	e	g_1	g_2	g_3	g_4	g_5
g_1	g_1	g_2	e	g_5	g_3	g_4
g_2	g_2	e	g_1	g_4	g_5	g_3
g_3	g_3	g_4	g_5	e	g_1	g_2
g_4	g_4	g_5	g_3	g_2	e	g_1
g_5	g_5	g_3	g_4	g_1	g_2	e

There is something interesting about this group. Notice that $g_3 \star g_1 = g_4$, whereas $g_1 \star g_3 = g_5$. So, we have the surprising result that in this group it is not necessarily true that $g_i \star g_j = g_j \star g_i$.

This leads to a new way of classifying groups. We say a group is **Abelian** if $g_i \star g_j = g_j \star g_i$ $\forall g_i, g_j \in G$. If a group is not Abelian, it is **Non-Abelian**.

Another term commonly used is **Commute**. If $g_i \star g_j = g_j \star g_i$, then we say that g_i and g_j commute. So, an Abelian group is **Commutative**, whereas a Non-Abelian group is **Non-Commutative**.

²We should be very careful to draw a distinction between the *elements* of the group and the *objects* the group acts on. The objects in this example are the eggs, and the permutations are the results of the group action. Neither the eggs nor the permutations of the eggs are the elements of the group. The elements of the group are abstract objects which we are assigning to some operation on the eggs, resulting in a new permutation

The Easter egg group of order 6 above is an example of a very important type of group. It is denoted S_3 , and is called the **Symmetric Group**. It is the group that takes three objects to all permutations of those three objects.

The more general group of this type is S_n , the group that takes n objects to all permutations of those objects. You can convince yourself that S_n will always have order $n!$ (n factorial).

The idea above with the 3 eggs is that S_3 is the *group*, while the eggs are the objects that the group *acts on*. The particular way an element of S_3 changes the eggs around is called the **Group Action** of that element. And each element of S_3 will move the eggs around while leaving them lined up. This ties in to our overarching concept of “what doesn’t change when something else changes”. The fact that there are 3 eggs with 3 particular colors lined up doesn’t change. The order they appear in does.

2.1.4 Representations

We suggested above that you think of groups as purely abstract things rather than trying to plug in actual numbers. Now, however, we want to talk about how to see groups, or the elements of groups, in terms of specific numbers. But, we will do this in a very systematic way. The name for a specific set of numbers or objects that form a group is a **Representation**. The remainder of this section (and the next) will primarily be about group representations.

We already discussed a few simple representations when we discussed $(\mathbb{Z}, +)$, $(\mathbb{R} - \{0\}, \cdot)$, and $(\mathbb{Z}_3, +)$. Let’s focus on $(\mathbb{Z}_3, +)$ for a moment (the integers mod 3, where $e = 0$, $g_1 = 1$, $g_2 = 2$, with addition). Notice that we could alternatively define $e = 1$, $g_1 = e^{\frac{2\pi i}{3}}$, and $g_2 = e^{\frac{4\pi i}{3}}$, and let \star be multiplication. So, in the “representation” with $(0, 1, 2)$ and addition, we had for example

$$g_1 \star g_2 = (1 + 2) \bmod 3 = 3 \bmod 3 = 0 = e$$

whereas now with the multiplicative representation we have

$$g_1 \star g_2 = e^{\frac{2\pi i}{3}} \cdot e^{\frac{4\pi i}{3}} = e^{2\pi i} = e^0 = 1 = e$$

So the structure of the group is preserved in both representations.

We have two completely different representations of the same group. This idea of different ways of expressing the same group is of extreme importance, and we will be using it throughout the remainder of these notes.

We now see a more rigorous way of coming up with representations of a particular group. We begin by introducing some notation. For a group (G, \star) with elements g_1, g_2, \dots , we call the Representation of that group $D(G)$, so that the elements of G are $D(e)$, $D(g_1)$,

$D(g_2)$ (where each $D(g_i)$ is a matrix of some dimension). We then choose \star to be matrix multiplication. So, $D(g_i) \cdot D(g_j) = D(g_i \star g_j)$.

It may not seem that we have done anything profound at this point, but we most definitely have. Remember above that we encouraged seeing groups as abstract things, rather than in terms of specific numbers. This is because a group is fundamentally an *abstract* object. A group is not a specific set of numbers, but rather a set of abstract objects with a well-defined structure telling you how those elements relate to each other.

And the beauty of a representation D is that, via normal matrix multiplication, we have a sort of “lens”, made of familiar things (like numbers, matrices, or Easter eggs), through which we can see into this abstract world. And because $D(g_i) \cdot D(g_j) = D(g_i \star g_j)$, we aren’t losing any of the structure of the abstract group by using a representation.

So now that we have some notation, we can develop a formalism to figure out exactly what D is for an arbitrary group.

We will use Dirac vector notation, where the column vector

$$\bar{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \end{pmatrix} = |v\rangle$$

and the row vector

$$\bar{v}^T = (v_1 \quad v_2 \quad v_3 \quad \cdots) = \langle v|$$

So, the dot product between two vectors is

$$\bar{v} \cdot \bar{u} = (v_1 \quad v_2 \quad v_3 \cdots) \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \end{pmatrix} = v_1 u_1 + v_2 u_2 + v_3 u_3 + \cdots \equiv \langle v|u\rangle$$

Now, we proceed by relating each element of a finite discrete group to one of the standard orthonormal unit vectors:

$$e \rightarrow |e\rangle = |\hat{e}_1\rangle \quad g_1 \rightarrow |g_1\rangle = |\hat{e}_2\rangle \quad g_2 \rightarrow |g_2\rangle = |\hat{e}_3\rangle$$

And we define the way an element in a representation $D(G)$ acts on these vectors to be

$$D(g_i)|g_j\rangle = |g_i \star g_j\rangle$$

Now, we can build our representation. We will (from now on unless otherwise stated) represent the elements of a group G using matrices of various sizes, and the group operation \star will be standard matrix multiplication. The specific matrices that represent a given

element g_k of our group will be given by

$$[D(g_k)]_{ij} = \langle g_i | D(g_k) | g_j \rangle \quad (2.1)$$

As an example, consider again the group of order 2 (we wrote out the multiplication table above on page 18). First, we find the matrix representation of the identity, $[D(e)]_{ij}$,

$$\begin{aligned} [D(e)]_{11} &= \langle e | D(e) | e \rangle = \langle e | e \star e \rangle = \langle e | e \rangle = 1 \\ [D(e)]_{12} &= \langle e | D(e) | g_1 \rangle = \langle e | e \star g_1 \rangle = \langle e | g_1 \rangle = 0 \\ [D(e)]_{21} &= \langle g_1 | D(e) | e \rangle = \langle g_1 | e \star e \rangle = \langle g_1 | e \rangle = 0 \\ [D(e)]_{22} &= \langle g_1 | D(e) | g_1 \rangle = \langle g_1 | e \star g_1 \rangle = \langle g_1 | g_1 \rangle = 1 \end{aligned}$$

So, the matrix representation of the identity is $D(e) \doteq \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. It shouldn't be surprising that the identity element is represented by the identity matrix.

Next we find the representation of $D(g_1)$:

$$\begin{aligned} [D(g_1)]_{11} &= \langle e | D(g_1) | e \rangle = \langle e | g_1 \star e \rangle = \langle e | g_1 \rangle = 0 \\ [D(g_1)]_{12} &= \langle e | D(g_1) | g_1 \rangle = \langle e | g_1 \star g_1 \rangle = \langle e | e \rangle = 1 \\ [D(g_1)]_{21} &= \langle g_1 | D(g_1) | e \rangle = \langle g_1 | g_1 \star e \rangle = \langle g_1 | g_1 \rangle = 1 \\ [D(g_1)]_{22} &= \langle g_1 | D(g_1) | g_1 \rangle = \langle g_1 | g_1 \star g_1 \rangle = \langle g_1 | e \rangle = 0 \end{aligned}$$

So, the matrix representation of g_1 is $D(g_1) \doteq \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. It is straightforward to check that this is a true representation,

$$\begin{aligned} e \star e &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = e \quad \checkmark \\ e \star g_1 &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = g_1 \quad \checkmark \\ g_1 \star e &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = g_1 \quad \checkmark \\ g_1 \star g_1 &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = e \quad \checkmark \end{aligned}$$

Instead of considering the next obvious example, the group of order 3, consider the group S_3 from above (the multiplication table is on page 19). The identity representation $D(e)$ is easy — it is just the 6×6 identity matrix. We encourage you to work out the representation of $D(g_1)$ on your own, and check to see that it is

$$D(g_1) \doteq \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \quad (2.2)$$

All 6 matrices can be found this way, and multiplying them out will confirm that they do indeed satisfy the group structure of S_3 .

2.1.5 Reducibility and Irreducibility — A Preview

You have probably noticed that equation (2.1) will always produce a set of $n \times n$ matrices, where n is the order of the group. There is actually a name for this particular representation. *The $n \times n$ matrix representation of a group of order n is called the **Regular Representation**.* More generally, *the $m \times m$ matrix representation of a group (of any order) is called the **m -dimensional representation**.*

But, as we have seen, there is more than one representation for a given group (in fact, there are an infinite number of representations).

One thing we can immediately see is that any group that is Non-Abelian cannot have a 1×1 matrix representation. This is because scalars (1×1 matrices) always commute, whereas matrices in general do not.

We saw above in equation (2.2) that we can represent the group S_n by $n! \times n!$ matrices. Or, more generally, we can represent any group using $m \times m$ matrices, where m equals $\text{order}(G)$. This is the regular representation. But it turns out that it is usually possible to find representations that are “smaller” than the regular representation.

To pursue how this might be done, note that we are working with matrix representations of groups. In other words, we are representing groups in *linear spaces*. We will therefore be using a great deal of linear algebra to find smaller representations. This process, of finding a smaller representation, is called **Reducing** a representation. Given an arbitrary representation of some group, the first question that must be asked is “is there a smaller representation?” If the answer is yes, then the representation is said to be **Reducible**. If the answer is no, then it is **Irreducible**.

Before we dive into the more rigorous approach to reducibility and irreducibility, let’s consider a more intuitive example, using S_3 . In fact, we’ll stick with our three painted Easter eggs, R , O , and Y :

1. $e(ROY) = (ROY)$
2. $g_1(ROY) = (OYR)$
3. $g_2(ROY) = (YRO)$
4. $g_3(ROY) = (ORY)$
5. $g_4(ROY) = (YOR)$
6. $g_5(ROY) = (RYO)$

We will represent the set of eggs by a column vector $|E\rangle = \begin{pmatrix} R \\ O \\ Y \end{pmatrix}$.

Now, by inspection, what matrix would do to $|E\rangle$ what g_1 does to (ROY) ? In other words, how can we fill in the ?'s in

$$\begin{pmatrix} ? & ? & ? \\ ? & ? & ? \\ ? & ? & ? \end{pmatrix} \begin{pmatrix} R \\ O \\ Y \end{pmatrix} = \begin{pmatrix} O \\ Y \\ R \end{pmatrix}$$

to make the equality hold? A few moments thought will show that the appropriate matrix is

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} R \\ O \\ Y \end{pmatrix} = \begin{pmatrix} O \\ Y \\ R \end{pmatrix}$$

Continuing this reasoning, we can see that the rest of the matrices are

$$D(e) \doteq \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad D(g_1) \doteq \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \quad D(g_2) \doteq \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

$$D(g_3) \doteq \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad D(g_4) \doteq \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad D(g_5) \doteq \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

You can do the matrix multiplication to convince yourself that this is in fact a representation of S_3 .

So, in equation (2.2), we had a 6×6 matrix representation. Here, we have a new representation of consisting of 3×3 matrices. We have therefore “reduced” the representation. In the next section, we will look at more mathematically rigorous ways of reducing representations.

2.1.6 Algebraic Definitions

Before moving on, we must spend this section learning the definitions of several terms which are used in group theory.

*If H is a subset of G , denoted $H \subset G$, such that the elements of H form a group, then we say that H forms a **Subgroup** of G . We make this more precise with examples.*

Example 7

Consider (as usual) the group S_3 , with the elements labeled as before:

1. $g_0(ROY) = (ROY)$
2. $g_1(ROY) = (OYR)$
3. $g_2(ROY) = (YRO)$
4. $g_3(ROY) = (ORY)$
5. $g_4(ROY) = (YOR)$
6. $g_5(ROY) = (RYO)$

(where we are relabeling $g_0 \equiv e$ for later convenience). The multiplication table is given on page 19.

Notice that $\{g_0, g_1, g_2\}$ form a subgroup. You can see this by noticing that the upper left 9 boxes in the multiplication table (the g_0, g_1, g_2 rows and columns) all have only g_0 's, g_1 's, and g_2 's. So, here is a group of order 3 contained in S_3 .

Example 8

Consider the subset of S_3 consisting of g_0 and g_3 only. Both g_0 and g_3 are their own inverses, so the identity exists, and the group is closed. Therefore, we can say that $\{g_0, g_3\} \subset S_3$ is a subgroup of S_3 .

In fact, if you write out the multiplication table for g_0 and g_3 only, you will see that it is exactly equivalent to the group of order 2 considered above. This means that we can say that S_3 *contains* the group of order 2 (and we know from last time that there is only one such group, though there are an infinite number of representations of it). The way we understand this is that the abstract entity S_3 , of which there is only one, contains the group of order 2, of which there is only one. However, the *representations* of S_3 , of which there are an infinite number, will *each* contain the group of order 2 (of which there are also an infinite number of representations).

Example 9

Notice that the sets $\{g_0, g_3\}$, $\{g_0, g_4\}$, and $\{g_0, g_5\}$ (all $\subset S_3$), are all the same as the group of order 2. This means that S_3 actually contains exactly three copies of the group of order 2 in addition to the single copy of the group of order 3.

Again, this is speaking in terms of the abstract entity S_3 . We can see this through the “lens” of representation by the fact that *any* representation of S_3 will contain three different copies of the group of order 2.

Example 10

As a final example of subgroups, there are two subgroups of *any* group, no matter what the group. One is the subgroup consisting of only the identity, $\{g_0\} \subset G$. All groups contain this, but it is never very interesting.

Secondly, $\forall G, G \subset G$, and therefore G is always a subgroup of itself. We call these subgroups the “trivial” subgroups.

We now introduce another important definition.

If G is a group, and H is a subgroup of G ($H \subset G$), then

- The set $gH = \{g \star h | h \in H\}$ is called the **Left Coset** of H in G
- The set $Hg = \{h \star g | h \in H\}$ is called the **Right Coset** of H in G

There is a right (or left) coset for each element $g \in G$, though they are not necessarily all unique. This definition should be understood as follows; a coset is a *set* consisting of the elements of H all multiplied on the right (or left) by some element of G .

Example 11

For the subgroup $H = \{g_0, g_1\} \subset S_3$ discussed above, the left cosets are

$$\begin{aligned} g_0\{g_0, g_1\} &= \{g_0 \star g_0, g_0 \star g_1\} = \{g_0, g_1\} \\ g_1\{g_0, g_1\} &= \{g_1 \star g_0, g_1 \star g_1\} = \{g_1, g_2\} \\ g_2\{g_0, g_1\} &= \{g_2 \star g_0, g_2 \star g_1\} = \{g_2, g_0\} \\ g_3\{g_0, g_1\} &= \{g_3 \star g_0, g_3 \star g_1\} = \{g_3, g_4\} \\ g_4\{g_0, g_1\} &= \{g_4 \star g_0, g_4 \star g_1\} = \{g_4, g_5\} \\ g_5\{g_0, g_1\} &= \{g_5 \star g_0, g_5 \star g_1\} = \{g_5, g_3\} \end{aligned}$$

So, the left cosets of $\{g_0, g_1\}$ in S_3 are $\{g_0, g_1\}, \{g_1, g_2\}, \{g_2, g_0\}, \{g_3, g_4\}, \{g_4, g_5\},$ and $\{g_5, g_3\}$.

We can now understand the following definition. H is a **Normal Subgroup** of G if $\forall h \in H, g^{-1} \star h \star g \in H$. Or, in other words, if H denotes the subgroup, it is a normal subgroup if $gH = Hg$, which says that the left and right cosets are all equal.

As a comment, saying gH and Hg are equal doesn't mean that each individual element in the coset gH is equal to the corresponding element in Hg , but rather that the two cosets *contain* the same elements, regardless of their order. For example, if we had the cosets $\{g_i, g_j, g_k\}$ and $\{g_j, g_k, g_i\}$, they would be equal because they contain the same three elements.

This definition means that if you take a subgroup H of a group G , and you multiply the *entire set* on the left by some element of $g \in G$, the resulting set will contain the exact same elements it would if you had multiplied *on the right* by the same element $g \in G$. Here is an example to illustrate.

Example 12

Consider the order 2 subgroup $\{g_0, g_3\} \subset S_3$. Multiplying on the left by, say, g_4 , gives

$$g_4 \star \{g_0, g_3\} = \{g_4 \star g_0, g_4 \star g_3\} = \{g_4, g_2\}$$

And multiplying on the right by g_4 gives

$$\{g_0, g_3\} \star g_4 = \{g_0 \star g_4, g_3 \star g_4\} = \{g_4, g_1\}$$

So, because the final sets do not contain the same elements, $\{g_4, g_2\} \neq \{g_4, g_1\}$, we conclude that the subgroup $\{g_0, g_3\}$ is *not* a normal subgroup of S_3 .

Example 13

Above, we found that $\{g_0, g_1, g_2\} \subset S_3$ is a subgroup of order 3 in S_3 . To use a familiar label, remember that we previously called the group of order 3 $(\mathbb{Z}_3, +)$. So, dropping the '+', we refer to the group of order 3 as \mathbb{Z}_3 . Is this subgroup normal? We leave it to you to show that it is.

Example 14

Consider the group of integers under addition, $(\mathbb{Z}, +)$. And, consider the subgroup $\mathbb{Z}_{\text{even}} \subset \mathbb{Z}$, the even integers under addition (we leave it to you to show that this is indeed a group).

Now, take some odd integer n_{odd} and act on the left:

$$n_{\text{odd}} + \mathbb{Z}_{\text{even}} = \{n_{\text{odd}} + 0, n_{\text{odd}} \pm 2, n_{\text{odd}} \pm 4, \dots\}$$

and then on the right:

$$\mathbb{Z}_{\text{even}} + n_{\text{odd}} = \{0 + n_{\text{odd}}, \pm 2 + n_{\text{odd}}, \pm 4 + n_{\text{odd}}, \dots\}$$

Notice that the final sets are the same (because addition is commutative). So, $\mathbb{Z}_{\text{even}} \subset \mathbb{Z}$ is a normal subgroup.

With a little thought, you can convince yourself that *all* subgroups of Abelian groups are normal.

If G is a group and $H \subset G$ is normal, then the **Factor Group** of H in G , denoted G/H (read “ $G \bmod H$ ”), is the group with elements in the set $G/H \equiv \{gH \mid g \in G\}$. The group operation \star is understood to be

$$(g_i H) \star (g_j H) = (g_i \star g_j) H$$

Example 15

Consider again \mathbb{Z}_{even} . Notice that we can call $\mathbb{Z}_{\text{even}} = 2\mathbb{Z}$ because $2\mathbb{Z} = 2\{0, \pm 1, \pm 2, \pm 3, \dots\} = \{0, \pm 2, \pm 4, \dots\} = \mathbb{Z}_{\text{even}}$. We know that $2\mathbb{Z} \subset \mathbb{Z}$ is normal, so we can build the factor group $\mathbb{Z}/2\mathbb{Z}$ as

$$\mathbb{Z}/2\mathbb{Z} = \{0 + 2\mathbb{Z}, \pm 1 + 2\mathbb{Z}, \pm 2 + 2\mathbb{Z}, \dots\}$$

But, notice that

$$\begin{aligned} n_{\text{even}} + 2\mathbb{Z} &= \mathbb{Z}_{\text{even}} \\ n_{\text{odd}} + 2\mathbb{Z} &= \mathbb{Z}_{\text{odd}} \end{aligned}$$

So, the group $\mathbb{Z}/2\mathbb{Z}$ only has 2 elements; the set of all even integers, and the set of all odd integers. And we know from before that there is only one group of order 2, which we denote \mathbb{Z}_2 . So, we have found that $\mathbb{Z}/2\mathbb{Z} = \mathbb{Z}_2$.

You can also convince yourself of the more general result

$$\mathbb{Z}/n\mathbb{Z} = \mathbb{Z}_n$$

Example 16

Finally, we consider the factor groups G/G and G/e .

- G/G — The set $G = \{g_0, g_1, g_2, \dots\}$ will be the same coset for any element of G multiplied by it. Therefore this factor group consists of only one element, and therefore $G/G = e$, the trivial group.
- G/e — The set $\{e\}$ will be a unique coset for any element of G , and therefore $G/e = G$.

Something that might help you understand factor groups better is this: the factor group G/H is the group that is left over when everything in H is “collapsed” to the identity element. Think about the above examples in terms of this picture.

If G and H are both groups (not necessarily related in any way), then we can form the **Product Group**, denoted $K \equiv G \otimes H$, where an arbitrary element of K is (g_i, h_j) . If the group operation of G is \star_G , and the group operation of H is \star_H , then two elements of K are multiplied according to the rule

$$(g_i, h_j) \star_K (g_k, h_l) \equiv (g_i \star_G g_k, h_j \star_H h_l)$$

2.1.7 Reducibility Revisited

Now that we understand subgroups, cosets, normal subgroups, and factor groups, we can begin a more formal discussion of reducing representations. Recall that in deriving equation (2.1), we made the designation

$$g_0 \rightarrow |\hat{e}_1\rangle \quad g_1 \rightarrow |\hat{e}_2\rangle \quad g_2 \rightarrow |\hat{e}_3\rangle \quad \text{etc.}$$

This was used to create an $\text{order}(G)$ -dimensional Euclidian space which, while not having any “physical” meaning, and while obviously not possessing any structure similar to the group, was and will continue to be of great use to us.

We have an n -dimensional space spanned by the orthonormal vectors $|g_0\rangle, |g_1\rangle, \dots, |g_{n-1}\rangle$, where g_0 is understood to always refer to the identity element. This brings us to the first definition of this section. *For a group $G = \{g_0, g_1, g_2, \dots\}$, we call the **Algebra** of G the set*

$$\mathbb{C}[G] \equiv \left\{ \sum_{i=0}^{n-1} c_i |g_i\rangle \mid c_i \in \mathbb{C} \forall i \right\}$$

In other words, $\mathbb{C}[G]$ is the set of all possible linear combinations of the vectors $|g_i\rangle$ with complex coefficients.

We could have defined the algebra over \mathbb{Z} or \mathbb{R} , but we used \mathbb{C} for generality at this point.

Addition of two elements of $\mathbb{C}[G]$ is merely normal addition of linear combinations,

$$\sum_{i=0}^{n-1} c_i |g_i\rangle + \sum_{i=0}^{n-1} d_i |g_i\rangle = \sum_{i=0}^{n-1} (c_i + d_i) |g_i\rangle$$

This definition amounts to saying that, in the n -dimensional Euclidian space we have created, with $n = \text{order}(G)$, you can choose any point in the space with complex coefficients, and this will correspond to a particular linear combination of elements of G .

Now that we have defined an algebra, we can talk about group actions. Recall that the g_i 's don't act on the $|g_j\rangle$'s, but rather the representation $D(g_i)$ does. We define the action $D(g_i)$ on an element of $\mathbb{C}[G]$ as follows:

$$\begin{aligned} D(g_i) \cdot \sum_{j=0}^{n-1} c_j |g_j\rangle &= D(g_i) \cdot (c_0 |g_0\rangle + c_1 |g_1\rangle + \dots + c_{n-1} |g_{n-1}\rangle) \\ &= c_0 |g_i \star g_0\rangle + c_1 |g_i \star g_1\rangle + \dots + c_{n-1} |g_i \star g_{n-1}\rangle = \sum_{j=0}^{n-1} c_j |g_i \star g_j\rangle \end{aligned}$$

Previously, we discussed how elements of a group act on each other, and we also talked about how elements of a group act on some other object or set of objects (like three painted

eggs). We now generalize this notion to a set of q abstract objects a group can act on, denoted $M = \{m_0, m_1, m_2, \dots, m_{q-1}\}$. Just as before, we build a vector space, similar to the one above used in building an algebra. The orthonormal vectors here will be

$$m_0 \rightarrow |m_0\rangle, \quad m_1 \rightarrow |m_1\rangle, \quad \dots \quad m_{q-1} \rightarrow |m_{q-1}\rangle$$

This allows us to understand the following definition.

The set

$$\mathbb{C}M \equiv \left\{ \sum_{i=0}^{q-1} c_i |m_i\rangle \mid c_i \in \mathbb{C} \forall i \right\}$$

is called the **Module** of M . (We don't use the square brackets here to distinguish modules from algebras). In other words, the space spanned by the $|m_i\rangle$ is the module.

Example 17

Consider, once again, S_3 . However, we generalize from three eggs to three "objects" m_0, m_1 , and m_2 . So, $\mathbb{C}M$ is all points in the 3-dimensional space of the form $c_0|m_0\rangle + c_1|m_1\rangle + c_2|m_2\rangle$ with $c_i \in \mathbb{C} \forall i$.

Then, operating on a given point with, say, g_1 gives

$$g_1(c_0|m_0\rangle + c_1|m_1\rangle + c_2|m_2\rangle) = (c_0|g_1m_0\rangle + c_1|g_1m_1\rangle + c_2|g_1m_2\rangle)$$

and from the multiplication table on page 19, we know

$$g_1m_0 = m_1, \quad g_1m_1 = m_0, \quad g_1m_2 = m_2$$

So,

$$\begin{aligned} (c_0|g_1m_0\rangle + c_1|g_1m_1\rangle + c_2|g_1m_2\rangle) &= (c_0|m_1\rangle + c_1|m_0\rangle + c_2|m_2\rangle) \\ &= c_1|m_0\rangle + c_0|m_1\rangle + c_2|m_2\rangle \end{aligned}$$

So, the effect of g_1 was to swap c_1 and c_0 . This can be visualized geometrically as a reflection in the $c_0 = c_1$ plane in the 3-dimensional module space. We can visualize every element of G in this way. They each move points around the module space in a well-defined way.

This allows us to give the following definition. *If $\mathbb{C}V$ is a module, and $\mathbb{C}W$ is a subspace of $\mathbb{C}V$ that is closed under the action of G , then $\mathbb{C}W$ is an **Invariant Subspace** of $\mathbb{C}V$.*

Example 18

Working with S_3 , we know that S_3 acts on a 3-dimensional space spanned by

$$|m_0\rangle = (1, 0, 0)^T, \quad |m_1\rangle = (0, 1, 0)^T, \quad \text{and} \quad |m_2\rangle = (0, 0, 1)^T$$

Now, consider the subspace spanned by

$$c(|m_0\rangle + |m_1\rangle + |m_2\rangle) \quad (2.3)$$

where $c \in \mathbb{C}$, and c ranges over all possible complex numbers. If we restrict c to \mathbb{R} , we can visualize this more easily as the set of all points in the line through the origin defined by $\lambda(\hat{i} + \hat{j} + \hat{k})$ (where $\lambda \in \mathbb{R}$). You can write out the action of any element of S_3 on any point in this subspace, and you will see that they are unaffected. This means that the space spanned by (2.3) is an invariant subspace.

As a note, all modules $\mathbb{C}V$ have two trivial invariant subspaces.

- $\mathbb{C}V$ is a trivial invariant subspace of $\mathbb{C}V$
- $\mathbb{C}e$ is a trivial invariant subspace of $\mathbb{C}V$

Finally, we can give a more formal definition of reducibility. *If a representation D of a group G acts on the space of a module $\mathbb{C}M$, then the representation D is said to be **Reducible** if $\mathbb{C}M$ contains a non-trivial invariant subspace. If a representation is not reducible, it is **Irreducible**.*

We encouraged you to write out the entire regular representation of S_3 above. If you have done so, you may have noticed that every 6×6 matrix appeared with non-zero elements only in the upper left 3×3 elements, and the lower right 3×3 elements. The upper right and lower left are all 0. This means that, for every element of S_3 , there will never be any mixing of the first 3 dimensions with the last 3. So, there are two 3-dimensional invariant subspaces in the module for this particular representation of S_3 (the regular representation).

We can now begin to take advantage of the fact that representations live in linear spaces with the following definition.

*If V is any n -dimensional space spanned by n linearly independent basis vectors, and U and W are both subspaces of V , then we say that V is the **Direct Sum** of U and W if every vector $\bar{v} \in V$ can be written as the sum $\bar{v} = \bar{u} + \bar{w}$, where $\bar{u} \in U$ and $\bar{w} \in W$, and every operator X acting on elements of V can be separated into parts acting individually on U and W . The notation for this is $V = U \oplus W$.*

In order to make this clearer, if X_n is an $n \times n$ matrix, it is the direct sum of $m \times m$ matrix A_m and $k \times k$ matrix B_k , denoted $X_n = A_m \oplus B_k$, iff X is in **Block Diagonal** form,

$$X_n = \begin{pmatrix} A_m & 0 \\ 0 & B_k \end{pmatrix}$$

where $n = m + k$, and A_m , B_k , and the 0's are understood as matrices of appropriate dimension.

We can generalize the previous definition as follows,

$$X_n = A_{n_1} \oplus B_{n_2} \oplus \cdots \oplus C_{n_k} = \begin{pmatrix} A_{n_1} & 0 & \cdots & 0 \\ 0 & B_{n_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \vdots & C_{n_k} \end{pmatrix}$$

where $n = n_1 + n_2 + \cdots + n_k$.

Example 19

Let $A_3 = \begin{pmatrix} 1 & 1 & -2 \\ -1 & 5 & \pi \\ -17 & 4 & 11 \end{pmatrix}$, and let $B_2 = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$. Then,

$$B_2 \oplus A_3 = \begin{pmatrix} 1 & 2 & 0 & 0 & 0 \\ 3 & 4 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & -2 \\ 0 & 0 & -1 & 5 & \pi \\ 0 & 0 & -17 & 4 & 11 \end{pmatrix}$$

To take stock of what we have done so far, we have talked about algebras, which are the vector spaces spanned by the elements of a group, and about modules, which are the vector spaces that representations of groups act on. We have also defined invariant subspaces as follows: Given some space and some group that acts on that space, moving the points around in a well-defined way, an invariant subspace is a subspace which always contains the same points. The group doesn't remove any points from that subspace, and it doesn't add any points to it. It merely moves the points around *inside* that subspace. Then, we defined a representation as reducible if there are any non-trivial invariant subspaces in the space that the group acts on.

And what this amounts to is the following: a representation of any group is reducible if it can be written in block diagonal form.

But this leaves the question of what we mean when we say "can be written". How can you "rewrite" a representation? This leads us to the following definition. *Given a matrix D and a non-singular matrix S , the linear transformation*

$$D \rightarrow D' = S^{-1}DS$$

*is called a **Similarity Transformation**.*

Then, we can give the following definition. *Two matrices related by a similarity transformation are said to be **Equivalent**.*

Because similarity transformations are linear transformations, if $D(G)$ is a representation of G , then so is $S^{-1}DS$ for literally *any* non-singular matrix S . To see this, if $g_i \star g_j = g_k$,

then $D(g_i)D(g_j) = D(g_k)$, and therefore

$$S^{-1}D(g_i)S \cdot S^{-1}D(g_j)S = S^{-1}D(g_i)D(g_j)S = S^{-1}D(g_k)S$$

So, if we have a representation that isn't in block diagonal form, how can we figure out if it is reducible? We must look for a matrix S that will transform it into block diagonal form.

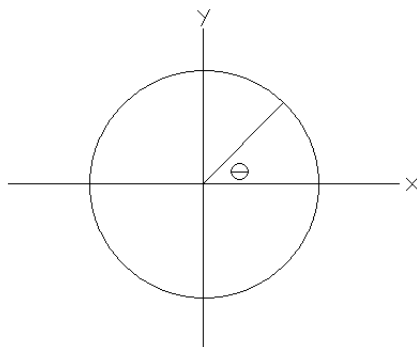
You likely realize immediately that this is not a particularly easy thing to do by inspection. It turns out that there is a very straightforward and systematic way of taking a given representation and determining whether or not it is reducible, and if so, what the irreducible representations are.

However, the details of how this can be done, while very interesting, are not necessary for the agenda of these notes. Therefore, for the sake of brevity, we will not pursue them. What is important is that you understand not only the details of general group theory and representation theory (which we outlined above), but also the concept of what it means for a group to be reducible or irreducible.

2.2 Introduction to Lie Groups

In section 2.1, we considered groups which are of finite order and discrete, which allowed us to write out a multiplication table.

Here, however, we examine a different type of group. Consider the unit circle, where each point on the circle is specified by an angle θ , measured from the positive x -axis.



We will refer to the point at $\theta = 0$ as the “starting point” (like *ROY* was for the Easter eggs). Now, just as we considered all possible orientations of (*ROY*) that left the eggs lined up, we consider all possible rotations the wheel can undergo. With the eggs there

were only 6 possibilities. Now however, for the wheel there are an infinite number of possibilities for θ (any real number $\in [0, 2\pi)$).

And note that if we denote the set of all angles as G , then all the rotations obey closure ($\theta_1 + \theta_2 = \theta_3 \in G, \forall \theta_1, \theta_2 \in G$), associativity (as usual), identity ($0 + \theta = \theta + 0 = \theta$), and inverse (the inverse of θ is $-\theta$).

So, we have a group that is *parameterized* by a continuous variable θ . So, we are no longer talking about g_i 's, but about $g(\theta)$.

Notice that this particular group (the circle) is Abelian, which is why we can (temporarily) use addition to represent it. Also, note that we obviously cannot make a multiplication table because the order of this group is ∞ .

One simple representation is the one we used above: taking θ and using addition. A more familiar (and useful) representation is the Euler matrix $g(\theta) \doteq \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}$ with the usual matrix multiplication:

$$\begin{pmatrix} \cos \theta_1 & \sin \theta_1 \\ -\sin \theta_1 & \cos \theta_1 \end{pmatrix} \begin{pmatrix} \cos \theta_2 & \sin \theta_2 \\ -\sin \theta_2 & \cos \theta_2 \end{pmatrix} \quad (2.4)$$

$$= \begin{pmatrix} \cos \theta_1 \cos \theta_2 - \sin \theta_1 \sin \theta_2 & \cos \theta_1 \sin \theta_2 + \sin \theta_1 \cos \theta_2 \\ -\sin \theta_1 \cos \theta_2 - \cos \theta_1 \sin \theta_2 & -\sin \theta_1 \sin \theta_2 + \cos \theta_1 \cos \theta_2 \end{pmatrix} \quad (2.5)$$

$$= \begin{pmatrix} \cos(\theta_1 + \theta_2) & \sin(\theta_1 + \theta_2) \\ -\sin(\theta_1 + \theta_2) & \cos(\theta_1 + \theta_2) \end{pmatrix} \quad (2.6)$$

This will prove to be a much more useful representation than θ with addition.

Groups that are parameterized by one or more continuous variables like this are called **Lie Groups**. Of course, the true definition of a Lie group is much more rigorous (and complicated), and that definition should eventually be understood. However, the definition we have given will suffice for the purposes of these notes.

2.2.1 Classification of Lie Groups

The usefulness of group theory is that groups represent a mathematical way to make changes to a system while leaving *something* about the system unchanged. For example, we moved (*ROY*) around, but the structure “3 eggs with different colors lined up” was preserved. With the circle, we rotated it, but it still maintained its basic structure as a circle. It is in this sense that group theory is a study of **Symmetry**. No matter which of “these” transformations you do to the system, “this” stays the same—this is symmetry.

To see the usefulness of this in physics, recall Noether’s Theorem (section 1.1.2). When you do a *symmetry* transformation to a Lagrangian, you get a conserved quantity. Think

back to the Lagrangian for the projectile (1.2). The transformation $x \rightarrow x + \epsilon$ was a symmetry because ϵ could take any value, and the Lagrangian was unchanged (note that ϵ forms the Abelian group $(\mathbb{R}, +)$).

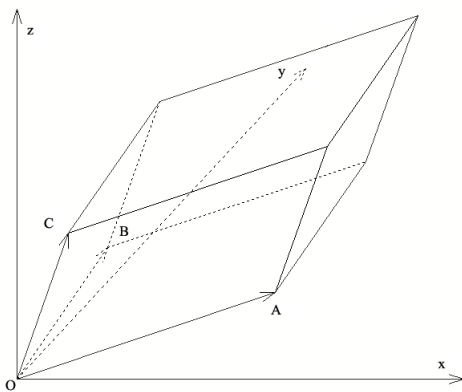
So, given a Lagrangian, which represents the structure of a physical system, a symmetry represents a way of changing the Lagrangian while preserving that structure. The particular preserved part of the system is the conserved quantity j we discussed in sections 1.1.2 and 1.1.6. And as you have no doubt noticed, nearly all physical processes are governed by **Conservation Laws**: conservation of momentum, energy, charge, spin, etc.

So, group theory, and in particular Lie group theory, gives us an extremely powerful way of understanding and classifying symmetries, and therefore conserved charges. And because it allows us to understand conserved charges, group theory can be used to understand the entirety of the physics in our universe.

We now begin to classify the major types of Lie groups we will be working with in these notes. To start, we consider the most general possible Lie group in an arbitrary number of dimensions, n . This will be the group that, for any point p in the n -dimensional space, can continuously take it *anywhere* else in the space. All that is preserved is that the points in the space stay in the space. This means that we can have literally any $n \times n$ matrix, or *linear* transformation, so long as the matrix is invertible (non-singular). Thus, in n dimensions the largest and most general Lie group is the group of all $n \times n$ non-singular matrices. We call this group $GL(n)$, or the **General Linear** group. The most general field of numbers to take the elements of $GL(n)$ from is \mathbb{C} , so we begin with $GL(n, \mathbb{C})$. This is the group of all $n \times n$ non-singular matrices with complex elements. The preserved quantity is that all points in \mathbb{C}^n stay in \mathbb{C}^n .

The most obvious subgroup of $GL(n, \mathbb{C})$ is $GL(n, \mathbb{R})$, or the set of all $n \times n$ invertible matrices with real elements. This leaves all points in \mathbb{R}^n in \mathbb{R}^n .

To find a further subgroup, recall from linear algebra and vector calculus that in n dimensions, you can take n vectors at the origin such that for a parallelepiped, we could obtain



Then, if you arrange the components of the n vectors into the rows (or columns) of a matrix, the determinant of that matrix will be the volume of the parallelepiped.

So, consider now the set of all General Linear transformations that transform all vectors from the origin (or in other words, points in the space) in such a way that the volume of the corresponding parallelepiped is preserved. This will demand that we only consider General Linear matrices with determinant 1. Also, the set of all General Linear matrices with unit determinant will form a group because of the general rule $\det |A \cdot B| = \det |A| \cdot \det |B|$. So, if $\det |A| = 1$ and $\det |B| = 1$, then $\det |A \cdot B| = 1$. We call this subgroup of $GL(n, \mathbb{C})$ the **Special Linear** group, or $SL(n, \mathbb{C})$. The natural subgroup of this is $SL(n, \mathbb{R})$. This group preserves not only the points in the space (as GL did), but also the volume, as described above.

Now, consider the familiar transformations on vectors in n -dimensional space of generalized Euler angles. These are transformations that rotate all points around the origin. These rotation transformations leave the radius squared (r^2) invariant. And, because $\bar{r}^2 = \bar{r}^T \cdot \bar{r}$, if we transform with a rotation matrix R , then $\bar{r} \rightarrow \bar{r}' = R\bar{r}$, and $\bar{r}^T \rightarrow \bar{r}'^T = \bar{r}^T R^T$, so $\bar{r}'^T \cdot \bar{r}' = \bar{r}^T R^T \cdot R\bar{r}$. But, as we said, we are demanding that the radius squared be invariant under the action of R , and so we demand $\bar{r}^T R^T \cdot R\bar{r} = \bar{r}^T \cdot \bar{r}$. So, the constraint we are imposing is $R^T \cdot R = \mathbb{I}$, which implies $R^T = R^{-1}$. This tells us that the rows and columns of R are orthogonal. Therefore, we call the group of generalized rotations, or generalized Euler angles in n dimensions, $O(n)$, or the **Orthogonal** group. We don't specify \mathbb{C} or \mathbb{R} here because it will be understood that we are always talking about \mathbb{R} .

Also, note that because $\det |R^T \cdot R| = \det |\mathbb{I}| \Rightarrow (\det |R|)^2 = 1 \Rightarrow \det |R| = \pm 1$. We again denote the subgroup with $\det |R| = +1$ the **Special Orthogonal** group, or $SO(n)$. To understand what this means, consider an orthogonal matrix with determinant -1 , such as

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

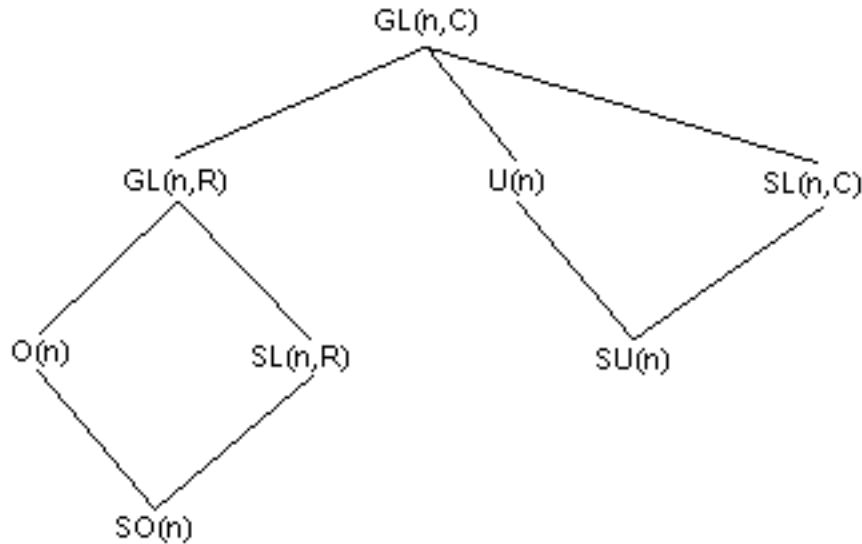
This matrix is orthogonal, and therefore is an element of the group $O(3)$, but the determinant is -1 . This matrix will take the point $(x, y, z)^T$ to the point $(x, y, -z)^T$. This changes the handedness of the system (the right hand rule will no longer work). So, if we limit ourselves to $SO(n)$, we are preserving the space, the radius, the volume, and the handedness of the space.

For vectors in \mathbb{C} space, we do not define orthogonal matrices (although we could). Instead, we discuss the complex version of the radius, where instead of $\bar{r}^2 = \bar{r}^T \cdot \bar{r}$, we have $\bar{r}^2 = \bar{r}^\dagger \cdot \bar{r}$, where the dagger denotes the Hermitian conjugate, $\bar{r}^\dagger = (\bar{r}^*)^T$, where \star denotes complex conjugate.

So, with the elements in R being in \mathbb{C} , we have $\bar{r} \rightarrow R\bar{r}$, and $\bar{r}^\dagger \rightarrow \bar{r}^\dagger R^\dagger$. So, $\bar{r}^\dagger \cdot \bar{r} \rightarrow \bar{r}^\dagger R^\dagger \cdot R\bar{r}$, and by the same argument as above with the orthogonal matrices, this demands that

$R^\dagger \cdot R = \mathbb{I}$, or $R^\dagger = R^{-1}$. We denote such matrices **Unitary**, and the set of all such $n \times n$ invertible matrices form the group $U(n)$. Again, we understand the unitary groups to have elements in \mathbb{C} , so we don't specify that. And, we will still have a subset of unitary matrices R with $\det |R| = 1$ called $SU(n)$, the **Special Unitary** groups.

We can summarize the hierarchy we have just described in the following diagram:



We will now describe one more category of Lie groups before moving on. We saw above that the group $SO(n)$ preserves the radius squared in real space. In coordinates, this means that $\vec{r}^2 = x_1^2 + x_2^2 + \cdots + x_n^2$, or more generally the dot product $\vec{x} \cdot \vec{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$ is preserved.

However, we can generalize this to form a group action that preserves not the radius squared, but the value (switching to indicial notation for the dot product) $x^a y_a = -x_1 y_1 - x_2 y_2 - \cdots - x_m y_m + x_{m+1} y_{m+1} + \cdots + x_{m+n} y_{m+n}$. We call the group that preserves this quantity

$SO(m, n)$. The space we are working in is still \mathbb{R}^{m+n} , but we are making transformations that preserve something different than the radius.

Note that $SO(m, n)$ will have an $SO(m)$ subgroup and an $SO(n)$ subgroup, consisting of rotations in the first m and last n components separately.

Finally, notice that the specific group of this type, $SO(1, 3)$, is the group that preserves the value $s^2 = -x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4$, or written more suggestively, $s^2 = -c^2t^2 + x^2 + y^2 + z^2$. Therefore, the group $SO(1, 3)$ is the **Lorentz Group**. Any action that is invariant under $SO(1, 3)$ is said to be a Lorentz Invariant theory (as all theories should be). We will find that thinking of Special Relativity in these terms, rather than in the terms of Part I, will be much more useful.

It should be noted that there are many other types of Lie groups. We have limited ourselves to the ones we will be working with in these notes.

2.2.2 Generators

Now that we have a good “birds eye view” of Lie groups, we can begin to pick apart the details of how they work.

As we said before, a Lie group is a group that is parameterized by a set of continuous parameters, which we call α_i for $i = 1, \dots, n$, where n is the number of parameters the group depends on. The elements of the group will then be denoted $g(\alpha_i)$.

Because all groups include an identity element, we will choose to parameterize them in such a way that $g(\alpha_i)|_{\alpha_i=0} = e$, the identity element. So, if we are going to talk about representations, $D_n(g(\alpha_i))|_{\alpha_i=0} = \mathbb{I}$, where \mathbb{I} is the $n \times n$ identity matrix for whatever dimension (n) representation we want.

Now, take α_i to be very small with $\delta\alpha_i \ll 1$. So, $D_n(g(0 + \delta\alpha_i))$ can be Taylor expanded:

$$D_n(g(\delta\alpha_i)) = \mathbb{I} + \delta\alpha_i \frac{\partial D_n(g(\alpha_i))}{\partial \alpha_i} \Big|_{\alpha_i=0} + \dots$$

The terms $\frac{\partial D_n}{\partial \alpha_i} \Big|_{\alpha_i=0}$ are extremely important, and we give them their own expression:

$$X_i \equiv -i \frac{\partial D_n}{\partial \alpha_i} \Big|_{\alpha_i=0} \quad (2.7)$$

(we have included the $-i$ in order to make X_i Hermitian, which will be necessary later).

So, the representation for infinitesimal $\delta\alpha_i$ is then

$$D_n(\delta\alpha_i) = \mathbb{I} + i\delta\alpha_i X_i + \dots$$

(where we have switched our notation from $D_n(g(\alpha))$ to $D_n(\alpha)$ for brevity).

The X_i 's are constant matrices which we will determine later.

Now, let's say that we want to see what the representation will look like for a finite value of α_i rather than an infinitesimal value. A finite transformation will be the result of an infinite number of infinitesimal transformations. Or in other words, $\alpha_i = N\delta\alpha_i$ as $N \rightarrow \infty$. So, $\delta\alpha_i = \frac{\alpha_i}{N}$, and an infinite number of infinitesimal transformations is

$$\lim_{N \rightarrow \infty} (1 + i\delta\alpha_i X_i)^N = \lim_{N \rightarrow \infty} \left(1 + i\frac{\alpha_i}{N} X_i\right)^N$$

If you expand this out for several values of N , you will see that it is exactly

$$\lim_{N \rightarrow \infty} \left(1 + i\frac{\alpha_i}{N} X_i\right)^N = e^{i\alpha_i X_i}$$

We call the X_i 's the **Generators** of the group, and there is one for each parameter required to specify a particular element of the group. For example, consider $SO(3)$, the group of rotations in 3 dimensions. We know from vector calculus that an element of $SO(3)$ requires 3 angles, usually denoted θ , ϕ , and ψ . Therefore, $SO(3)$ will require 3 generators, which will be denoted X_θ , X_ϕ , and X_ψ . We will discuss how the generators can be found soon.

In general, there will be several (in fact, infinite) different sets of X_i 's that define a given group (just as there are an infinite number of representations of any finite group). What we will find is that up to a similarity transformation, a particular set of generators defines a particular representation of a group.

So, $D_n(\alpha_i) = e^{i\alpha_i X_i}$ for any group (the i index in the exponent is understood to be summed over all parameters and generators). The best way to think of the parameter space for the group is as a vector space, where the generators describe the behavior near the identity, but form a basis for the entire vector space. By analogy, think of the unit vectors \hat{i} , \hat{j} , and \hat{k} in \mathbb{R}^3 . They are defined at the origin, but they can be combined with real numbers/parameters to specify any arbitrary point in \mathbb{R}^3 . In the same way, the generators are the "unit vectors" of the parameter space (which in general is a much more complicated space than Euclidian space), and the parameters (like θ , ϕ , and ψ) specify where in the parameter space you are in terms of the generators. That point in the parameter space will then correspond to a particular element of the group.

We call the number of generators of a group (or equivalently the number of parameters necessary to specify an element), the **Dimension** of the group. For example, the dimension of $SO(3)$ is 3. The dimension of $SO(2)$ (rotations in the plane) however is only 1 (only θ is needed), so there will be only one generator.

2.2.3 Lie Algebras

In section 2.1 we discussed algebras. An algebra is a space spanned by elements of the group with \mathbb{C} coefficients parameterizing the Euclidian space we defined. Obviously we can't define an algebra in the same way for Lie groups, because the elements are continuous. But, as discussed in the last section, a particular element of a Lie group is defined by the values of the parameters in the parameter space spanned by the generators. We will see that the generators will form the algebras for Lie groups.

Consider two elements of the same group with generators X_i , one with parameter values α_i and the other with parameter values β_i . The product of the 2 elements will then be $e^{i\alpha_i X_i} e^{i\beta_j X_j}$. Because we are assuming this is a group, we know that the product must be an element of the group (due to closure), and therefore the product must be specified by some set of parameters δ_k , so $e^{i\alpha_i X_i} e^{i\beta_j X_j} = e^{i\delta_k X_k}$. Note that the product won't necessarily simply be $e^{i\alpha_i X_i} e^{i\beta_j X_j} = e^{i(\alpha_i X_i + \beta_j X_j)}$ because the generators are matrices and therefore don't in general commute.

So, we want to figure out what δ_i will be in terms of α_i and β_i . We do this as follows.

$$i\delta_i X_k = \ln(e^{i\delta_k X_k}) = \ln(e^{i\alpha_i X_i} e^{i\beta_j X_j}) = \ln(1 + e^{i\alpha_i X_i} e^{i\beta_j X_j} - 1) \equiv \ln(1 + x)$$

where we have defined $x \equiv e^{i\alpha_i X_i} e^{i\beta_j X_j} - 1$. We will proceed by expanding only to second order in α_i and β_j , though the result we will obtain will hold at arbitrary order. By Taylor expanding the exponential terms,

$$\begin{aligned} e^{i\alpha_i X_i} e^{i\beta_j X_j} - 1 &= (1 + i\alpha_i X_i + \frac{1}{2}(i\alpha_i X_i)^2 + \dots)(1 + i\beta_j X_j + \frac{1}{2}(i\beta_j X_j)^2 + \dots) - 1 \\ &= 1 + i\beta_j X_j - \frac{1}{2}(\beta_j X_j)^2 + i\alpha_i X_i - \alpha_i X_i \beta_j X_j - \frac{1}{2}(\alpha_i X_i)^2 - 1 \\ &= i(\alpha_i X_i + \beta_j X_j) - \alpha_i X_i \beta_j X_j - \frac{1}{2}((\alpha_i X_i)^2 + (\beta_j X_j)^2) \end{aligned}$$

Then, using the general Taylor expansion $\ln(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots$, and again

keeping terms only to second order in α and β , we have

$$\begin{aligned}
x - \frac{x^2}{2} &= \left[i(\alpha_i X_i + \beta_j X_j) - \alpha_i X_i \beta_j X_j - \frac{1}{2}[(\alpha_i X_i)^2 + (\beta_j X_j)^2] \right] \\
&\quad - \frac{1}{2} \left[i(\alpha_i X_i + \beta_j X_j) - \alpha_i X_i \beta_j X_j - \frac{1}{2}[(\alpha_i X_i)^2 + (\beta_j X_j)^2] \right]^2 \\
&= i(\alpha_i X_i + \beta_j X_j) - \alpha_i X_i \beta_j X_j - \frac{1}{2}[(\alpha_i X_i)^2 + (\beta_j X_j)^2] \\
&\quad - \frac{1}{2} \left[-(\alpha_i X_i + \beta_j X_j)(\alpha_i X_i + \beta_j X_j) \right] \\
&= i(\alpha_i X_i + \beta_j X_j) - \alpha_i X_i \beta_j X_j - \frac{1}{2}[(\alpha_i X_i)^2 + (\beta_j X_j)^2] \\
&\quad + \frac{1}{2} \left[(\alpha_i X_i)^2 + (\beta_j X_j)^2 + \alpha_i \beta_j (X_i X_j + X_j X_i) \right] \\
&= i(\alpha_i X_i + \beta_j X_j) + \frac{1}{2} \alpha_i \beta_j (X_j X_i - X_i X_j) \\
&= i(\alpha_i X_i + \beta_j X_j) - \frac{1}{2} \alpha_i \beta_j [X_i, X_j] \\
&= i(\alpha_i X_i + \beta_j X_j) - \frac{1}{2} [\alpha_i X_i, \beta_j X_j]
\end{aligned}$$

So finally we can see

$$i\delta_k X_k = i(\alpha_i X_i + \beta_j X_j) - \frac{1}{2} [\alpha_i X_i, \beta_j X_j]$$

or

$$e^{i\alpha_i X_i} e^{i\beta_j X_j} = e^{i(\alpha_i X_i + \beta_j X_j) - \frac{1}{2} [\alpha_i X_i, \beta_j X_j]} \quad (2.8)$$

Equation (2.8) is called the **Baker-Campbell-Hausdorff** formula, and it is one of the most important relations in group theory and in physics. Notice that, if the generators commute, this reduces to the normal equation for multiplying exponentials. You can think of equation (2.8) as the generalization of the normal exponential multiplication rule.

Now, it is clear that the commutator $[X_i, X_j]$ must be proportional to some linear combination of the generators of the group (because of closure). So, it must be the case that

$$[X_i, X_j] = i f_{ijk} X_k \quad (2.9)$$

for some set of constants f_{ijk} . These constants are called the **Structure Constants** of the group, and if they are completely known, the commutation relations between all the generators are known, and so the entire group can be determined in any representation you want.

The generators, under the specific commutation relations defined by the structure constants, form the **Lie Algebra** of the group, and it is this commutation structure which forms the structure of the Lie group.

2.2.4 The Adjoint Representation

We will talk about several representations for each group we discuss, but we will mention a very important one now. We mentioned before that the structure constants f_{ijk} completely determine the entire structure of the group.

We begin by using the Jacobi identity,

$$[X_i, [X_j, X_k]] + [X_j, [X_k, X_i]] + [X_k, [X_i, X_j]] = 0 \quad (2.10)$$

(if you aren't familiar with this identity, try multiplying it out. You will find that it is identically true — all the terms cancel exactly). But, from equation (2.9), we can write

$$[X_i, [X_j, X_k]] = if_{jka}[X_i, X_a] = if_{jka}f_{iab}X_b$$

Plugging this into (2.10) we get

$$\begin{aligned} if_{jka}f_{iab}X_b + if_{kia}f_{jab}X_b + if_{ija}f_{kab}X_b &= 0 \\ \Rightarrow (f_{jka}f_{iab} + f_{kia}f_{jab} + f_{ija}f_{kab})iX_b &= 0 \\ \Rightarrow f_{jka}f_{iab} + f_{kia}f_{jab} + f_{ija}f_{kab} &= 0 \end{aligned} \quad (2.11)$$

So, if we define the matrices

$$[T^a]_{bc} \equiv -if_{abc} \quad (2.12)$$

then it is easy to show that (2.11) leads to

$$[T^a, T^b] = if_{abc}T^c$$

So, the structure constants themselves form a representation of the group (as defined by (2.12)). We call this representation the **Adjoint Representation**, and it will prove to be extremely important.

Notice that the indices labeling the rows and columns in (2.12) each run over the same values as the indices labeling the T matrices. This tells us that the adjoint representation is made of $n \times n$ matrices, where n is the dimension of the group, or the number of parameters in the group. For example, $SO(3)$ requires 3 parameters to specify an element (θ, ϕ, ψ) , so the adjoint representation of $SO(3)$ will consist of 3×3 matrices. $SO(2)$ on the other hand is Abelian, and therefore all of the structure constants vanish. Therefore there is no adjoint representation of $SO(2)$.

We now go on to consider several specific groups in detail.

2.2.5 $SO(2)$

We start by looking at an extremely simple group, $SO(2)$. This is the group of rotations in the plane that leaves $\bar{r}^2 = x^2 + y^2 = \begin{pmatrix} x & y \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} = \bar{v}^T \cdot \bar{v}$ invariant. So for some generator X (which we will now find) of $SO(2)$, $\bar{v} \rightarrow R(\theta)\bar{v} = e^{i\theta X}\bar{v}$, and $\bar{v}^T \rightarrow \bar{v}^T e^{i\theta X^T}$. So, expanding to first order only, $\bar{v}^T e^{i\theta X^T} e^{i\theta X} \bar{v} = \bar{v}^T (1 + i\theta X^T + i\theta X) \bar{v} = \bar{v}^T \cdot \bar{v} + \bar{v}^T i\theta (X + X^T) \bar{v}$. And because we demand that r^2 be invariant, we demand that $X + X^T = 0 \Rightarrow X = -X^T$. So, X must be antisymmetric. Therefore we take

$$X \equiv \frac{1}{i} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

(the $\frac{1}{i}$ is included to balance the i we inserted in equation (2.7) to ensure that X is Hermitian).

So, an arbitrary element of $SO(2)$ will be

$$\begin{aligned} e^{i\theta X} &= e^{i\theta \frac{1}{i} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}} = e^{\theta \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}} \\ &= \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^0 + \theta \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^1 + \frac{1}{2}\theta^2 \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^2 + \dots \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \theta \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} - \frac{1}{2}\theta^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \frac{1}{3!}\theta^3 \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} + \dots \\ &= \begin{pmatrix} 1 - \frac{1}{2}\theta^2 + \dots & \theta - \frac{1}{3!}\theta^3 + \dots \\ -(\theta - \frac{1}{3!}\theta^3 + \dots) & 1 - \frac{1}{2}\theta^2 + \dots \end{pmatrix} \\ &= \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \end{aligned}$$

which is exactly what we would expect for a matrix describing rotations in the plane.

Also, notice that because $SO(2)$ is Abelian, the commutation relations trivially vanish ($[X, X] \equiv 0$), and so all of the structure constants are zero.

Now that we have found an explicit example of a generator, and seen an example of how generators relate to group elements, we move on to slightly more complicated examples.

2.2.6 $SO(3)$

We could easily generalize the argument from the proceeding section and find the generators of $SO(3)$ in the same way, but in order to illustrate more clearly how generators work, we will approach $SO(3)$ differently by working backwards. Above, we found the generators and used them to calculate the group elements. Here, we begin with the known

group elements of $SO(3)$, which are just the standard Euler matrices for rotations in 3-dimensional space:

$$R_x(\phi) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \phi & \sin \phi \\ 0 & -\sin \phi & \cos \phi \end{pmatrix} \quad (2.13)$$

$$R_y(\psi) = \begin{pmatrix} \cos \psi & 0 & -\sin \psi \\ 0 & 1 & 0 \\ \sin \psi & 0 & \cos \psi \end{pmatrix} \quad (2.14)$$

$$R_z(\theta) = \begin{pmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (2.15)$$

Now, recall the definition of the generators, equation (2.7). We can use it to find the generators of $SO(3)$, which we will denote J_x , J_y , and J_z .

$$J_x = \frac{1}{i} \frac{dR_x(\phi)}{d\phi} \Big|_{\phi=0} = \frac{1}{i} \begin{pmatrix} 0 & 0 & 0 \\ 0 & -\sin \phi & \cos \phi \\ 0 & -\cos \phi & \sin \phi \end{pmatrix} \Big|_{\phi=0} = \frac{1}{i} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}$$

And similarly

$$J_y = \frac{1}{i} \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad J_z = \frac{1}{i} \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

You can plug these into the exponentials with the appropriate parameters (ϕ , ψ , or θ) and find that $e^{i\phi J_x}$, $e^{i\psi J_y}$, and $e^{i\theta J_z}$ reproduce (2.13), (2.14), and (2.15), respectively.

Furthermore, you can multiply out the commutators to find

$$[J_x, J_y] = iJ_z, \quad [J_y, J_z] = iJ_x, \quad [J_z, J_x] = iJ_y$$

or

$$[J_i, J_j] = i\epsilon_{ijk}J_k$$

which tells us that the structure constants for $SO(3)$ are

$$f_{ijk} = \epsilon_{ijk} \quad (2.16)$$

where ϵ_{ijk} is the totally antisymmetric tensor. The structure constants being non-zero is consistent with $SO(3)$ being a non-Abelian group.

2.2.7 $SU(2)$

We will approach $SU(2)$ yet another way: by starting with the structure constants. It turns out they are the same as the structure constants for $SO(3)$:

$$f_{ijk} = \epsilon_{ijk} \quad (2.17)$$

To see why, recall that $SU(2)$ are rotations in two complex dimensions. The most general form of such a matrix $U \in SU(2)$ is $U = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. The “Special” part of $SU(2)$ demands that the determinant be equal to 1, or

$$ad - bc = 1$$

and the “Unitary” part demands that $U^{-1} = U^\dagger$. So,

$$U^{-1} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = U^\dagger = \begin{pmatrix} a^* & c^* \\ b^* & d^* \end{pmatrix}$$

or in other words,

$$U = \begin{pmatrix} a & b \\ -b^* & a^* \end{pmatrix}$$

where we demand $|a|^2 + |b|^2 = 1$.

Both a and b are in \mathbb{C} , and therefore have 2 real components each, so U has 4 real parameters. The constraint $|a|^2 + |b|^2 = 1$ fixes one of them, leaving 3 real parameters, just like in $SO(3)$. This is a loose explanation of why $SU(2)$ and $SO(3)$ have the same structure constants. They are both rotational groups with 3 real parameters.

This also tells us that $SU(2)$ will have 3 generators.

2.2.8 $SU(2)$ and Physical States

The elements of any Lie group (in a d -dimensional representation consisting of $d \times d$ matrices) will act on vectors, just like the 3×3 matrices representing S_3 acted on $(R \ O \ Y)^T$ in section 2.1.5. The most natural way to understand the space a Lie group acts on is to study the eigenvectors and eigenvalues of the generators of the representation you are using (the reason for this is beyond the scope of these notes at this point, but will become more clear as we proceed). These eigenvectors will obviously form a basis of the eigenspace of the physical space the group is acting on.

Using similarity transformations, one or more of the generators of a Lie group can be diagonalized. For now, trust us that with $SU(2)$, it is only possible to diagonalize one of the

three generators at a time (you may convince yourself of this by studying the commutation relations). We will call the generators of $SU(2)$ J^1 , J^2 , and J^3 , and by convention we take J^3 to be the diagonal one. So, consequently, the eigenvectors of J^3 will be the basis vectors of the physical vector space upon which $SU(2)$ acts.

Now, we know that J^3 (whatever it is ... we don't know at this point) will in general have more than one eigenvalue. Let's call the greatest eigenvalue of J^3 (whatever it is) j , and the eigenvectors of J^3 will be denoted $|j; m\rangle$ (the first j is merely a label — the second value describes the vector), where m is the eigenvalue of the eigenvector. The eigenvector corresponding to the greatest eigenvalue j will obviously then be $|j; j\rangle$. So, $J^3|j; j\rangle = j|j; j\rangle$, or more generally $J^3|j; m\rangle = m|j; m\rangle$.

Now let's assume that we know $|j; j\rangle$. There is a trick we can employ to find the rest of the states. Define the following linear combinations of the generators:

$$J^\pm \equiv \frac{1}{\sqrt{2}}(J^1 \pm iJ^2) \quad (2.18)$$

Now, using the fact that the $SU(2)$ generators obey the commutation relations in equation (2.16), it is easy to show the following relations,

$$[J^2, J^\pm] = \pm J^\pm \quad \text{and} \quad [J^+, J^-] = J^3 \quad (2.19)$$

Notice that, because by definition J^i are all Hermitian, we have

$$(J^-)^\dagger = J^+ \quad (2.20)$$

Consider some arbitrary eigenvector $|j; m\rangle$. We know the eigenvalue of this will be m , so $J^3|j; m\rangle = m|j; m\rangle$. But now let's create some new state by acting on $|j; m\rangle$ with either of the operators (2.18). The new state will be $J^\pm|j; m\rangle$, but what will the J^3 eigenvalue be? Using the commutation relations in (2.19),

$$J^3 J^\pm|j; m\rangle = (\pm J^\pm + J^\pm J^3)|j; m\rangle = (m \pm 1)J^\pm|j; m\rangle$$

So, the vector $J^+|j; m\rangle$ is the eigenvector with eigenvalue $m + 1$, and the vector $J^-|j; m\rangle$ is the eigenvector with eigenvalue $m - 1$.

If we have some arbitrary eigenvector $|j; m\rangle$, we can use J^\pm to move up or down to the eigenvector with the next highest or lowest eigenvalue. For this reason, J^\pm are called the **Raising** and **Lowering** operators. They raise and lower the eigenvalue of the state by one.

Clearly, the eigenvector with the greatest eigenvalue j , with eigenvector $|j; j\rangle$, cannot be raised any higher, so we define $J^+|j; j\rangle \equiv 0$. We will see that there is also a lowest eigenvalue j' , so we similarly define $J^-|j; j'\rangle \equiv 0$.

Now, considering once again $|j; j\rangle$. We know that if we operate on this state with J^- , we will get the eigenvector with the eigenvalue $j - 1$. But, we don't know exactly what this

state will be (knowing the eigenvalue doesn't mean we know the actual state). But, we know it will be proportional to $|j; j-1\rangle$. So, we set $J^-|j; j\rangle = N_j|j; j-1\rangle$, where N_j is the proportionality constant. To find N_j , we take the inner product (and using (2.20)):

$$\langle j; j|J^+J^-|j; j\rangle = |N_j|^2\langle j; j-1|j; j-1\rangle$$

But we can also write

$$\begin{aligned}\langle j; j|J^+J^-|j; j\rangle &= \langle j; j|(J^+J^- - J^-J^+)|j; j\rangle = \langle j; j|[J^+, J^-]|j; j\rangle \\ &= \langle j; j|J^3|j; j\rangle = j\langle j; j|j; j\rangle = j\end{aligned}\quad (2.21)$$

where we used the fact that $J^+|j; j\rangle = 0$ to get the first equality, and (2.19) to get the third equality. We also assumed that $|j; j\rangle$ is normalized.

So, (2.21) tells us

$$\langle j; j-1|j; j-1\rangle = 1 \iff N_j \equiv \sqrt{j} \quad (2.22)$$

And our normalized state is therefore

$$\frac{J^-}{N_j}|j; j\rangle = \frac{J^-}{\sqrt{j}}|j; j\rangle = |j; j-1\rangle$$

Repeating this to find N_{j-1} , we have

$$\begin{aligned}|N_{j-1}|^2\langle j; j-2|j; j-2\rangle &= \langle j; j-1|J^+J^-|j; j-1\rangle \\ &= \langle j; j|\frac{J^+}{\sqrt{j}}J^+J^-\frac{J^-}{\sqrt{j}}|j; j\rangle \\ &= \frac{1}{j}\langle j; j|J^+J^+J^-J^-|j; j\rangle \\ &= \frac{1}{j}\langle j; j|J^+(J^3 + J^-J^+)J^-|j; j\rangle \\ &= \frac{1}{j}\langle j; j|(J^+J^3J^- + J^+J^-J^+J^-)|j; j\rangle \\ &= \frac{1}{j}\langle j; j|(J^+(-J^- + J^-J^3) + J^+J^-(J^3 + J^-J^+))|j; j\rangle \\ &= \frac{1}{j}[\langle j; j|(-J^+J^- + jJ^+J^- + jJ^+J^-)|j; j\rangle] \\ &= \frac{1}{j}\langle j; j|(-[J^+, J^-] + 2j[J^+, J^-])|j; j\rangle \\ &= \frac{1}{j}\langle j; j|(-J^3 + 2jJ^3)|j; j\rangle \\ &= \frac{1}{j}(2j^2 - j) = 2j - 1\end{aligned}\quad (2.23)$$

So, $|N_{j-1}|^2 = 2j - 1$, or $N_{j-1} = \sqrt{2j - 1}$.

We can continue this process, and we will find that the general result is

$$N_{j-k} = \frac{1}{\sqrt{2}} \sqrt{(2j-k)(k+1)} \quad (2.24)$$

and the general states are defined by

$$|j; j-k\rangle = \frac{1}{N_{j-k}} (J^-)^k |j; j\rangle$$

Notice that these expressions recover (2.22) and (2.23) for $k=0$ and $k=1$, respectively.

Furthermore, notice that when $k=2j$,

$$N_{j-2j} = \frac{1}{\sqrt{2}} \sqrt{(2j-2j)(2j+1)} \equiv 0$$

So, the state $|j; j-k\rangle|_{k=2j} = |j; -j\rangle$ is the state with the lowest eigenvalue, and by definition $J^-|j; -j\rangle \equiv 0$.

So, in a general representation of $SU(2)$, we have $2j+1$ states:

$$\{j, j-1, j-2, \dots, -j+2, -j+1, -j\}$$

This therefore demands that $j = \frac{n}{2}$ for some integer n . In other words, the highest eigenvalue of an $SU(2)$ eigenvector can be $0, \frac{1}{2}, 1, \frac{3}{2}, 2$, etc.

Furthermore, using these states, it is easy to show

$$\begin{aligned} \langle j; m' | J^3 | j; m \rangle &= m \delta_{m', m} \\ \langle j; m' | J^+ | j; m \rangle &= \frac{1}{\sqrt{2}} \sqrt{(j+m+1)(j-m)} \delta_{m', m+1} \\ \langle j; m' | J^- | j; m \rangle &= \frac{1}{\sqrt{2}} \sqrt{(j+m)(j-m+1)} \delta_{m', m-1} \end{aligned} \quad (2.25)$$

2.2.9 $SU(2)$ for $j = \frac{1}{2}$

We will skip the $j=0$ case because it is trivial (though we will discuss it later when we return to physics).

For $j = \frac{1}{2}$, the two eigenvalues of J^3 will be $\frac{1}{2}$ and $\frac{1}{2} - 1 = -\frac{1}{2}$. So, denoting the J^3 generator of $SU(2)$ when $j = \frac{1}{2}$ as $J_{1/2}^3$, we have

$$J_{1/2}^3 = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$$

Now, inverting (2.18) to get

$$J^1 = \frac{1}{\sqrt{2}}(J^- + J^+) \quad \text{and} \quad J^2 = \frac{i}{\sqrt{2}}(J^- - J^+)$$

and using the standard matrix equation $[J_j^a]_{m',m} = \langle j, m' | J^a | j, m \rangle$, and the explicit products in (2.25), we can find (for example)

$$\left\langle \frac{1}{2}; -\frac{1}{2} \left| J^1 \right| \frac{1}{2}; -\frac{1}{2} \right\rangle = \left\langle \frac{1}{2}; -\frac{1}{2} \left| \frac{1}{\sqrt{2}}(J^- + J^+) \right| \frac{1}{2}; -\frac{1}{2} \right\rangle = \cdots = 0$$

So $[J^1]_{11} = 0$. Then,

$$\left\langle \frac{1}{2}; -\frac{1}{2} \left| J^1 \right| \frac{1}{2}; \frac{1}{2} \right\rangle = \left\langle \frac{1}{2}; -\frac{1}{2} \left| \frac{1}{\sqrt{2}}(J^- + J^+) \right| \frac{1}{2}; \frac{1}{2} \right\rangle = \cdots = \frac{1}{2}$$

So $[J^1]_{12} = \frac{1}{2}$.

We can continue this to find all the elements for each generator for $j = 1/2$. The final result will be

$$J_{1/2}^1 = \frac{1}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \frac{\sigma^1}{2}, \quad J_{1/2}^2 = \frac{1}{2} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} = \frac{\sigma^2}{2}, \quad J_{1/2}^3 = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \frac{\sigma^3}{2} \quad (2.26)$$

where the σ^i matrices are the **Pauli Spin Matrices**. This is no accident! We will discuss this in much, much more detail later, but for now recall that we said that $SU(2)$ is the group of transformations in 2-dimensional complex space (with one of the real parameters fixed, leaving 3 real parameters). We are going to see that $SU(2)$ is the group which represents quantum mechanical spin, where j is the value of the spin of the particle. In other words, particles with spin $1/2$ are described by the $j = 1/2$ representation (the 2×2 representation in (2.26)), and particles with spin 1 are described by the $j = 1$ representation, and so on. In other words, $SU(2)$ describes quantum mechanical spin in 3 dimensions in the same way that $SO(3)$ describes normal “spin” in 3 dimensions. We will talk about the physical implications, reasons, and meaning of this later.

However, as a warning, be careful at this point not to think too much in terms of physics. You have likely covered $SU(2)$ in great detail in a quantum mechanics course (though you may not have known it was called “ $SU(2)$ ”), but the approach we are taking here has a different goal than what you have likely seen before. The properties of $SU(2)$ we are seeing here are actually very, very specific and simplified illustrations of much deeper concepts in Lie groups, and in order to understand particle physics we must understand Lie groups in this way. So for now, try to fight the temptation to merely understand everything we are doing in terms of the physics you have seen before and learn this as we are presenting it: pure mathematics. We will focus on how it applies to physics later, in its fuller and more fundamental way than introductory quantum mechanics makes apparent.

2.2.10 $SU(2)$ for $j = 1$

You can follow the same procedure we used above to find

$$J_1^1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad J_1^2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad J_1^3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} \quad (2.27)$$

Notice that only J_1^3 is diagonal (as before), and that the eigenvalues are $\{1, 0, -1\}$, or $\{j, j-1, j-2 = -j\}$ as we'd expect.

2.2.11 $SU(2)$ for Arbitrary j

For any given j , we have 3 generators J_j^1, J_j^2 , and J_j^3 , and for whatever dimension ($d = 2j + 1$) the physical space we are working in, we have d eigenvectors

$$|j; j\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad |j; j-1\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad |j; j-2\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \dots \quad |j; -j\rangle = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

with eigenvalues $\{j, j-1, j-2, \dots, -j\}$, respectively.

Then, for any j , we can form the linear combinations $J_j^\pm \equiv \frac{1}{\sqrt{2}}(J_j^1 \pm iJ_j^2)$. For example, for $j = 1/2$ these are

$$J_{1/2}^+ = \frac{1}{\sqrt{2}} \left[\frac{1}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + \frac{i}{2} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \right] = \frac{1}{2\sqrt{2}} \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

and similarly

$$J_{1/2}^- = \dots = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

So, the two $j = 1/2$ eigenvectors will be $|\frac{1}{2}; \frac{1}{2}\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $|\frac{1}{2}; -\frac{1}{2}\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. So,

$$J_{1/2}^+ \left| \frac{1}{2}; \frac{1}{2} \right\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 0$$

$$J_{1/2}^- \left| \frac{1}{2}; -\frac{1}{2} \right\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 0$$

and similarly

$$J_{1/2}^- \left| \frac{1}{2}; \frac{1}{2} \right\rangle = \left| \frac{1}{2}; -\frac{1}{2} \right\rangle$$

$$J_{1/2}^+ \left| \frac{1}{2}; -\frac{1}{2} \right\rangle = \left| \frac{1}{2}; \frac{1}{2} \right\rangle$$

which is exactly what we would expect.

The same calculation can be done for the $j = 1$ case and we will find the same results, except that the $j = 1$ state (the first eigenvector) can be lowered *twice*. The first time $J_{1/2}^-$ acts it takes it to the state with eigenvalue 0, and the second time it acts it takes it to the state with eigenvalue -1 . Acting a third time will destroy the state (take it to 0). Analogously, the lowest state, with eigenvalue $j = -1$ can be raised twice.

We can do the same analysis for any $j = \text{integer or half integer}$.

As we said before, we interpret j as the quantum mechanical spin of a particle, and the group $SU(2)$ describes that rotation. It is important to recognize that quantum spin is not a rotation through spacetime (it would be described by $SO(3)$ if it was), but rather through the mathematically constructed spinor space. We will talk more about this space later.

So for a given particle with spin, we can talk about both its rotation through physical spacetime using $SO(3)$, as well as its rotation through complex spinor space using $SU(2)$. Both values will be physically measurable and will be conserved quantities. The total angular momentum of the particle will be the combination of both spin and spacetime angular momentum. Again, we will talk much more about the spin of physical particles when we return to a discussion of physics. We only mention this now to give a preview of where this is going. However, spin is not the only thing $SU(2)$ describes. We will also find that it is the group which governs the weak nuclear force (whereas $U(1)$ describes the electromagnetic force, and $SU(3)$ describes the strong force ... much, much more on this later).

2.2.12 Root Space

As a comment before beginning this section, it is likely that you will find this to be the most difficult section of these notes. The material here is both extremely difficult (especially the first time it is encountered), and extremely important to the development of particle physics. In fact, this section is the most central to what will come later in these notes. If the contents are not clear you are encouraged to read this section multiple times until it becomes clear. It may also be helpful to study this section while looking closely at the examples in the sections forming the remainder of this part of these notes. They illustrate the point of where we are going with all of this.

We saw in the previous section that we can view the physical space that a group is acting on by using the eigenvectors of the diagonal generators as a basis. These eigenvectors can be arranged in order of decreasing eigenvalue. Then, the non-diagonal generators can be used to form linear combinations that act as raising and lowering operators, which transform one eigenvector to another, changing the eigenvalue by an amount defined by the commutation relations of the generators.

We now see that this generalizes very nicely.

An arbitrary Lie group is defined in terms of its generators. As we said at the end of section 2.2.2, it is best to think of the generators as being analogous to the basis vectors spanning some space. Of course, the space the generators span is much more complicated than \mathbb{R}^n in general, but the generators span the space the same way. In this sense, the generators form a linear vector space. So, we must define an inner product for them. For reasons that are beyond the scope of these notes, we will choose the generators and inner product so that, for generators T^a and T^b ,

$$\langle T^a, T^b \rangle \equiv \frac{1}{\kappa} \text{Tr} (T^a T^b) = \delta^{ab} \quad (2.28)$$

where κ is some normalization constant.

Also, in the set of generators of a Lie group, there will be a closed subalgebra of generators which all commute with each other, but not with generators outside of this subalgebra. In other words, this is the set of generators which can be simultaneously diagonalized through some similarity transformation. For $SU(2)$, we saw that there was only one generator in this subalgebra which we chose to be J_j^3 (recall that a matrix will only commute with all other matrices if it is equal to the identity matrix times a constant, whereas two diagonal matrices will always commute regardless of what their diagonal elements are).

Let's say that a particular Lie group has N generators total, or is an N -dimensional group. Then, let's say that there are $M < N$ generators in the mutually commuting subalgebra. We call those M generators the **Cartan Subalgebra**, and the generators in it are called **Cartan Generators**. We define the number M as the **Rank** of the group.

By convention we will label the Cartan generators H^i ($i = 1, \dots, M$) and the non-Cartan generators E^i ($i = 1, \dots, N - M$).

For example, with $SU(2)$ we had $H^1 = J_j^3$, and $E^1 = J_j^1$, $E^2 = J_j^2$.

Before moving on, we point out that this should seem familiar. If you think back to an introductory class in quantum mechanics, recall that we always choose some set of variables that all commute with each other (usually we choose *either* position or momentum because $[x, p] \neq 0$). Then, we expand the physical states in terms of the position *or* momentum eigenvectors. Here, we are doing the exact same thing, only in a much more general context.

Now, the H^i 's are simultaneously diagonalized, so we will write the physical states in

terms of their eigenvalues. In an n -dimensional representation D_n , the generators are $n \times n$ matrices, so the eigenvectors are n -dimensional. So, there will be a total of n eigenvectors, and each will have one eigenvalue with each of the M Cartan generators H^i . So, for each of these eigenvectors, which we temporarily denote $|j\rangle$, for $j = 1, \dots, n$, we have the M eigenvalues with M Cartan generators, which we call t_j^i (where $j = 1, \dots, n$ labels the eigenvectors, and $i = 1, \dots, M$ labels the eigenvalues), and we form what is called a **Weight Vector**

$$\bar{t}_j \equiv \begin{pmatrix} t_j^1 \\ t_j^2 \\ t_j^3 \\ \vdots \\ t_j^M \end{pmatrix} \quad (2.29)$$

where $j = 1, \dots, n$. The individual components of these vectors, the t_j^i 's, are called the **Weights**.

So for a given representation D_n , we now denote the state $|D_n; \bar{t}_j\rangle$ (instead of $|j\rangle$). So, our eigenvalues will be

$$H^i |D_n; \bar{t}_j\rangle = t_j^i |D_n; \bar{t}_j\rangle \quad (2.30)$$

As we mentioned before, the adjoint representation is a particularly important representation. If you do not remember the details of the adjoint representation, go reread section 2.2.4. Here, the generators are defined by equation (2.12), $[T^a]_{bc} \equiv -if_{abc}$. Recall that each index runs from 1 to N , so that the generators in the adjoint representation are $N \times N$ matrices, and the eigenvectors are N -dimensional.

Also, as a point of nomenclature, weights in the adjoint representation are called **Roots**, and the corresponding vectors (as in (2.29)) are called **Root Vectors**.

This means that there is exactly one eigenvector for each generator, and therefore one root vector for each generator. So, in equation (2.29), $j = 1, \dots, N$. We make this more obvious by explicitly assigning each eigenvector to a generator as follows. First, because we now have the same number of generators, eigenvectors, and root vectors, we label the generators by the root vectors $T^{\bar{t}_j}$ instead of T^j . Also, we now refer to general eigenstates as $|Adj; T^{\bar{t}_j}\rangle$, where $j = 1, \dots, N$ and \bar{t}_j is the M -dimensional root vector corresponding to $T^{\bar{t}_j}$. And, we also divide the states $|Adj; T^{\bar{t}_j}\rangle$ into two groups: those corresponding to the M Cartan generators $|Adj; H^{\bar{h}_j}\rangle$ (where $j = 1, \dots, M$ and \bar{h}_j is the M -dimensional root vector corresponding to $H^{\bar{h}_j}$), and those corresponding to the $N - M$ non-Cartan generators $|Adj; E^{\bar{e}_j}\rangle$ (where $j = 1, \dots, N - M$ and \bar{e}_j is the M -dimensional root vector corresponding to $E^{\bar{e}_j}$).

Don't be alarmed by the superscripts being vectors. We are using this notation for later convenience, and $T^{\bar{t}_i}$ here means the same thing T^j did before (the j^{th} generator). This

notation, which we use only for the adjoint representation, is simply taking advantage of the fact that in the adjoint representation, the total number of generators, the number of eigenvectors of the Cartan generators, the dimension of the representation, and the number of weight/root vectors is the same.

Also, with the adjoint representation states $|Adj; T^{\bar{t}_j}\rangle$, we can use equation (2.28) to define the inner product between states as

$$\langle Adj; T^{\bar{t}_j} | Adj; T^{\bar{t}_k} \rangle = \frac{1}{\kappa} \text{Tr} (T^{\bar{t}_j} T^{\bar{t}_k}) = \delta^{jk} \quad (2.31)$$

We will make use of this equation soon.

The matrix elements of a given generator will then be given by the familiar equation

$$-if_{abc} = [T^{\bar{t}_a}]_{bc} \equiv \langle Adj; T^{\bar{t}_b} | T^{\bar{t}_a} | Adj; T^{\bar{t}_c} \rangle$$

We want to know what an arbitrary generator $T^{\bar{t}_a}$ will do to an arbitrary state $|Adj; T^{\bar{t}_b}\rangle$ in the adjoint representation. So,

$$\begin{aligned} T^{\bar{t}_a} |Adj; T^{\bar{t}_b}\rangle &= \sum_c |Adj; T^{\bar{t}_c}\rangle \langle Adj; T^{\bar{t}_c} | T^{\bar{t}_a} | Adj; T^{\bar{t}_b}\rangle = \sum_c |Adj; T^{\bar{t}_c}\rangle [T^{\bar{t}_a}]_{cb} \\ &= \sum_c |Adj; T^{\bar{t}_c}\rangle (-if_{acb}) = \sum_c if_{abc} |Adj; T^{\bar{t}_c}\rangle \end{aligned}$$

And, because there is exactly one eigenvector for each generator, the state $|Adj; T^{\bar{t}_c}\rangle$ corresponds to the generator T^c . And because we know that

$$if_{abc} T^{\bar{t}_c} = [T^{\bar{t}_a}, T^{\bar{t}_b}]$$

(where c is understood to be summed) by definition of the structure constants, we can infer that

$$T^{\bar{t}_a} |Adj; T^{\bar{t}_b}\rangle = \sum_c if_{abc} |Adj; T^{\bar{t}_c}\rangle = |Adj; [T^{\bar{t}_a}, T^{\bar{t}_b}]\rangle \quad (2.32)$$

where $[T^{\bar{t}_a}, T^{\bar{t}_b}]$ is simply the commutator.

The derivation of equation (2.32) is extremely important, and it is vital that you understand it. However, it is also one of the more difficult results of this already difficult section. You are therefore encouraged (again) to read through this section, comparing it with examples several times until it becomes clear.

So, let's apply this to combinations of the two types of generators we have, $H^{\bar{h}_a}$'s and $E^{\bar{e}_a}$'s. If we have a Cartan generator acting on a state corresponding to a Cartan generator, we have (from equation (2.30))

$$H^{\bar{h}_a} |Adj; H^{\bar{h}_b}\rangle = h_b^a |Adj; H^{\bar{h}_b}\rangle$$

But from (2.32) we have

$$H^{\bar{h}_a} |Adj; H^{\bar{h}_b}\rangle = |Adj; [H^{\bar{h}_a}, H^{\bar{h}_b}]\rangle$$

By definition, the Cartan generators commute, so $[H^{\bar{t}_a}, H^{\bar{t}_b}] \equiv 0$, and therefore

$$\bar{h}_b \equiv 0 \quad (2.33)$$

So we can drop them from our notation, leaving the eigentstates corresponding to non-Cartan generators denoted $|Adj; H^j\rangle$.

On the other hand, if we have a Cartan generator acting on an eigenstate corresponding to a non-Cartan generator, equation (2.30) gives

$$H^a |Adj; E^{\bar{e}_b}\rangle = e_b^a |Adj; [H^a, E^{\bar{e}_b}]\rangle \quad (2.34)$$

And equation (2.32) gives

$$H^a |Adj; E^{\bar{e}_b}\rangle = |Adj; [H^a, E^{\bar{e}_b}]\rangle \quad (2.35)$$

Now, we don't know *a priori* what $[H^a, E^{\bar{e}_b}]$ is, but comparing (2.34) and (2.35), we see

$$|Adj; e_b^a E^{\bar{e}_b}\rangle = |Adj; [H^a, E^{\bar{e}_b}]\rangle$$

And because we know that each of these vectors corresponds directly to the generators, we have the final result

$$[H^a, E^{\bar{e}_b}] = e_b^a E^{\bar{e}_b} \quad (2.36)$$

Now we want to know what a non-Cartan generator does to a given eigentstate. Consider an arbitrary state $|Adj; T^{\bar{t}_b}\rangle$ with H^c eigenvalue t_b^c . We can act on this with $E^{\bar{e}_a}$ to create the new state $E^{\bar{e}_a} |Adj; T^{\bar{t}_b}\rangle$. So what will the H^c eigenvalue of this new state be? Using (2.36),

$$\begin{aligned} H^c E^{\bar{e}_a} |Adj; T^{\bar{t}_b}\rangle &= (H^c E^{\bar{e}_a} - E^{\bar{e}_a} H^c + E^{\bar{e}_a} H^c) |Adj; T^{\bar{t}_b}\rangle = ([H^c, E^{\bar{e}_a}] + E^{\bar{e}_a} H^c) |Adj; T^{\bar{t}_b}\rangle \\ &= (e_a^c E^{\bar{e}_a} + E^{\bar{e}_a} t_b^c) |Adj; T^{\bar{t}_b}\rangle = (t_b^c + e_a^c) E^{\bar{e}_a} |Adj; T^{\bar{t}_b}\rangle \\ &= (\bar{t}_b + \bar{e}_a)^c E^{\bar{e}_a} |Adj; T^{\bar{t}_b}\rangle \end{aligned} \quad (2.37)$$

So, by acting on the one of the eigenstates with a non-Cartan generator $E^{\bar{e}_a}$, we have shifted the H^c eigenvalue by one of the coordinates of the root vector. What this means is that the non-Cartan generators play a role analogous to the raising and lowering operators we saw in $SU(2)$, except instead of merely shifting the state “up” and “down”, it moves the states around through some M -dimensional space.

From this, we can also see that if there is an operator that can transform from one state to another, there must be a corresponding operator that will make the opposite transformation. Therefore, for every operator $E^{\bar{e}_a}$, we expect to have the operator $E^{-\bar{e}_a}$, and corresponding eigenstate $|Adj; E^{-\bar{e}_a}\rangle$.

Finally, consider the state $E^{\bar{e}_a}|Adj; E^{-\bar{e}_a}\rangle$. We know from (2.32) that $E^{\bar{e}_a}|Adj; E^{-\bar{e}_a}\rangle = |Adj; [E^{\bar{e}_a}, E^{-\bar{e}_a}]\rangle$. The eigenvalue of this state can be found using equation (2.37):

$$H^b E^{\bar{e}_a}|Adj; E^{-\bar{e}_a}\rangle = (-\bar{e}_a + \bar{e}_a)^b E^{\bar{e}_a}|Adj; E^{-\bar{e}_a}\rangle \equiv 0$$

But according to equation (2.33), states with 0 eigenvalue are states corresponding to Cartan generators. Therefore we conclude that the state $E^{\bar{e}_a}|Adj; E^{-\bar{e}_a}\rangle$ is proportional to some linear combination of the Cartan states,

$$E^{\bar{e}_a}|Adj; E^{-\bar{e}_a}\rangle = \sum_b N_b |Adj; H^b\rangle \quad (2.38)$$

where the N_b 's are the constants of proportionality. To find the constants N_b , we follow an approach similar to the one we used in deriving (2.24). Taking the inner product and using (2.32),

$$\begin{aligned} \langle Adj; H^c | E^{\bar{e}_a} | Adj; E^{-\bar{e}_a} \rangle &= \sum_b N_b \langle Adj; H^c | Adj; H^b \rangle = \sum_b N_b \delta^{cb} = N_c \quad (2.39) \\ \Rightarrow \langle Adj; H^c | Adj; [E^{\bar{e}_a}, E^{-\bar{e}_a}] \rangle &= N_c \end{aligned}$$

Then, using (2.31)

$$\begin{aligned} \langle Adj; H^c | Adj; [E^{\bar{e}_a}, E^{-\bar{e}_a}] \rangle &= \frac{1}{\kappa} \text{Tr} (H^c [E^{\bar{e}_a}, E^{-\bar{e}_a}]) \\ &= \frac{1}{\kappa} \text{Tr} (E^{-\bar{e}_a} [H^c, E^{\bar{e}_a}]) \\ &= \frac{1}{\kappa} e_a^c \text{Tr} (E^{-\bar{e}_a} E^{\bar{e}_a}) \\ &= e_a^c \delta^{aa} \\ &= e_a^c \end{aligned}$$

So,

$$N_c = e_a^c$$

And therefore equation (2.38) is now

$$E^{\bar{e}_a}|Adj; E^{-\bar{e}_a}\rangle = |Adj; [E^{\bar{e}_a}, E^{-\bar{e}_a}]\rangle = e_a^b |Adj; H^b\rangle$$

where the sum over b is understood. This leads to our final result,

$$[E^{\bar{e}_a}, E^{-\bar{e}_a}] = e_a^b H^b \quad (2.40)$$

Though we did all of this using the adjoint representation we have seen before, this structure is the same in any representation, and therefore everything we have said is valid in any D_n . We worked in the adjoint simply because that makes the results easiest to obtain. The extensive use we made of labeling the eigenvectors with the generators can

only be done in the adjoint representation because only in the adjoint does the number of eigenvectors equal the number of eigenstates. However, this will not be a problem. The important results from this section are (2.36) and (2.40), which are true in any representation. Part of what we will do later is find these structures in other representations.

The importance of the ideas in this section cannot be stressed enough. However, the material is somewhat abstract. So, we consider a few examples of how all this works.

2.2.13 Adjoint Representation of $SU(2)$

We now illustrate what we did in section 2.2.12 with $SU(2)$. We will work in the adjoint representation to make the correspondence with section 2.2.12 as transparent as possible.

$SU(2)$ has 3 generators, and therefore the adjoint representation will consist of 3×3 matrices. This is simply the $j = 1$ representation, which we wrote out in equation (2.27).

First, it is easy to verify that (2.28) and (2.31) hold for $\kappa = 2$.

Next we look at the eigenstates. We know they will be the normal vectors

$$v_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

(we will relabel them to be consistent with section 2.2.12 shortly).

Obviously only J_1^3 is diagonal, so $SU(2)$ has rank $M = 1$. We define

$$H^1 = J_1^3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

$$E^1 = J_1^1 = \frac{1}{2} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad E^2 = J_1^2 = \frac{1}{2} \begin{pmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{pmatrix}$$

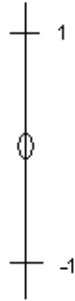
Because the rank is 1, the root vectors will be 1-dimensional vectors, or scalars. We find them easily by finding the eigenvalues of each eigenvector with H^1 :

$$H^1 v_1 = (+1)v_1, \quad H^1 v_2 = (0)v_2, \quad H^1 v_3 = (-1)v_3$$

So the root vectors are

$$\bar{t}_1 = t_1 = +1 \quad \bar{t}_2 = t_2 = 0 \quad \bar{t}_3 = t_3 = -1 \quad (2.41)$$

We can graph these on the real line as shown below,



Now our initial guess will be to associate v_3 with $J_1^3 = H^1$, and then $v_1 = E^1$ and $v_2 = E^2$. But we want to exploit what we learned in section 2.2.12, and therefore we must make sure that (2.36) and (2.40) hold.

Starting with (2.36), we check (leaving the tedious matrix multiplication up to you)

$$[H^1, E^1] = \dots = \frac{1}{2} \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} \quad (2.42)$$

$$[H^1, E^2] = \dots = -\frac{i}{2} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad (2.43)$$

But we have a problem. According to (2.36), $[H^1, E^i]$ should be proportional to E^i , but this is not the case here. However notice that in (2.42),

$$\frac{1}{2} \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} = iE^2$$

and in (2.43),

$$-\frac{i}{2} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} = -iE^1$$

Writing this more suggestively,

$$[H^1, E^1] = iE^2, \quad [H^1, iE^2] = E^1 \quad (2.44)$$

So, if we take the linear combinations of equations (2.44), we get $[H^1, \alpha E^1 \pm \beta iE^2] = \beta E^1 \pm \alpha iE^2$, which has the correct form of equation (2.36) as long as $\alpha = \beta$. Therefore we are now working with the operators $E^\pm \equiv \alpha(E^1 \pm iE^2)$.

Now we seek to impose (2.40). We start by evaluating

$$\begin{aligned} [E^+, E^-] &= \alpha^2 [E^1 + iE^2, E^1 - iE^2] \\ &= \alpha^2 ([E^1, E^1] - i[E^1, E^2] + i[E^2, E^1] + [E^2, E^2]) \\ &= -2i\alpha^2 [E^1, E^2] = \dots \\ &= -2i\alpha^2 iH^1 = 2\alpha^2 H^1 \end{aligned}$$

Then, from equations (2.41) and the definition of E^\pm , we see that $\pm e_1^1 = \pm(t_1 - t_2) = \pm(1 - 0) = \pm 1$. So we therefore set $\alpha^2 = \frac{1}{2} \Rightarrow \alpha = \frac{1}{\sqrt{2}}$, and we find that the appropriate non-Cartan generators (including the 1 to be consistent with the notation in section 2.2.12) are

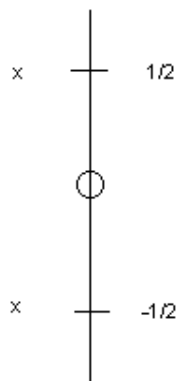
$$E^{\pm 1} = \frac{1}{\sqrt{2}}(E^1 \pm iE^2) \quad (2.45)$$

which is exactly what we had in equation (2.18) above. So, we have derived the trick used to understand quantum mechanical spin in introductory quantum courses!

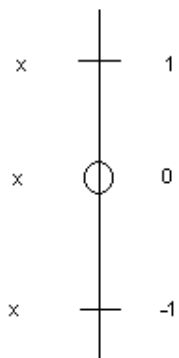
2.2.14 $SU(2)$ for Arbitrary j ... Again

Now that we have our operators in the adjoint representation, we can consider any arbitrary representation. As we saw in section 2.2.11, we can form the linear combinations in equation (2.45) for any $j = \text{integer or half integer}$. The weight vectors will always look like those in the diagram on page 58 (in other words, raising and lowering operators always raise or lower their eigenvalue by 1).

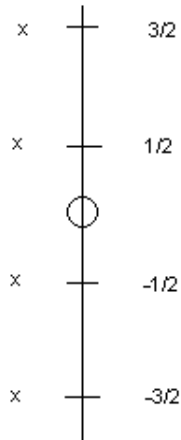
The space of physical states, on the other hand, changes for each representation. For $j = \frac{1}{2}$, we have



For $j = 1$,



For $j = 3/2$,



and so on.

Notice that the vectors graphed in the diagram on page 58 are the exact vectors required to move from point to point in each of these graphs. This is obviously not a coincidence.

2.2.15 $SU(3)$

Now that we have said pretty much everything we can about $SU(2)$, which is only Rank 1 (and therefore not all that interesting), we move on to $SU(3)$. However, we will expedite the process by stating the structure constants up front. The non-zero structure constants are

$$f_{123} = 1, \quad f_{147} = f_{165} = f_{246} = f_{257} = f_{345} = f_{376} = \frac{1}{2}, \quad f_{458} = f_{678} = \frac{\sqrt{3}}{2}$$

The most convenient representation is the **Fundamental Representation** (consisting of

3×3 matrices). They are $T^a = \frac{1}{2}\lambda^a$ for $a = 1, \dots, 8$, where

$$\begin{aligned} \lambda^1 &= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, & \lambda^2 &= \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, & \lambda^3 &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, & \lambda^4 &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \\ \lambda^5 &= \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, & \lambda^6 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, & \lambda^7 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, & \lambda^8 &= \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix} \end{aligned} \quad (2.46)$$

Clearly, only two of these are diagonal, λ^3 and λ^8 . So, $SU(3)$ is a rank 2 group.

Before moving on, we summarize a few results (without proofs). An arbitrary $SU(n)$ group will always have $n^2 - 1$ generators, and will be rank $n - 1$. An arbitrary $SO(n)$ group (for n even) will always have $\frac{n(n-1)}{2}$ generators. We won't worry about the rank of the orthogonal groups.

Working in the adjoint representation of $SU(3)$ would involve 8×8 matrices, which would obviously be very tedious. So, we exploit the fact that the techniques we developed in section 2.2.12 are valid in any representation, and stick with the Fundamental Representation defined by the generators in (2.46).

Proceeding as in section 2.2.13, we note that the eigenvectors will again be

$$v_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

(we will relabel them to be consistent with section 2.2.12 shortly).

Then, the Cartan generators are

$$H^1 = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad H^2 = \frac{1}{2\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}$$

and the non-Cartan Generators are simply

$$E^1 = T^1, \quad E^2 = T^2, \quad E^3 = T^4, \quad E^4 = T^5, \quad E^5 = T^6, \quad E^6 = T^7$$

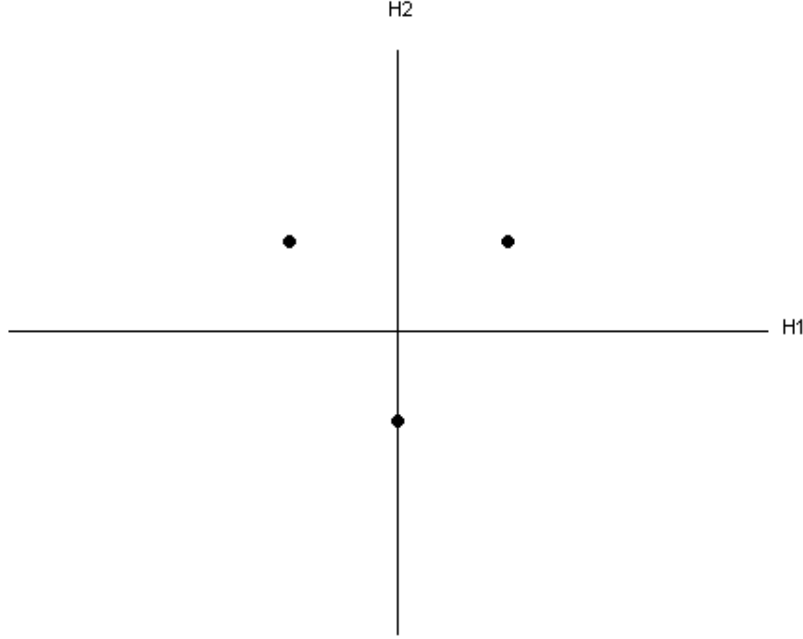
So we have 6 eigenvalues to find,

$$\begin{aligned} H^1 v_1 &= \left(\frac{1}{2}\right) v_1, & H^1 v_2 &= \left(-\frac{1}{2}\right) v_2, & H^1 v_3 &= (0) v_3 \\ H^2 v_1 &= \left(\frac{1}{2\sqrt{3}}\right) v_1, & H^2 v_2 &= \left(\frac{1}{2\sqrt{3}}\right) v_2, & H^2 v_3 &= \left(-\frac{1}{\sqrt{3}}\right) v_3 \end{aligned}$$

So the weight vectors will be 2-dimensional (because the rank is 2). They are

$$\bar{t}^1 = \left(\frac{1}{2} \quad \frac{1}{2\sqrt{3}} \right)^T, \quad \bar{t}^2 = \left(-\frac{1}{2} \quad \frac{1}{2\sqrt{3}} \right)^T, \quad \bar{t}^3 = \left(0 \quad -\frac{1}{\sqrt{3}} \right)^T \quad (2.47)$$

We can graph these in \mathbb{R}^2 as shown below,



Now, repeating nearly the identical argument we started with equation (2.42) and repeating it for all 6 non-Cartan generators, we find that in order to maintain (2.36) and (2.40), we must work with the operators

$$\begin{aligned} \frac{1}{\sqrt{2}}(T^1 \pm iT^2) &= \frac{1}{\sqrt{2}}(E^1 \pm iE^2) \\ \frac{1}{\sqrt{2}}(T^4 \pm iT^5) &= \frac{1}{\sqrt{2}}(E^3 \pm iE^4) \\ \frac{1}{\sqrt{2}}(T^6 \pm iT^7) &= \frac{1}{\sqrt{2}}(E^5 \pm iE^6) \end{aligned} \quad (2.48)$$

The weight vectors associated with these will be, respectively,

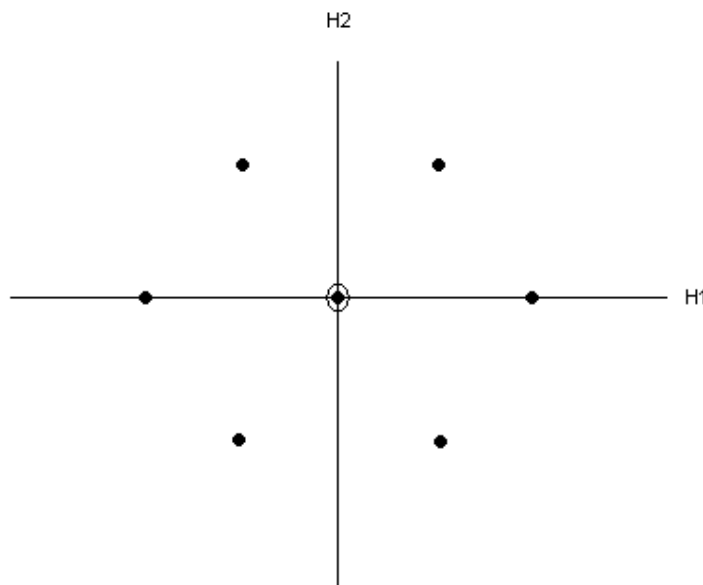
$$\pm(\bar{t}_1 - \bar{t}_2) = \pm \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \pm(\bar{t}_1 - \bar{t}_3) = \pm \begin{pmatrix} 1/2 \\ \sqrt{3}/2 \end{pmatrix}, \quad \text{and} \quad \pm(\bar{t}_2 - \bar{t}_3) = \pm \begin{pmatrix} -1/2 \\ \sqrt{3}/2 \end{pmatrix} \quad (2.49)$$

So, the non-Cartan generators are

$$E^{\pm} \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad E^{\pm} \begin{pmatrix} 1/2 \\ \sqrt{3}/2 \end{pmatrix}, \quad E^{\pm} \begin{pmatrix} -1/2 \\ \sqrt{3}/2 \end{pmatrix}$$

We are no longer in the adjoint representation, so we had to be more deliberate about choosing these linear combinations than we could be in section 2.2.13. What we did here is more general; we chose them to be the differences in the three weight vectors in equation (2.47), *so that* these vectors would naturally transform from one eigenvector to another (just as the raising and lowering operators do, as we found for $SU(2)$ and more generally in section 2.2.12). The remarkable property of Lie groups is that this is always possible in any representation.

We can graph the 6 vectors in (2.49), along with the two Cartan weight vectors, which we know from (2.33) are 0:



And again, just as with $SU(2)$, notice that the 6 non-zero vectors are the exact vectors that would be necessary to move from point to point on the diagram on page 63. So once again, we see that the non-Cartan generators act as raising and lowering operators which transform between the eigenstates of the Cartan generators. Notice that there were 6 non-Cartan generators, and they formed linear combinations to form 6 raising and lowering operators.

2.2.16 What is the Point of All of This?

Before finally getting back to physics, we give a spoiler of how Lie theory is used in physics. What we are going to find is that some physical interaction (electromagnetism, weak force, strong force) will ultimately be described by a Lie group in some particular

representation. The particles that interact with that force will be described by the eigenvectors of the Cartan generators of the group, and the eigenvalues of those eigenvectors will be the physically measurable charges. Clearly, the number of charges associated with the interaction is equal to the number of dimensions of the representation. For example, you likely are aware that the strong force has 3 charges, called “colors” (red, green, and blue). So, the strong force (we will see) will be in a 3-dimensional representation of the group that describes it.

We will find that all forces carrying particles (photons, gluons, W and Z bosons) will be described by the generators of their respective Lie group. The Cartan generators will be force-carrying particles which can interact with any particle charged under that group by transferring energy and momentum, but do not change the charge (photons and Z bosons). This makes sense because Cartan generators are *not* raising or lowering operators. On the other hand, the non-Cartan generators will be force carrying particles which interact with any particle charged under that group by not only transferring energy and momentum, but also changing the charge (W bosons and gluons).

We won’t be able to come back to discussing how this works until much later, and until examples are worked out, this may not be clear. We merely wanted to give an idea of where we are going with this.

2.3 References and Further Reading

The material in section 2.1 came primarily from [9] and [30]. The material in 2.2 came from [9], [10], and [15]. The sections on $SU(2)$ also came from [31].

For further reading, we recommend [2], [8], [16], [17], and [28].

3 Part III — Quantum Field Theory

3.1 A Primer to Quantization

3.1.1 Quantum Fields

Our ultimate goal in the exposition that follows is to formulate a *relativistic quantum mechanical* theory of *interactions*. So, beginning with the fundamental equation of quantum mechanics, Schroedinger's equation,

$$H\Psi = i\hbar \frac{\partial \Psi}{\partial t}$$

we know that for a non-interacting, non-relativistic particle, $H = \frac{\vec{p}^2}{2m} = -\frac{\hbar}{2m} \bar{\nabla}^2$, so

$$-\frac{\hbar}{2m} \bar{\nabla}^2 \Psi = i\hbar \frac{\partial \Psi}{\partial t} \quad (3.1)$$

Of course, Ψ is in this case a **Scalar Field**, and therefore only has one state. So, it describes a spin-0 particle (or, in the language we have learned in the previous sections, it sits in a $j = 0$ representation of $SU(2)$, which is the trivial representation). And, since Ψ does not have any spacetime indices, it also transforms trivially under the Lorentz group $SO(1, 3)$.

Notice, however, that we have a fundamental barrier in making a relativistic theory - the spatial derivative in (3.1) acts quadratically ($\bar{\nabla}^2$), whereas the time derivative is linear. Clearly, treating space and time differently in this way is unacceptable for a relativistic theory. That is a hint of a much more fundamental problem with quantum mechanics; space is always treated as an operator, but time is always treated as a parameter. This fundamental asymmetry is what ultimately prevents a straightforward generalization to relativistic quantum theory.

To fix this problem, we have two choices: either promote time to an operator along with space, or demote space back to a parameter and quantize in a new way.

The first option would result in the Hermitian operators \hat{X} , \hat{Y} , \hat{Z} , and \hat{T} . It turns out that this approach is very difficult and less useful as far as building a relativistic quantum theory. So, we will take the second option.

In demoting position to a parameter along with time, we obviously have sacrificed the operators which we imposed commutation relations on to get a “quantum” theory in the first place. And because we obviously can't impose commutation relations on parameters (because they are scalars), quantization appears impossible. So, we are going to have to make a fairly radical reinterpretation.

Rather than letting the coordinates be Hermitian operators that act on the state in the Hilbert space representing a particle, *we now interpret the particle as the Hermitian operator*, and *this* operator (or particle) will be parameterized by the spacetime coordinates. The physical state that the particle operators act on is then the vacuum itself, $|0\rangle$. So, whereas before you acted on the “electron” $|\Psi\rangle$ with the operator \hat{x} , now the “electron” (parameterized by x) $\Psi(x^\mu)$ acts on the vacuum $|0\rangle$, creating the state $(\Psi(x^\mu)|0\rangle)$. In other words, the operator representing an electron excites the vacuum (empty space) resulting in an electron. We will see that all quantum fields contain appropriate raising and lowering operators to do just this.

This approach, where the quantum mechanical entities are no longer the coordinates acting on the fields, but the fields themselves, is called **Quantum Field Theory (QFT)**.

So, whereas before, quantization was defined by imposing commutation relations on the coordinate operators $[x, p] \neq 0$, we now quantize by imposing commutation relations on the field operators, $[\Psi_1, \Psi_2] \neq 0$.

Because we must still write down the equations of motion which govern the dynamics of the fields, we will need to spend the rest of this section coming up with the classical equations governing the fields we want to work with. We will quantize them in the next section.

3.1.2 Spin-0 Fields

As we said above, Schroedinger’s equation (3.1) describes the time evolution of a spin-0 field, or a scalar field. Generalizing to higher spins will come later. Now, we see how to make this description relativistic.

The most obvious guess for a relativistic form is to simply plug in the standard relativistic Hamiltonian

$$H = \sqrt{\vec{p}^2 c^2 + m^2 c^4} \quad (3.2)$$

Note that, using the standard Taylor expansion $\sqrt{1+x^2} \approx 1 + \frac{1}{2}x^2$ for $x \ll 1$ gives $H \approx mc^2 + \frac{\vec{p}^2}{2m}$, for $\vec{p}^2 \ll c^2$, which is the standard non-relativistic form (plus a constant) we’d expect from a low speed limit.

Plugging (3.2) into (3.1), we have

$$i\hbar \frac{\partial \phi}{\partial t} = \sqrt{-\hbar^2 c^2 \nabla^2 + m^2 c^4} \phi$$

But there are two problems with this:

1. The space and time derivatives are still treated differently, so this is inadequate as a relativistic equation, and

2. Taylor expanding the square root will give an infinite number of derivatives acting on ϕ , making this theory non-local.

One solution is to square the operator on both sides, giving

$$\begin{aligned} -\hbar^2 \frac{\partial^2 \phi}{\partial t^2} &= (-\hbar^2 c^2 \bar{\nabla}^2 + m^2 c^4) \phi \\ \Rightarrow (-\partial^0 \partial_0 + \bar{\nabla}^2 - \frac{m^2 c^2}{\hbar^2}) \phi &= 0 \end{aligned}$$

Or, if we choose the so called “natural units” or “God units”, where $c = \hbar = 1$, we have

$$(\partial^2 - m^2) \phi = 0 \tag{3.3}$$

Equation (3.3) is called the **Klein Gordon** equation. It is nothing more than an operator version of the standard relativistic relation $E^2 = m^2 c^4 + \vec{p}^2 c^2$.

Note that because we will be quantizing fields and not coordinates, there is absolutely nothing “quantum” about the Klein Gordon equation. It is, at this point, merely a relativistic wave equation for a classical, spinless, non-interacting field.

Finally, we note one major problem with the Klein Gordon equation. When we squared the Hamiltonian $H = \sqrt{m^2 c^4 + \vec{p}^2 c^2}$ to get $H^2 = m^2 c^4 + \vec{p}^2 c^2$, the energy eigenvalues became $E = \pm \sqrt{m^2 c^4 + \vec{p}^2 c^2}$. It appears that we have a negative energy eigenvalue! Obviously this is unacceptable in a physically meaningful theory, because negative energy means that we don’t have a true vacuum, and therefore a particle can cascade down forever, giving off an infinite amount of radiation.

We will see that this problem plagues the spin-1/2 particles as well, so we wait to talk about the solution until then.

3.1.3 Why $SU(2)$ for Spin?

Because we are talking about particles “spinning”, a common question is why don’t we use $SO(3)$ instead of $SU(2)$? The original answer to the question is historical. The experiments done in the early days of quantum mechanics were not consistent with the particles having a rotational degree of freedom in spacetime. Rather, the data indicated that, along any given axis, the spin could have only one of two possible values, and $SO(3)$ does not explain this. Here, however, we consider a more mathematical explanation.

First, recall that spin is a purely quantum mechanical phenomenon, with no classical analogue. Because the data demanded two possible spin states, the field describing the particle had to have 2 spin components, $\Psi = \begin{pmatrix} \Psi_1 \\ \Psi_2 \end{pmatrix}$. Now, if we seek a 2-dimensional

representation of $SO(3)$, we find that there is only one: $D_0 \oplus D_0$, the trivial representation consisting of all 1's. This means

$$\begin{pmatrix} \Psi'_1 \\ \Psi'_2 \end{pmatrix} = D_0 \oplus D_0 \begin{pmatrix} \Psi_1 \\ \Psi_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \Psi_1 \\ \Psi_2 \end{pmatrix} = \begin{pmatrix} \Psi_1 \\ \Psi_2 \end{pmatrix}$$

which is no transformation. This is the only 2-dimensional representation of $SO(3)$ that is possible.

The solution to the problem is found in one of the many peculiarities of quantum mechanics. The only physically measurable quantity in quantum theory is the probability amplitude, which is proportional to the square of Ψ . Therefore, the state $\begin{pmatrix} \Psi_1 \\ \Psi_2 \end{pmatrix}$ is physically identical to $\begin{pmatrix} -\Psi_1 \\ -\Psi_2 \end{pmatrix}$.

Now consider a general element of $SO(3)$: $e^{i(\phi J_x + \psi J_y + \theta J_z)}$. On the other hand, a general element of $SU(2)$ will be $e^{i(\phi \frac{\sigma_x}{2} + \psi \frac{\sigma_y}{2} + \theta \frac{\sigma_z}{2})}$.

Now consider rotating the system by an angle of 2π around, say, the z axis. The $SO(3)$ element corresponding to this rotation will be $e^{i2\pi J_z}$, while the $SU(2)$ element will be $e^{i\pi\sigma_z}$. The factor of $1/2$ difference means that the spinor space rotates through only half the angle of the $SO(3)$ does. So, in the 2π rotation, $U \in SU(2) \rightarrow -U$, whereas $R \in SO(3) \rightarrow R$. Therefore, both U and $-U$ correspond to R . There is a 2 to 1 correspondence between $SU(2)$ and $SO(3)$.

And, as we said above, spin is a purely quantum mechanical effect and experimentally only allows 2 values, but $SO(3)$ has no such representation, whereas the $j = 1/2$ representation of $SU(2)$ does. We therefore use $SU(2)$. And, because $SU(2)$ is $2 \rightarrow 1$ with $SO(3)$, but spin is quantum mechanical, both U and $-U$ can consistently correspond to the same $R \in SO(3)$. The minus sign difference is not subject to measurement; only $|\Psi|^2$ is physically measurable.

An important thing to understand is that “spin” is not a rotation through spacetime in any meaningful way. It is a rotation in “spinor space”, which is an *internal* degree of freedom. Like many things in quantum mechanics, spinor space is a mathematical structure. All we can say for certain is what we can measure, or know ($|\Psi|^2$), not what “is”.

3.1.4 Spin $\frac{1}{2}$ Particles

Finding equation (3.3) was easy because scalar fields have no spacetime indices and no spinor indices, and they therefore transform trivially under $SU(2)$ and the Lorentz group.

A particle of spin $1/2$ however, will have two complex components, one for spin $+1/2$,

and the other for spin $-1/2$. So, we describe such a particle as the two-component **Spinor**

$$\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}$$

where ψ_1 and ψ_2 are both $\in \mathbb{C}$. So, we want some differential operator in the form of 2×2 matrices to act on such a field to form the equation of motion.

Following Dirac's approach, he reasoned that given such a 2×2 operator, the equation of motion should somehow "imply" the Klein Gordon equation (which merely makes the theory relativistic). So his goal (and our goal) is to find an equation with a 2×2 matrix differential operator acting on ψ that results in (3.3).

Dirac's approach was to find an operator of the form

$$\mathcal{D} = \gamma^\mu \partial_\mu = \gamma^0 \partial_0 + \gamma^1 \partial_1 + \gamma^2 \partial_2 + \gamma^3 \partial_3$$

where the γ 's are 2×2 matrices, and the equation of motion is then $\mathcal{D}\psi = -im\psi$. The challenge is in finding the appropriate 2×2 γ matrices. Dirac reasoned that, in order to be properly relativistic, operating twice with \mathcal{D} should give the Klein Gordon equation. In other words,

$$\begin{aligned} \mathcal{D}\psi = -im\psi &\Rightarrow \mathcal{D}\mathcal{D}\psi = -im\mathcal{D}\psi \\ &\Rightarrow \gamma^\mu \partial_\mu \gamma^\nu \partial_\nu \psi = -im(-im\psi) \\ &\Rightarrow \gamma^\mu \gamma^\nu \partial_\mu \partial_\nu \psi = -m^2 \psi \\ &\Rightarrow (\gamma^\mu \gamma^\nu \partial_\mu \partial_\nu + m^2)\psi = 0 \end{aligned} \tag{3.4}$$

This will yield the Klein Gordon equation if $\gamma^\mu \gamma^\nu = -\eta^{\mu\nu} \mathbb{I}$. Or, using the symmetry of the sum in (3.4), it will yield the Klein Gordon equation if we demand $\frac{1}{2}(\gamma^\mu \gamma^\nu + \gamma^\nu \gamma^\mu) = -\eta^{\mu\nu} \mathbb{I}$. Consider

$$\{\gamma^\mu, \gamma^\nu\} = \gamma^\mu \gamma^\nu + \gamma^\nu \gamma^\mu = -2\eta^{\mu\nu} \mathbb{I} \tag{3.5}$$

If the γ matrices satisfy (3.5), then (3.4) gives

$$(\gamma^\mu \gamma^\nu \partial_\mu \partial_\nu + m^2)\psi = 0 \Rightarrow (-\eta^{\mu\nu} \partial_\mu \partial_\nu + m^2)\psi = 0 \Rightarrow (\partial^2 - m^2)\psi = 0$$

which is exactly the Klein Gordon equation (3.3).

So, we have the Dirac equation

$$(\mathcal{D} + im)\psi = 0 \tag{3.6}$$

but we still have a problem. Namely, there does not exist a set of 2×2 matrices that solve (3.5). Nor does there exist a set of 3×3 matrices. The smallest possible size where this is possible is 4×4 . Obviously, if we want to describe a spin-1/2 particle with exactly 2 spin

states, using 4 spin components does not seem right. But, we will accept the necessity of 4×4 Dirac matrices and move on.

Instead of using $\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}$, we will define the two 2-dimensional spinors

$$\psi_L \equiv \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} \quad \text{and} \quad \psi_R \equiv \begin{pmatrix} \psi_3 \\ \psi_4 \end{pmatrix} \quad (3.7)$$

and the 4-component spinor

$$\psi \equiv \begin{pmatrix} \psi_L \\ \psi_R \end{pmatrix} \quad (3.8)$$

Now it is possible to solve (3.5). Such a problem is actually very familiar to algebraists, and we will not delve into the details of how this is done. Instead, we merely state one solution (there are many, up to a similarity transformation). We define the 4×4 matrices

$$\gamma^i = \begin{pmatrix} 0 & -\sigma^i \\ \sigma^i & 0 \end{pmatrix} \quad \text{and} \quad \gamma^0 = \begin{pmatrix} 0 & \sigma^0 \\ \sigma^0 & 0 \end{pmatrix} \quad (3.9)$$

where σ^0 is the 2×2 identity matrix, and σ^i are the Pauli spin matrices. It should be no surprise that they show up in attempting to describe spin-1/2 particles. What is interesting is that we did not assume them—we derived them using (3.5).

Before moving on, notice that we have initiated a convention that will be used throughout the rest of these notes. Whenever a greek index is used, it runs over all spacetime indices. Whenever a latin index is used, it runs over only the spatial part. So in (3.9), i runs 1, 2, 3.

Now that we have an explicit form of the Dirac gamma matrices, we can write out (3.6) explicitly:

$$\begin{pmatrix} 0 & 0 & \partial_0 - \partial_3 & -\partial_1 + i\partial_2 \\ 0 & 0 & -\partial_1 - i\partial_2 & \partial_0 + \partial_3 \\ \partial_0 + \partial_3 & \partial_1 - i\partial_2 & 0 & 0 \\ \partial_1 + i\partial_2 & \partial_0 - \partial_3 & 0 & 0 \end{pmatrix} \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{pmatrix} = -im \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{pmatrix}$$

Or, in terms of ψ_L and ψ_R ,

$$\begin{aligned} i\bar{\sigma}^\mu \partial_\mu \psi_R &= +m\psi_L \\ i\sigma^\mu \partial_\mu \psi_L &= +m\psi_R \end{aligned}$$

where we have defined the 4-vectors $\sigma^\mu = (\sigma^0, \sigma^1, \sigma^2, \sigma^3)$ and $\bar{\sigma}^\mu = (\sigma^0, -\sigma^1, -\sigma^2, -\sigma^3)$.

3.1.5 The Lorentz Group

This section is intended to give a deeper understanding of why we were unable to find a 2×2 matrix representation to solve (3.5).

Recall that the driving idea behind the derivation of the Dirac equation (3.6) was to make it imply the Klein Gordon equation, or in other words to be a relativistic theory. Put another way, it was to create a theory that was invariant under the Lorentz group $SO(1, 3)$. So, let's take a closer look at the Lorentz group.

We know from section 1.1.5 that the Lorentz group consists of 3 rotations and 3 boosts. We gave the general forms of these transformations in equations (1.5) and (1.6). It is easy, using those general expressions in addition to (2.7), to find all 6 generators, and then multiply them out to get the commutation relations. We spare the (easy but tedious) details and simply state the commutation relations. If we label the generators of rotation J^i ($i = 1, 2, 3$), and the generators of boosts K^i ($i = 1, 2, 3$), then the commutation relations are

$$\begin{aligned} [J^i, J^j] &= i\epsilon^{ijk} J^k \\ [J^i, K^j] &= i\epsilon^{ijk} K^k \\ [K^i, K^j] &= -i\epsilon^{ijk} J^k \end{aligned}$$

In order to make the actual structure of this group more obvious, we define two new linear combinations of these generators:

$$N^i = \frac{1}{2}(J^i - iK^i) \quad N^{i\dagger} = \frac{1}{2}(J^i + iK^i)$$

Writing out the commutation relations for N^i and $N^{i\dagger}$, we get

$$\begin{aligned} [N^i, N^j] &= i\epsilon^{ijk} N^k \\ [N^{i\dagger}, N^{j\dagger}] &= i\epsilon^{ijk} N^{k\dagger} \\ [N^i, N^{j\dagger}] &= 0 \end{aligned}$$

So, both N^i and $N^{i\dagger}$ separately form an $SU(2)$. In more mathematical terms, we say that $SO(1, 3)$ is **isomorphic** to $SU(2) \otimes SU(2)$, which we denote $SO(1, 3) \cong SU(2) \otimes SU(2)$. While the idea of an isomorphism is a very rich mathematical idea, for now you can simply think of it as a way of saying that two groups have the same group structure.

So, because a given representation of $SU(2)$ is defined by the value of j , we can see that a particular representation of the Lorentz group $SO(1, 3) \cong SU(2) \otimes SU(2)$ is defined by *two* values of j , or by the doublet (j, j') . The smallest possible representation then is $(j, j') = (0, 0)$. This has one state from $j = 0$ and one state from $j' = 0$, and therefore has $1 \times 1 = 1$ state total. Therefore, this representation describes a scalar field.

Then, there is the state $(0, 1/2)$, which will have one state from $j = 0$, but two states from $j = 1/2$, for a total of $1 \times 2 = 2$ states. Therefore, this describes a single spin-1/2 field. We call this field ψ_L , and the $(0, 1/2)$ representation the **Left-Handed Spinor Representation** of the Lorentz group.

Clearly, we will also have the representation $(1/2, 0)$, which also has 2 states, corresponding to the ψ_R field. This is called the **Right-Handed Spinor Representation** of the Lorentz group. This is the reason for the notation used in (3.7). The left-handed $(0, 1/2)$ representation acts on ψ_L and the right-handed $(1/2, 0)$ representation acts on ψ_R .

Next is the representation $(1/2, 1/2)$, which has two states from $j = 1/2$ and two from the $j' = 1/2$ for a total of $2 \times 2 = 4$ states. It turns out that this representation is the spacetime vector representation we use to act on spacetime vectors for the standard Lorentz transformations discussed in section 1.1.5.

Now, an $SU(2)$ representation specified by some j is an irreducible representation, and therefore the tensor products $SU(2) \otimes SU(2)$ specified by the doublet (j, j') are irreducible. This means that there are no irreducible subspaces, and so given a representation (j, j') , there is a particular transformation taking the state (j, j') to (j', j) . For the $(0, 0)$ and $(1/2, 1/2)$ representations this doesn't affect anything. However, this fact means that the $(0, 1/2)$ and $(1/2, 0)$ representations must always appear together. To put this in more mathematical language, our choices for representations of the Lorentz group are

$$(0, 0), \quad (1/2, 1/2) \quad \text{and} \quad (1/2, 0) \oplus (0, 1/2)$$

which are 1, 4, and 4-dimensional representations, respectively. Furthermore, they are the representations which transform Klein Gordon scalar/spinor-0 fields, spacetime 4-vectors, and spin-1/2 spinors, respectively.

The physical meaning of this fact is that relativity demands that if you want a theory with spin-1/2 particles, you cannot have them existing by themselves. They must come in pairs, each transforming under an $SU(2)$ representation of opposite handedness. In the next two sections we will discuss ways of interpreting this fact, starting with Dirac's original approach which, while brilliant, didn't ultimately work. Then we will consider what appears to be the correct view.

3.1.6 The Dirac Sea Interpretation of Antiparticles

Initially, it may seem that the impossibility of finding a 2×2 matrix solution to (3.5) means that we can't have fields with 2 spinor states. However, we saw in the last section that we aren't limited to scalars and spacetime 4-component spinors. We can also have two fields, ψ_L and ψ_R , which can be paired together to form two spin-1/2 fields in a 4-component spinor $\psi = \begin{pmatrix} \psi_L \\ \psi_R \end{pmatrix}$. So, Dirac was faced with the challenge of both interpreting this, while at the same time dealing with the negative energy states mentioned in section 3.1.2.

Dirac's solution, though today abandoned, was brilliant enough to mention. He suggested that because spin-1/2 particles obey the **Pauli Exclusion Principle**, there could be

an infinite number of particles *already* in the negative energy levels, and so they are already occupied, preventing any more particles from falling down and giving off infinite energy. Thus, the negative energy problem was solved.

Furthermore, he said that it is possible for one of the particles in this infinite negative sea to be excited and jump up into a positive energy state, leaving behind a hole. This would appear to us, experimentally, as a particle with the same mass, but the opposite charge. He called such particles **Antiparticles**. For example, the antiparticle of the electron is the antielectron, or the positron (same mass, opposite charge). The positron is not a particle in the same sense as the electron, but rather is a hole in an infinite sea of electrons. And where this negative charge is missing, all that is left is a hole which appears as a positively charged particle.

So, ψ_L describes a particle, and due to the infinite sea of negative particles, there can always be a hole, which will be described by ψ_L . Everything about this worked out mathematically, and when antiparticles were detected about 5 years after Dirac's prediction of them, it appeared that Dirac's suggestion was correct.

However, there were two major problems with Dirac's idea, and they ultimately proved fatal to the "Dirac Sea" interpretation:

1. This theory, which was supposed to be a theory of single particles, now requires an infinite number of them.
2. Particles like photons, pions, mesons, or Klein-Gordon scalars don't obey the Pauli Exclusion Principle, but still have negative energy states, and therefore Dirac's argument doesn't work.

However, his labeling them "antiparticles" has stuck, and we therefore still refer to the right-handed part of the spin-1/2 field as the antiparticle, whereas the left-handed part is still the particle.

For these reasons, we must have some other way of understanding the existence of the antiparticles.

3.1.7 The QFT Interpretation of Antiparticles

In presenting the problem of negative energy states, we have been somewhat intentionally sloppy. To take stock, we have two equations of motion: the Klein Gordon (3.3) for scalar/spin-0 fields, and the Dirac equation (3.6) for spin-1/2 particles.

And in our discussion of negative energy states, we were "pretending" that the ψ 's and ϕ 's are "states" with negative energy. But, as we said in section 3.1.1, QFT offers a different interpretation of the fields. Namely, the fields are not states — they are operators. And

consequently they can't have energy. A state is *made* by acting on the vacuum with either of the operators ϕ or ψ , and then the state $\phi|0\rangle$ or $\psi|0\rangle$ has some energy.

So, QFT allows us to see the antiparticle as a real, actual particle, rather than the absence of a particle. And, we do not need the conceptually difficult idea of an infinite sea of negative energy particles. The vacuum $|0\rangle$, with no particles in it, is now our state with the lowest possible energy level. And, as we will see, there are never negative energy states with these particles.

How exactly $|0\rangle$ works will become clearer when we quantize. The point to be understood for now is that QFT solves the problem of negative energy by reinterpreting what is a state and what is an operator. The fields ϕ and ψ are operators, not states, and therefore they do not have energy associated with them (any more than the operator \hat{x} or \hat{p}_x did in non-relativistic quantum mechanics). So, without any problems of negative energy, we merely accept that nature, due to relativity, demands that particles come in particle/antiparticle pairs, and we move on.

3.1.8 Lagrangians for Scalars and Dirac Particles

Now that we have the equations of motion (3.3) and (3.6), we want to know the actions that lead to these equations of motion. In order to save time, we will merely write down the answers and let you take the variations to see that they do indeed lead to the Klein Gordon and Dirac equations of motion for ϕ and ψ .

They are

$$\mathcal{L}_{KG} = -\frac{1}{2}\partial^\mu\phi\partial_\mu\phi - \frac{1}{2}m^2\phi \quad (3.10)$$

$$\mathcal{L}_D = i\psi_L^\dagger\bar{\sigma}^\mu\partial_\mu\psi_L + i\psi_R^\dagger\sigma^\mu\partial_\mu\psi_R - m(\psi_L^\dagger\psi_R + \psi_R^\dagger\psi_L) \quad (3.11)$$

where the dagger represents the Hermitian conjugate, $\psi_L^\dagger = (\psi_1^*, \psi_2^*)$, $\psi_R^\dagger = (\psi_3^*, \psi_4^*)$, as usual. You can actually take the variation of \mathcal{L}_D with respect to either ψ_L^\dagger and ψ_R^\dagger to get the equations of motion for ψ_L and ψ_R , or you can take the variations with respect to ψ_L and ψ_R to get the equations for ψ_L^\dagger and ψ_R^\dagger . The two sets of equations are simply the conjugates of each other, and therefore represent a single set of equations.

In order to simplify (3.11), the convention is to use the Dirac gamma matrices (3.9) to define $\bar{\psi} \equiv \psi^\dagger\gamma^0$ (where ψ here is the 4-component spinor in equation (3.8)). Using this, all 4 terms in (3.11) can be summarized as

$$\mathcal{L}_D = \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi \quad (3.12)$$

3.1.9 Conserved Currents

In Part I we discussed how symmetries and conserved quantities are related. Let's consider a few examples of this using the Lagrangians we have now defined.

Consider a massless Klein Gordon scalar particle, described by $\mathcal{L} = -\frac{1}{2}\partial^\mu\phi\partial_\mu\phi$. Following what we did starting with equation (1.2), consider the transformation $\phi \rightarrow \phi + \epsilon$, where ϵ is a constant. Because $\partial^\mu\phi \rightarrow \partial^\mu\phi + \partial^\mu\epsilon = \partial^\mu\phi$, the Lagrangian is invariant. So (using $\delta\phi = 1$), our conserved quantity is

$$j^\mu = \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\delta\phi = -\partial^\mu\phi$$

Or, consider the Klein Gordon Lagrangian with complex scalar fields ϕ and ϕ^\dagger , which we write as $\mathcal{L} = -\partial^\mu\phi^\dagger\partial_\mu\phi - m^2\phi^\dagger\phi$. We can make the transformation $\phi \rightarrow e^{i\alpha}\phi$ and $\phi^\dagger \rightarrow \phi^\dagger e^{-i\alpha}$ (where α is an arbitrary real constant). This type of transformation is called a $U(1)$ transformation, because $e^{i\alpha}$ is an element of the group of all 1×1 unitary matrices, as discussed in section 2.2.1.

The conserved quantity associated with this $U(1)$ symmetry is

$$j^\mu = \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\delta\phi + \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi^\dagger)}\delta\phi^\dagger = i(\phi\partial^\mu\phi^\dagger - \phi^\dagger\partial^\mu\phi)$$

Or consider the Dirac Lagrangian. Notice that it is invariant under the $U(1)$ transformation $\psi \rightarrow e^{i\alpha}\psi$, with current

$$j^\mu = \bar{\psi}\gamma^\mu\psi \quad (3.13)$$

In both of the previous examples, notice that the $U(1)$ symmetry changes the field at *all* points in space at once, and all in the same way. In other words, it is a single overall constant phase $e^{i\alpha}$. We therefore call such a symmetry a **Global Symmetry**. The implications of this are likely not clear at this point. We merely wish to call your attention to the fact that $e^{i\alpha}$ has no spacetime dependence.

3.1.10 The Dirac Equation with an Electromagnetic Field

Previously we found the Lagrangian for an electromagnetic field (1.13). Our goal now is to find a Lagrangian that describes the electromagnetic field and a spin-1/2 particle that couples to the electromagnetic field, and additionally the interaction between them. We start by writing down a Lagrangian without any interaction. This will simply be the sum of the two terms,

$$\mathcal{L} = \mathcal{L}_D + \mathcal{L}_{EM} = \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - J^\mu A_{\mu\nu} \quad (3.14)$$

But, because the Dirac part has no terms in common with the electromagnetic part, the equations of motion and the conserved quantities for both ψ and A^μ will be exactly the same, as if the other weren't present at all. In other words, both fields go about their way as if the other weren't there—there is no interaction in this theory. Because this makes for a boring universe (and horrible phenomenology), we need to find some way of coupling the two fields together to produce some sort of interaction.

Interaction is added to a physical theory by adding another term to the Lagrangian called the **Interaction Term**. So, the final Lagrangian will have the form $\mathcal{L} = \mathcal{L}_D + \mathcal{L}_{EM} + \mathcal{L}_{int}$.

Now, for reasons that will become clear in the next section (and even more clear when we quantize), we do this by coupling the electromagnetic field A^μ to the current resulting from the $U(1)$ symmetry in \mathcal{L}_D , which we discussed in section 3.1.9, and wrote out in equation (3.13). In other words, our interaction term will be proportional to $A^\mu j_\mu$.

So, adding a constant of proportionality q (which we will see has the physical interpretation of a coupling constant, weighting the probability of an interaction to take place, or equivalently the physical interpretation of electric charge), our Lagrangian is now

$$\begin{aligned}\mathcal{L} &= \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - J^\mu A_\mu - qj^\mu A_\mu \\ &= \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - (J^\mu + q\bar{\psi}\gamma^\mu\psi)A_\mu\end{aligned}\tag{3.15}$$

Notice that \mathcal{L} is still invariant under the global $U(1)$ symmetry, and the $U(1)$ current is still $J^\mu = \bar{\psi}\gamma^\mu\psi$.

Also, notice that the Lagrangians in (3.14) and (3.15) are the same except for a shift in the current term, $J^\mu \rightarrow J^\mu + qj^\mu$. Recall that physically, $J^\mu = (\rho, \vec{J})$ represents the charge and current creating the field. The fact that J^μ has shifted in (3.15) simply means that the spin-1/2 particle in this theory contributes to the field, which is exactly what we would expect it to do.

If we set $q = e$, the electric charge, this Lagrangian becomes upon quantization the Lagrangian of Quantum Electrodynamics (QED), which to date makes the most accurate experimental predictions ever.

In the next section, we will re-derive this Lagrangian in a more fundamental way.

3.1.11 Gauging the Symmetry

Physically speaking, this section is among the most important in these notes. Read this section again and again until you understand every step.

Consider once again the Dirac Lagrangian (3.6). As we said in section 3.1.9, it is invariant under the *global* $U(1)$ transformation $\psi \rightarrow e^{i\alpha}\psi$. It is global in that it acts on the field the

exact same way at every point in spacetime. The idea behind this section is that we are going to make this symmetry **Local**, so that α depends on spacetime ($\alpha = \alpha(x^\mu)$), and then try to force the Lagrangian to maintain its invariance under the *local* $U(1)$ transformation. Making a global symmetry local is referred to as **Gauging** the symmetry.

We start by making the local $U(1)$ transformation:

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi \rightarrow \bar{\psi}e^{-i\alpha(x)}(i\gamma^\mu\partial_\mu - m)e^{i\alpha(x)}\psi$$

and because the differential operators will now act on $\alpha(x)$ as well as ψ , we get extra terms:

$$\begin{aligned}\mathcal{L} \rightarrow \bar{\psi}e^{-i\alpha(x)}(i\gamma^\mu\partial_\mu - m)e^{i\alpha(x)}\psi &= \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi - \bar{\psi}\gamma^\mu\psi\partial_\mu\alpha(x) \\ &= \bar{\psi}(i\gamma^\mu\partial_\mu - m - \gamma^\mu\partial_\mu\alpha(x))\psi\end{aligned}$$

If we want to demand that \mathcal{L} still be invariant under this local $U(1)$ transformation, we must find a way of canceling the $\bar{\psi}\gamma^\mu\psi\partial_\mu\alpha(x)$ term. We do this in the following way.

Define some arbitrary field A_μ which under the $U(1)$ transformation $e^{i\alpha(x)}$ transforms according to

$$A_\mu \rightarrow A_\mu - \frac{1}{q}\partial_\mu\alpha(x) \quad (3.16)$$

We call A_μ the **Gauge Field** for reasons that will be clear soon, and q is a constant we have included for later convenience.

We introduce A_μ by replacing the standard derivative ∂_μ with the **Covariant Derivative**

$$D_\mu \equiv \partial_\mu + iqA_\mu \quad (3.17)$$

If you have studied general relativity or differential geometry at any point, you are familiar with covariant derivatives. There is an incredibly rich geometric picture of all of this, but it is beyond the scope of these notes. We will deal with it later in this series, however.

As a comment regarding vocabulary, to say that a particle “carries charge” mathematically means that it has the corresponding term in its covariant derivative. So, if a particle’s covariant derivative is equal to the normal differential operator ∂^μ , then the particle has no charge, and it will not interact with anything. But if it carries charge, it will have a term corresponding to that charge in its covariant derivative. This will become clearer as we proceed.

So, our Lagrangian is now

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi = \bar{\psi}(i\gamma^\mu[\partial_\mu + iqA_\mu] - m)\psi = \bar{\psi}(i\gamma^\mu\partial_\mu - m - q\gamma^\mu A_\mu)\psi$$

And under the local $U(1)$ we have

$$\begin{aligned}
\mathcal{L} &\rightarrow \bar{\psi} e^{-i\alpha(x)} (i\gamma^\mu \partial_\mu - m - q\gamma^\mu [A_\mu - \frac{1}{q} \partial_\mu \alpha(x)]) e^{i\alpha(x)} \psi \\
&= \bar{\psi} (i\gamma^\mu \partial_\mu - m - \gamma^\mu \partial_\mu \alpha(x) - q\gamma^\mu A_\mu + \gamma^\mu \partial_\mu \alpha(x)) \psi \\
&= \bar{\psi} (i\gamma^\mu \partial_\mu - m - q\gamma^\mu A_\mu) \psi = \bar{\psi} (i\gamma^\mu D_\mu - m) \psi \\
&= \mathcal{L}
\end{aligned}$$

So, the addition of the field A_μ has indeed restored the $U(1)$ symmetry. Notice that now it is not only invariant under this local $U(1)$, but also still under the global $U(1)$ we started with, with the same conserved $U(1)$ current $j^\mu = \bar{\psi} \gamma^\mu \psi$. This allows us to rewrite the Lagrangian as

$$\mathcal{L} = \bar{\psi} (i\gamma^\mu D_\mu - m) \psi = \bar{\psi} (i\gamma^\mu \partial_\mu - m) \psi - q j^\mu A_\mu \quad (3.18)$$

But we have a problem. If we want to know what the dynamics of A_μ will be, we naturally take the variation of the Lagrangian with respect to A_μ . But because there are no derivatives of A_μ , the Euler-Lagrange equation is merely $\frac{\partial \mathcal{L}}{\partial A_\mu} = -q \bar{\psi} \gamma^\mu \psi = 0$. But $-q \bar{\psi} \gamma^\mu \psi = -q j^\mu$. So the equation of motion for A_μ says that the current vanishes, or that $j^\mu = 0$, and so the Lagrangian is reduced back to (3.12), which was not invariant under the local $U(1)$.

We can state this problem in another way. All physical fields have some sort of dynamics. If they don't then they are merely a constant background field that never changes and does nothing. As it is written, equation (3.18) has a field A_μ but A_μ has no kinetic term, and therefore no dynamics.

So, to fix this problem we must include some sort of dynamics, or kinetic terms, for A_μ .

The way to do this turns out to involve a considerable amount of geometry which would be out of place in these notes. We will cover the necessary ideas in a later paper in this series and derive the following expressions. For now we merely give the results and ask you for patience until we have the machinery to derive them.

For an arbitrary field A_μ , the appropriate gauge-invariant kinetic term is

$$\mathcal{L}_{Kin,A} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu}$$

where

$$F^{\mu\nu} \equiv \frac{i}{q} [D^\mu, D^\nu] \quad (3.19)$$

and q is the constant of proportionality introduced in the transformation of A_μ in equation (3.16). D^μ is the covariant derivative defined in (3.17).

Writing out (3.19) (and using an arbitrary test function $f(x)$),

$$\begin{aligned}
F^{\mu\nu} f(x) &= \frac{i}{q} [D^\mu, D^\nu] f(x) \\
&= \frac{i}{q} [(\partial^\mu + iqA^\mu)(\partial^\nu + iqA^\nu) - (\partial^\nu + iqA^\nu)(\partial^\mu + iqA^\mu)] f(x) \\
&= \frac{i}{q} [\partial^\mu \partial^\nu f(x) + iq\partial^\mu (A^\nu f(x)) + iqA^\mu \partial^\nu f(x) - q^2 A^\mu A^\nu f(x) \\
&\quad - \partial^\nu \partial^\mu f(x) + iq\partial^\nu (A^\mu f(x)) + iqA^\nu \partial^\mu f(x) - q^2 A^\nu A^\mu f(x)] \\
&= \frac{i}{q} [iqf(x)\partial^\mu A^\nu + iqA^\nu \partial^\mu f(x) + iqA^\mu \partial^\nu f(x) - q^2 A^\mu A^\nu f(x) \\
&\quad - iqf(x)\partial^\nu A^\mu - iqA^\mu \partial^\nu f(x) - iqA^\nu \partial^\mu f(x) + q^2 A^\nu A^\mu f(x)] \\
&= [\partial^\mu A^\nu - \partial^\nu A^\mu + iq[A^\mu, A^\nu]] f(x)
\end{aligned}$$

But for each value of μ , A^μ is a scalar function, so the commutator term vanishes, leaving (dropping the test function $f(x)$)

$$F^{\mu\nu} = \frac{i}{q} [D^\mu, D^\nu] = \partial^\mu A^\nu - \partial^\nu A^\mu \quad (3.20)$$

So, writing out the entire Lagrangian we have

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4} F_{\mu\nu} F^{\mu\nu}$$

And finally, because A^μ is obviously a physical field, we can naturally assume that there is some source term causing it, which we simply call J^μ . This makes our final Lagrangian

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} - J^\mu A_\mu$$

Comparing this to (3.15) we see that they are exactly the same. So what have we done? We started with nothing but a Lagrangian for a spin-1/2 particle, which had a global $U(1)$ symmetry. Then, all we did was promote the $U(1)$ symmetry to a local symmetry (we gauged the symmetry), and then imposed what we had to impose to get a consistent theory. The gauge field A_μ was forced upon us, and the form of the kinetic term for A_μ is demanded automatically by geometric considerations we did not delve into.

In other words, we started with nothing but a non-interacting particle, and by specifying *nothing* but $U(1)$ we have created a theory with not only that same particle, but also electromagnetism. The A_μ field, which upon quantization will be the photon, is a direct consequence of the $U(1)$.

This is what we meant at the end of section 2.2.11 when we said that electromagnetism is described by $U(1)$. We will talk more about the weak and strong forces later, as well as the groups that give rise to them.

Theories of this type, where we generate forces by specifying a Lie group, are called **Gauge Theories**, or **Yang-Mills Theories**.

Finally, notice that (3.16) has exactly the same form as (1.15). This is why we call A_μ a gauge field. The gauge symmetry in electromagnetism is a sort of remnant of the much deeper and more fundamental $U(1)$ structure of the theory.

3.2 Quantization

3.2.1 Review of What Quantization Means

In quantum mechanics (not QFT), quantization is done by taking certain dynamical quantities and making use of the **Heisenberg Uncertainty Principle**. Normally we take position \bar{x} and momentum \bar{p} and, according to Heisenberg, the measurement of the particle's position will effect its momentum and vice-versa.

To make this more precise, we promote x and p from merely being variables to being Hermitian operators \hat{x} and \hat{p} (which can be represented by matrices) acting on some vector space. Calling a vector in this space $|\psi\rangle$, physically measurable quantities (like position or momentum) become the eigenvalues of the operators \hat{x} and \hat{p} ,

$$\begin{aligned}\hat{x}|\psi\rangle &= x|\psi\rangle \\ \hat{p}|\psi\rangle &= p|\psi\rangle\end{aligned}$$

Heisenberg Uncertainty says that measuring x will affect the value of p , and vice-versa. It is the act of measuring which enacts this effect. It is not an engineering problem in the sense that there is no better measurement technique which would undo this. It is a fundamental fact of quantum mechanics (and therefore the universe) that measurement of one variable affects another.

So, if we measure x (using \hat{x}) and then p (using \hat{p}), we will in general get different values for both than if we measured p and then x . More mathematically, $\hat{x}\hat{p} \neq \hat{p}\hat{x}$. Put another way,

$$[\hat{x}, \hat{p}] \equiv \hat{x}\hat{p} - \hat{p}\hat{x} \neq 0$$

For reasons learned in an introductory quantum course, the actual relation is

$$[\hat{x}, \hat{p}] = i\hbar \tag{3.21}$$

where \hbar is Planck's Constant. We call (3.21) the **Canonical Commutation Relation**, and it is this structure which allows us to determine the physical structure of the theory.

More generally, we choose some set of operators that all commute with each other, and then label a physical state by its eigenvectors. For example \hat{x} , \hat{y} and \hat{z} all commute with

each other, so we may label a physical state by its eigenvectors $|\psi_r\rangle = |x, y, z\rangle$. Or, because \hat{p}_x , \hat{p}_y , and \hat{p}_z all commute, we may call the state $|\psi_p\rangle = |p_x, p_y, p_z\rangle$. We may also include some other values like spin and angular momentum, to have (for example) $|\psi\rangle = |x, y, z, s_z, L_z, \dots\rangle$.

As discussed in section 3.1.1, when we make the jump to QFT, the fields are no longer the states but the operators. We are therefore going to impose commutation relations on the fields, not on the coordinates.

Furthermore, whereas before the states were eigenvectors of the coordinate operators, we now will expand the fields in terms of the eigenvectors of the Hamiltonian.

3.2.2 Canonical Quantization of Scalar Fields

We begin with the Klein Gordon Lagrangian in equation (3.10), but we make the slight modification of adding an arbitrary constant Ω ,

$$\mathcal{L}_{KG} = -\frac{1}{2}\partial^\mu\phi\partial_\mu\phi - \frac{1}{2}m^2\phi^2 + \Omega$$

Note that Ω has absolutely no affect whatsoever on the physics.

Quantization then comes about by defining the field momentum and Hamiltonian (using (1.7) and (1.8)) to get

$$\Pi = \frac{\partial\mathcal{L}}{\partial\dot{\phi}(x)} = \dot{\phi} \quad (3.22)$$

$$\mathcal{H} = \Pi\dot{\phi} - \mathcal{L} = \frac{1}{2}\Pi^2 + \frac{1}{2}(\bar{\nabla}\phi)^2 + \frac{1}{2}m^2\phi^2 - \Omega \quad (3.23)$$

Now, using the canonical commutation relations (3.21) as guides, we impose

$$\begin{aligned} [\phi(t, \bar{x}), \phi(t', \bar{x}')] &= 0 \\ [\Pi(t, \bar{x}), \Pi(t', \bar{x}')] &= 0 \\ [\phi(t, \bar{x}), \Pi(t', \bar{x}')] &= i\delta(t - t')\delta(\bar{x} - \bar{x}') \end{aligned} \quad (3.24)$$

(where we have set $\hbar = 1$).

We can see more clearly what this means if we expand the solutions of the Klein Gordon equation. One solution is plane waves, $e^{i\bar{k}\cdot\bar{x} \pm i\omega t}$, where

$$\omega = +\sqrt{\bar{k}^2 + m^2} \quad (3.25)$$

and \bar{k} is the standard wave vector.

So, we write the field ϕ as

$$\phi(t, \bar{x}) = \int \frac{d^3 \bar{k}}{f(\bar{k})} [a(\bar{k})e^{i\bar{k} \cdot \bar{x} - i\omega t} + b(\bar{k})e^{i\bar{k} \cdot x + i\omega t}]$$

where $f(x)$ is a redundant function which we have included for later convenience. For now, both $a(\bar{k})$ and $b(\bar{k})$ are merely arbitrary coefficients (integration constants) used to expand $\phi(t, \bar{x})$ in terms of individual solutions.

We demand that $\phi(t, \bar{x})$ be Hermitian. This requires

$$\phi^\dagger = \phi \Rightarrow \phi^* = \phi \Rightarrow b^*(\bar{k}) = a(-\bar{k})$$

Then, changing the sign of the integration variable \bar{k} on the second term in the integral allows us to use 4-vector notation, so

$$\phi(x) = \int \frac{d^3 \bar{k}}{f(\bar{k})} [a(\bar{k})e^{ik \cdot x} + a^*(\bar{k})e^{-ik \cdot x}]$$

where $k \cdot x = k^\mu x_\mu$.

Now notice that the integration measure, $d^3 \bar{k}$, is not invariant under Lorentz transformations (because it integrates over the spatial part but not over the time part). We therefore choose $f(\bar{k})$ to restore Lorentz invariance.

We know that the measure $d^3 k$ would be invariant, as would δ functions and Θ (step) functions. So, consider the invariant combination

$$d^4 k \delta(k^2 + m^2) \Theta(k^0) \quad (3.26)$$

The δ function merely requires that relativity hold ($k^2 + m^2$ is simply the relativistic relation (3.2)), and the Θ function preserves causality. So this is a physically acceptable Lorentz invariant integration measure.

Recall the general δ function identity,

$$\int_{-\infty}^{\infty} dx \delta(g(x)) = \sum_i \frac{1}{\left| \frac{dg(x)}{dx} \right|_{x=x_i}}$$

where the x_i 's are the zeros of the function $g(x)$. We can do the k^0 integral over measure (3.26), and using the fact that the zeros of $k^2 + m^2 = \bar{k}^2 - k^0 k_0 + m^2$ in terms of k^0 are $k^0 k_0 = \bar{k}^2 + m^2 = \omega^2$, we get

$$\int d^3 \bar{k} dk^0 \delta(k^2 + m^2) \Theta(k^0) = \int \frac{d^3 \bar{k}}{2\omega}$$

So, adding a factor of $(2\pi)^3$ for later convenience, we take our invariant measure to be

$$\frac{d^3 \bar{k}}{(2\pi)^3 2\omega}$$

So finally,

$$\phi(x) = \int \widetilde{dk} [a(\bar{k})e^{ik \cdot x} + a^\star(\bar{k})e^{-ik \cdot x}] \quad (3.27)$$

where we have defined $\widetilde{dk} \equiv \frac{d^3\bar{k}}{(2\pi)^3 2\omega}$.

The commutation relations we defined in (3.24) will now hold provided we impose

$$\begin{aligned} [a(\bar{k}), a(\bar{k}')] &= 0 \\ [a^\dagger(\bar{k}), a^\dagger(\bar{k}')] &= 0 \\ [a(\bar{k}), a^\dagger(\bar{k}')] &= (2\pi)^3 2\omega \delta^3(\bar{k} - \bar{k}') \end{aligned} \quad (3.28)$$

(showing this is fairly tedious, but we encourage you to work it out). We are using \dagger instead of \star to emphasize that, in the quantum theory, we are talking about Hermitian operators. The operators $a(\bar{k})$ and $a^\dagger(\bar{k})$ are scalars, so in this case $a^\star = a^\dagger$.

Furthermore, we can write the Hamiltonian H in terms of (3.27):

$$\begin{aligned} H &= \int d^3x \mathcal{H} = \int d^3x \left(\frac{1}{2} \Pi^2 + \frac{1}{2} (\nabla \phi)^2 + \frac{1}{2} m^2 \phi^2 - \Omega \right) \\ &= \frac{1}{2} \int \widetilde{dk} \widetilde{dk'} d^3x [(-i\omega a(\bar{k})e^{ik \cdot x} + i\omega a^\star(\bar{k})e^{-ik \cdot x})(-i\omega' a(\bar{k}')e^{ik' \cdot x} + i\omega' a^\star(\bar{k}')e^{-ik' \cdot x}) \\ &\quad + (i\bar{k} a(\bar{k})e^{ik \cdot x} - i\bar{k} a^\star(\bar{k})e^{-ik \cdot x}) \cdot (i\bar{k}' a(\bar{k}')e^{ik' \cdot x} - i\bar{k}' a^\star(\bar{k}')e^{-ik' \cdot x}) \\ &\quad m^2(a(\bar{k})e^{ik \cdot x} + a^\star(\bar{k})e^{-ik \cdot x})(a(\bar{k}')e^{ik' \cdot x} + a^\star(\bar{k}')e^{-ik' \cdot x})] - \int d^3x \Omega \\ &= \frac{1}{2} \int \widetilde{dk} \widetilde{dk'} d^3x [(-\omega\omega' a(\bar{k})a(\bar{k}')e^{i(k+k') \cdot x} + \omega\omega' a(\bar{k})a^\star(\bar{k}')e^{i(k-k') \cdot x} \\ &\quad + \omega\omega' a^\star(\bar{k})a(\bar{k}')e^{-i(k-k') \cdot x} - \omega\omega' a^\star(\bar{k})a^\star(\bar{k}')e^{-i(k+k') \cdot x}) \\ &\quad + (-\bar{k} \cdot \bar{k}' a(\bar{k})a(\bar{k}')e^{i(k+k') \cdot x} + \bar{k} \cdot \bar{k}' a(\bar{k})a^\star(\bar{k}')e^{i(k-k') \cdot x} \\ &\quad + \bar{k} \cdot \bar{k}' a^\star(\bar{k})a(\bar{k}')e^{-i(k-k') \cdot x} - \bar{k} \cdot \bar{k}' a^\star(\bar{k})a^\star(\bar{k}')e^{-i(k+k') \cdot x}) \\ &\quad + m^2(a(\bar{k})a(\bar{k}')e^{i(k+k') \cdot x} + a(\bar{k})a^\star(\bar{k}')e^{i(k-k') \cdot x} \\ &\quad + a^\star(\bar{k})a(\bar{k}')e^{-i(k-k') \cdot x} + a^\star(\bar{k})a^\star(\bar{k}')e^{-i(k+k') \cdot x}) \\ &\quad - V\Omega] \end{aligned}$$

where V is the volume of the space resulting from the $\int d^3x$ integral. Then, from the fact that $\int d^3x e^{i\bar{x} \cdot \bar{y}} = (2\pi)^3 \delta^3(\bar{y})$, we have

$$\begin{aligned} H &= \frac{1}{2} (2\pi)^3 \int \widetilde{dk} \widetilde{dk'} [\delta^3(\bar{k} - \bar{k}')(\omega\omega' + \bar{k} \cdot \bar{k}' + m^2)(a^\star(\bar{k})a(\bar{k}')e^{-i(\omega-\omega')t} + a(\bar{k})a^\star(\bar{k}')e^{-i(\omega-\omega')t}) \\ &\quad + \delta^3(\bar{k} + \bar{k}')(-\omega\omega' - \bar{k} \cdot \bar{k}' + m^2)(a(\bar{k})a(\bar{k}')e^{-i(\omega+\omega')t} + a^\star(\bar{k})a^\star(\bar{k}')e^{i(\omega+\omega')t}) \\ &\quad - V\Omega] \\ &= \frac{1}{2} \int \widetilde{dk} \frac{1}{2\omega} [(\omega^2 + \bar{k}^2 + m^2)(a^\star(\bar{k})a(\bar{k}) + a(\bar{k})a^\star(\bar{k})) \\ &\quad + (-\omega^2 + \bar{k}^2 + m^2)(a(\bar{k})a(-\bar{k})e^{-2i\omega t} + a^\star(\bar{k})a^\star(-\bar{k})e^{2i\omega t})] - V\Omega \end{aligned}$$

and finally, using the definition of ω (equation (3.25)), this becomes

$$H = \frac{1}{2} \int \widetilde{dk} \, \omega(a^\star(\bar{k})a(\bar{k}) + a(\bar{k})a^\star(\bar{k})) - V\Omega$$

And now, using (3.28), we can rewrite this as (switching from \star to \dagger to emphasize the Hermitian nature)

$$\begin{aligned} H &= \frac{1}{2} \int \widetilde{dk} \, \omega(a^\dagger(\bar{k})a(\bar{k}) + a(\bar{k})a^\dagger(\bar{k})) - V\Omega \\ &= \frac{1}{2} \int \widetilde{dk} \, \omega(a^\dagger(\bar{k})a(\bar{k}) + (2\pi)^3 2\omega\delta^3(\bar{k} - \bar{k}) + a^\dagger(\bar{k})a(\bar{k})) - V\Omega \\ &= \int \widetilde{dk} \, \omega a^\dagger(\bar{k})a(\bar{k}) + \int \widetilde{dk} \, \omega(2\pi)^3 \delta^3(0) - V\Omega \\ &= \int \widetilde{dk} \omega a^\dagger(\bar{k})a(\bar{k}) + \int \frac{d^3\bar{k}}{(2\pi)^3 2\omega} \omega(2\pi)^3 \delta^3(0) - V\Omega \\ &= \int \widetilde{dk} \omega a^\dagger(\bar{k})a(\bar{k}) + \frac{1}{2} \delta^3(0) \int d^3\bar{k} - V\Omega \end{aligned}$$

Notice that both the second and third terms are infinite (assuming the volume V of the space we are in is infinite). This may be troubling, but remember that Ω is an arbitrary constant we can set to be anything we want. So, let's define

$$\Omega \equiv \frac{1}{2V} \delta^3(0) \int d^3\bar{k}$$

leaving

$$H = \int \widetilde{dk} \, \omega a^\dagger(\bar{k})a(\bar{k}) \tag{3.29}$$

Remember that measurement can only detect *changes* in energy, and therefore the infinity we subtracted off does not affect the value we will measure experimentally. What we have done here, by subtracting off the infinite part in a way that doesn't change the physics, is a very primitive example of **Renormalization**. Often, for various reasons, measurable quantities in QFT are plagued by different types of infinities. However, it is possible to subtract off those infinities in a well-defined way, leaving a finite part. It turns out that this finite part is the correct value seen in nature. The reasons for this are very deep, and we will not discuss them (or general renormalization theory) in much depth in these notes. For correlating theoretical results with experiment, being able to renormalize results correctly is vital. However, our goal is not to understand the subtleties of renormalization, but to understand the overall structure of particle physics. When you take a course on QFT you will spend a great, great deal of time on renormalization, and a deeper understanding of it will emerge.

So, we have our field expansion (3.27) and commutation relations (3.28). Notice that (3.28) have the exact form of a simple harmonic oscillator, which you learned about in introductory quantum mechanics. Therefore, because they have the same structure as the harmonic oscillator, they will have the same physics. By doing nothing but imposing relativity, we have found that scalar fields, which are Hermitian operators, act as raising and lowering (or synonymously creation and annihilation) operators on the vacuum (just like the simple harmonic oscillator).

Comparing (3.28) with the standard harmonic oscillator operators, it is clear that $a^\dagger(\bar{k})$ creates a ϕ particle with momentum \bar{k} and energy ω , whereas $a(\bar{k})$ annihilates a ϕ particle with momentum \bar{k} and energy ω . A normalized state will be

$$|\bar{k}\rangle = \sqrt{2\omega} a^\dagger(\bar{k})|0\rangle \quad (3.30)$$

The entire spectrum of states can be studied by acting on $|0\rangle$ with creation operators, and probability amplitudes for one state to be found in another, $\langle \bar{k}_f | \bar{k}_i \rangle$, are straightforward to calculate (and positive semi-definite). Naturally this theory does not discuss any interactions between particles, and therefore we will have to do a great deal of modification before we are done. But this simple exercise of merely imposing the standard commutation relations (3.24) between the field and its momentum, we have gained complete knowledge of the quantum mechanical states of the theory.

3.2.3 The Spin-Statistics Theorem

Notice that the states coming from (3.30) will include the two particle state

$$|\bar{k}; \bar{k}'\rangle = 2\sqrt{\omega\omega'} a^\dagger(\bar{k}) a^\dagger(\bar{k}')|0\rangle \quad (3.31)$$

But the commutation relations (3.28) tell us that $a^\dagger(\bar{k}) a^\dagger(\bar{k}') = a^\dagger(\bar{k}') a^\dagger(\bar{k})$. So, this theory also allows the state

$$|\bar{k}'; \bar{k}\rangle = 2\sqrt{\omega'\omega} a^\dagger(\bar{k}') a^\dagger(\bar{k})|0\rangle \quad (3.32)$$

Recall from a chemistry or modern physics course that particles with half-integer spin obey the Pauli Exclusion Principle, whereas particles of integer spin do not. Our Klein Gordon scalar fields ϕ are spinless ($j = 0$), and therefore we would expect that they do not obey Pauli exclusion. The fact that our commutation relations have allowed both states (3.31) and (3.32) is therefore expected. This is an indication that we quantized correctly.

But notice that this statistical result (that the scalar fields do not obey Pauli exclusion) is entirely a result of the commutation relations. Therefore, if we attempt to quantize a spin-1/2 field in the same way, they will obviously not obey Pauli exclusion either. We must therefore quantize spin-1/2 differently.

It turns out that the correct way to quantize spin-1/2 fields is to use, instead of commutation relations like we used for scalar fields, *anticommutation relations*. If the operators of our spin-1/2 fields obey

$$\{a_1^\dagger, a_2^\dagger\} = a_1^\dagger a_2^\dagger = 0 \Rightarrow a_1^\dagger a_2^\dagger = -a_2^\dagger a_1^\dagger$$

then if we try to act twice with the same operator, we have

$$a_1^\dagger a_1^\dagger |0\rangle = -a_1^\dagger a_1^\dagger |0\rangle \Rightarrow a_1^\dagger a_1^\dagger |0\rangle = 0$$

In other words, if we quantize with anticommutation relations, it is not possible for two particles to occupy the same state simultaneously.

This relationship between the spin of a particle and the statistics it obeys (which demands that integer spin particles be quantized by commutation relations and half-integer spin particles to be quantized with anticommutation relations) is called the **Spin-Statistics Theorem**.

And, because particles obeying Pauli exclusion are said to have **Bose-Einstein** statistics, and particles that do not obey Pauli exclusion are said to have **Fermi-Dirac** statistics, we call particles with integer spin **Bosons**, and particles with half-integer spin **Fermions**.

3.2.4 Left-Handed and Right-Handed Fields

Recall that in the Dirac Lagrangian (3.12), our fundamental field was the 4-component spinor $\psi = \begin{pmatrix} \psi_L \\ \psi_R \end{pmatrix}$ where ψ_L transforms under the left-handed $(0, 1/2)$ representation of the Lorentz group, and ψ_R transforms under the right-handed $(1/2, 0)$ representation.

In general, we refer to these 2-component spinors as **Weyl** fields (usually pronounced “vile”). So, the fermion is the spinor combination of two Weyl fields, one being the left-handed particle, and the other being the right-handed antiparticle.

Also in (3.12) was the field we defined as $\bar{\psi} = \psi^\dagger \gamma^0 = (\psi_R^\dagger, \psi_L^\dagger)$. If we interpret $\bar{\psi}$ as the conjugate of ψ (which the form of the Dirac Lagrangian implies we should), then we see that the right-handed field is the conjugate of the left, and vice versa. Or, in other words,

$$\psi_L^\dagger = \psi_R \quad \text{and} \quad \psi_R^\dagger = \psi_L$$

We take advantage of the fact by writing all fields in terms of left-handed Weyl fields. For example, given the two left-handed Weyl fields χ and ξ , we can form the 4-component spinor field $\psi = \begin{pmatrix} \chi \\ \xi^\dagger \end{pmatrix}$, and so $\bar{\psi} = (\xi, \chi^\dagger)$. We will refer to such a field as a **Dirac Field**, and denote it ψ_D .

On the other hand, we could define a 4-component spinor in terms of a single left-handed Weyl field χ , or $\psi = \begin{pmatrix} \chi \\ \chi^\dagger \end{pmatrix}$. But now notice that $\bar{\psi} = (\chi, \chi^\dagger)$, which is equal simply to the transpose of ψ . We refer to such a field (whose conjugate is equal to its transpose) as a **Majorana Field**, and denote it ψ_M .

Recall that an antiparticle has the same mass but opposite charge and opposite handedness of its particle. So, working with the Dirac field ψ_D , we can change the charge by merely swapping χ and ξ , using the **Charge Conjugation** operator \mathcal{C} defined by

$$\mathcal{C}\psi_D = \mathcal{C} \begin{pmatrix} \chi \\ \xi^\dagger \end{pmatrix} = \begin{pmatrix} \xi \\ \chi^\dagger \end{pmatrix}$$

Also, consider the transpose of $\bar{\psi}_D$ (which is just returning the conjugate of ψ_D to column form), $\bar{\psi}_D^T = \begin{pmatrix} \xi \\ \chi^\dagger \end{pmatrix}$. Acting on this with \mathcal{C} gives

$$\mathcal{C}\bar{\psi}_D^T = \mathcal{C} \begin{pmatrix} \xi \\ \chi^\dagger \end{pmatrix} = \begin{pmatrix} \chi \\ \xi^\dagger \end{pmatrix} = \psi_D$$

So, we have

$$\mathcal{C}\psi_D = \bar{\psi}_D^T \quad \text{and} \quad \mathcal{C}\bar{\psi}_D^T = \psi_D$$

We therefore say that ψ_D and $\bar{\psi}_D^T$ are **Charge Conjugate** to each other.

However, notice that with the Majorana field,

$$\mathcal{C}\psi_M = \mathcal{C} \begin{pmatrix} \chi \\ \chi^\dagger \end{pmatrix} = \begin{pmatrix} \chi \\ \chi^\dagger \end{pmatrix} = \psi_M$$

and

$$\mathcal{C}\bar{\psi}_M^T = \bar{\psi}_M^T = \psi_M$$

So in summary, Dirac fields are not equal to their charge conjugate, while Majorana fields are. By analogy with scalars (where the complex conjugate of a real number is equal to itself, whereas the complex conjugate of a complex number is not), we often refer to Majorana fields as **Real**, and to Dirac fields as **Complex**.

So, we can now write out the Lagrangian for Dirac and Majorana fields in terms of their Weyl fields:

$$\mathcal{L}_D = i\chi^\dagger \bar{\sigma}^\mu \partial_\mu \chi + i\xi^\dagger \bar{\sigma}^\mu \partial_\mu \xi - m(\chi\xi + \chi^\dagger \xi^\dagger) \quad (3.33)$$

$$\mathcal{L}_M = i\chi^\dagger \bar{\sigma}^\mu \partial_\mu \chi - \frac{1}{2}m(\chi\chi + \chi^\dagger \chi^\dagger) \quad (3.34)$$

3.2.5 Canonical Quantization of Fermions

We first quantize the Dirac fermion. The general solution to the Dirac equation is

$$\psi_D(x) = \sum_{s=1}^2 \int \widetilde{dk} [b_s(\bar{k}) u_s(\bar{k}) e^{ik \cdot x} + d_s^\dagger(\bar{k}) v_s(\bar{k}) e^{-ik \cdot x}]$$

where $s=1, 2$ are the two spin states, b_s and d_s^\dagger are (respectively) the lowering operator for the particle and the raising operator for the antiparticle. The charge conjugate of ψ_D will have the raising operator for the particle and the lowering operator for the antiparticle.

The u_s and v_s are constant 4-component vectors which act as a basis for all particle/antiparticle states in the spinor space (for our purposes, they are merely present to make ψ_D a 4-component field).

We quantize, as we said in section 3.2.3, using *anti*-commutation relations. Writing only the non-zero relation,

$$\{\psi_\alpha(t, \bar{x}), \bar{\psi}_\beta(t, \bar{x})\} = \delta^3(\bar{x} - \bar{x}') (\gamma^0)_{\alpha\beta}$$

These imply that the only non-zero commutation relations in terms of the operators are

$$\begin{aligned} \{b_s(\bar{k}), b_{s'}^\dagger(\bar{k}')\} &= (2\pi)^3 \delta^3(\bar{k} - \bar{k}') 2\omega \delta_{ss'} \\ \{d_s^\dagger(\bar{k}), d_{s'}(\bar{k}')\} &= (2\pi)^3 \delta^3(\bar{k} - \bar{k}') 2\omega \delta_{ss'} \end{aligned}$$

Once again, these form the algebra of a simple harmonic oscillator, and we can therefore find the entire spectrum of states by acting on $|0\rangle$ with b_s^\dagger and d_s^\dagger .

Then, following a series of calculations nearly identical to the ones in section 3.2.2, we arrive at the Hamiltonian

$$H = \sum_{s=1}^2 \int \widetilde{dk} \omega [b_s^\dagger(\bar{k}) b_s(\bar{k}) + d_s^\dagger(\bar{k}) d_s(\bar{k})] - \lambda \quad (3.35)$$

where λ is an infinite constant we can merely subtract off and therefore ignore.

Comparing (3.29) and (3.35), we see that they both have essentially the same form; ω (which is energy) to the left of the creation operator, which is to the left of the annihilation operator. To understand the meaning of this, we will see how it generates energy eigenvalues. We will use equation (3.29) for simplicity. Consider acting with the Hamiltonian

operator on some arbitrary state $|\bar{p}\rangle$ with momentum \bar{p} . Using (3.30),

$$\begin{aligned}
H|\bar{p}\rangle &= \int \widetilde{dk} \omega_k a^\dagger(\bar{k}) a(\bar{k}) |\bar{p}\rangle = \int \widetilde{dk} \omega_k a^\dagger(\bar{k}) a(\bar{k}) \sqrt{2\omega_p} a^\dagger(\bar{p}) |0\rangle \\
&= \int \widetilde{dk} \omega_k \sqrt{2\omega_p} a^\dagger(\bar{k}) ((2\pi)^3 2\omega_p \delta^3(\bar{k} - \bar{p}) + a^\dagger(\bar{p}) a(\bar{k})) |0\rangle \\
&= \int \widetilde{dk} \omega_k \sqrt{2\omega_p} a^\dagger(\bar{k}) (2\pi)^3 2\omega_p \delta^3(\bar{k} - \bar{p}) |0\rangle \\
&= \int \frac{d^3\bar{k}}{(2\pi)^3 2\omega_k} \omega_k \sqrt{2\omega_p} a^\dagger(\bar{k}) (2\pi)^3 2\omega_p \delta^3(\bar{k} - \bar{p}) |0\rangle \\
&= \int d^3\bar{k} \sqrt{2\omega_p} a^\dagger(\bar{k}) \omega_p \delta^3(\bar{k} - \bar{p}) |0\rangle \\
&= \omega_p \sqrt{2\omega_p} a^\dagger |0\rangle = \omega_p |\bar{p}\rangle
\end{aligned}$$

So, $H|\bar{p}\rangle = \omega_p |\bar{p}\rangle$, where $\omega_p = \bar{p}^2 + m^2$, which is the relativistic equation for energy as in equation (3.25). So, the Hamiltonian operator gives the appropriate energy eigenvalue on our physical quantum states.

For the Dirac Hamiltonian the eigenvalue will be a linear combination of the energies of each type of particle. If we denote the states as $|\bar{p}_b, s_b; \bar{p}_d, s_d\rangle$, where the first two elements give the state of a b type particle and the second of the d type particle, we have

$$H|\bar{p}_b, s_b; \bar{p}_d, s_d\rangle = \dots = (\omega_{p_b} + \omega_{p_d})|\bar{p}_b, s_b; \bar{p}_d, s_d\rangle$$

For Majorana fields things are simpler. We only have one type of particle, so

$$\psi_M(x) = \sum_{s=1}^2 \int \widetilde{dk} [b_s(\bar{k}) u_s(\bar{k}) e^{ik \cdot x} + b_s^\dagger(\bar{k}) v_s(\bar{k}) e^{-ik \cdot x}]$$

And quantization with anticommutation relations will give

$$H = \sum_{s=1}^2 \int \widetilde{dk} \omega b_s^\dagger(\bar{k}) b_s(\bar{k})$$

3.2.6 Insufficiencies of Canonical Quantization

While the Canonical Quantization procedure we have carried out in the past several sections has given us a tremendous amount of information (the entire spectrum of states for bosons, Dirac fermions, and Majorana fermions), it is still lacking quite a bit. As we said at the beginning of section 3.1.1, we ultimately want a relativistic quantum mechanical theory of interactions. Canonical Quantization has provided a relativistic quantum mechanical theory, but we aren't close to being able to incorporate interactions into our theory. While it is possible to incorporate interactions, it is very difficult, and in order to simplify we will need a new way of quantizing.

3.2.7 Path Integrals and Path Integral Quantization

Perhaps the most fundamental experiment in quantum mechanics is the **Double Slit** experiment. In brief, what this experiment tells us is that, when a single electron moves through a screen with two slits, and no observation is made regarding which slit it goes through, it actually goes through *both* slits, and until a measurement is made (for example, when it hits the observation screen behind the double slit), it exists in a superposition of *both* paths. As a result, the particle exhibits a wave nature, and the pattern that emerges on the observation screen is an interference pattern—the same as if a classical wave was passing through the double slit—all paths in the superposition of the single electron are interfering with each other, both destructively and constructively. Once the electron is observed on the observation screen, it collapses probabilistically into one of its possible states (a particular location on the observation screen).

If, on the other hand, you set up some mechanism to observe *which* of the two slits the electron travels through, then the observation has been made *before* the observation screen, and you no longer have the superposition, and therefore you no longer see any indication of an interference pattern. The electrons are behaving, in a sense, classically from the double slit to the observation screen in this case.

The meaning of this is that a particle that has not been observed will actually take every possible path at once. Once an observation has been made, there is some probability associated with each path. Some paths are very likely, and others are less likely (some are nearly impossible). But until observation, it actually exists in a superposition of all possible states/paths.

So, to quantize, we will create a mathematical expression for a “sum over all possible paths”. This expression is called a **Path Integral**, and will prove to be a much more useful way to quantize a physical system.

We begin this construction by considering merely the amplitude for a particle at position q_1 at time t_1 to propagate to q_2 at time t_2 . This amplitude will be given by

$$\langle q_2, t_2 | q_1, t_1 \rangle = \langle q_2 | e^{iH(t_2-t_1)} | q_1 \rangle$$

To evaluate this, we begin by dividing the time interval $T \equiv t_2 - t_1$ into $N + 1$ equal intervals of length $\delta t = \frac{T}{N+1}$ each. So, we can insert N complete sets of position eigenstates,

$$\langle q_2, t_2 | q_1, t_1 \rangle = \int_{-\infty}^{\infty} \prod_{i=1}^N dQ_i \langle q_2 | e^{-iH\delta t} | Q_N \rangle \langle Q_N | e^{-iH\delta t} | Q_{N-1} \rangle \cdots \langle Q_1 | e^{-iH\delta t} | q_1 \rangle \quad (3.36)$$

Let’s look at a single one of these amplitudes. We know that in nearly all physical theories, we can break the Hamiltonian up as $H = \frac{p^2}{2m} + V(Q)$. So, using the completeness of

momentum eigenstates,

$$\begin{aligned}
\langle Q_{i+1} | e^{-iH\delta t} | Q_i \rangle &= \langle Q_{i+1} | e^{-i\left(\frac{P^2}{2m} + V(Q)\right)\delta t} | Q_i \rangle \\
&= \langle Q_{i+1} | e^{-i\delta t \frac{P^2}{2m}} e^{-i\delta t V(Q)} | Q_i \rangle \\
&= \int dP' \langle Q_{i+1} | e^{-i\delta t \frac{P'^2}{2m}} | P' \rangle \langle P' | e^{-i\delta t V(Q)} | Q_i \rangle \\
&= \int dP' e^{-i\delta t \frac{P'^2}{2m}} e^{-i\delta t V(Q_i)} \langle Q_{i+1} | P' \rangle \langle P' | Q_i \rangle \\
&= \int dP' e^{-i\delta t \frac{P'^2}{2m}} e^{-i\delta t V(Q_i)} \frac{e^{iP'Q_{i+1}}}{\sqrt{2\pi}} \frac{e^{-iP'Q_i}}{\sqrt{2\pi}} \\
&= \int \frac{dP'}{2\pi} e^{iH\delta t} e^{iP'(Q_{i+1}-Q_i)} \\
&= \int \frac{dP'}{2\pi} e^{i\left[P'(Q_{i+1}-Q_i)-H\delta t\right]} \\
&= \int \frac{dP'}{2\pi} e^{i\delta t \left[P'\left(\frac{Q_{i+1}-Q_i}{\delta t}\right)-H\right]}
\end{aligned}$$

And taking the limit as $\delta t \rightarrow 0$, $\frac{Q_{i+1}-Q_i}{\delta t} \rightarrow \dot{Q}_i$. So,

$$\int \frac{dP'}{2\pi} e^{i\delta t \left[P'\left(\frac{Q_{i+1}-Q_i}{\delta t}\right)-H\right]} = \int \frac{dP'}{2\pi} e^{idt_{i+1}[P'\dot{Q}_i-H]}$$

where the subscript on dt merely indicates where the infinitesimal time interval “ends”. So, we can plug this into (3.36) and taking the limit as $\delta t \rightarrow 0$,

$$\begin{aligned}
\langle q_2, t_2 | q_1, t_1 \rangle &= \int_{-\infty}^{\infty} \prod_{i=1}^N dQ_i \langle q_2 | e^{-iH\delta t} | Q_N \rangle \langle Q_N | e^{-iH\delta t} | Q_{N-1} \rangle \cdots \langle Q_1 | e^{-iH\delta t} | q_1 \rangle \\
&= \lim_{N \rightarrow \infty} \int_{-\infty}^{\infty} \prod_{i=1}^N dQ_i \int \frac{dP'_i}{2\pi} e^{idt_2[P'_N\dot{Q}_N-H]} e^{idt_N[P'_{N-1}\dot{Q}_{N-1}-H]} \cdots e^{idt_1[P'_1\dot{Q}_1-H]} \\
&\quad \int_{-\infty}^{\infty} \mathcal{D}p \mathcal{D}q e^{-i \int_{t_1}^{t_2} dt (p\dot{q}-H)}
\end{aligned}$$

where $\mathcal{D}p = \prod_{i=1}^{\infty} dp_i$ and $\mathcal{D}q = \prod_{i=1}^{\infty} dq_i$.

And if p shows up quadratically (as it always does; $\frac{p^2}{2m}$), then we can merely do the Gaussian integral over p , resulting in an overall constant which we merely absorb back into the measure when we normalize. Then, recognizing that the integrand in the exponent is $p\dot{q} - H = \mathcal{L}$, we have

$$\langle q_2, t_2 | q_1, t_1 \rangle = \int \mathcal{D}q e^{i \int_{t_1}^{t_2} dt \mathcal{L}} = \int \mathcal{D}q e^{iS} \quad (3.37)$$

Formally, the measure of (3.37) has an infinite number of differentials, and therefore evaluating it would require doing an infinite number of integrals. This is to be expected, since

the point of the path integral is a sum over every possible path, of which there are an infinite number. So, because we obviously can't do an infinite number of integrals, we will have to find a clever way of evaluating (3.37). But before doing so, we discuss what the path integral means.

3.2.8 Interpretation of the Path Integral

Equation (3.37) says that, given an initial and final configuration (q_1, t_1) and (q_2, t_2) , absolutely *any* path between them is possible. This is the content of the $\mathcal{D}q$ part: it is the sum over all paths.

Then, for each of those paths, the integral assigns a *statistical weight* of e^{iS} to it, where the action S is calculated using *that* path (recall our comments in section 1.1.1 about S being a functional, not a function).

So, consider an arbitrary path q_0 , which receives statistical weight $e^{iS[q_0]}$. Now, consider a path q' very close to q_0 , only varying by a small amount: $q' = q_0 + \epsilon \delta q_0$. This will have statistical weight $e^{iS[q_0 + \epsilon \delta q_0]} = e^{iS[q_0] + i\epsilon \delta q_0 \frac{\delta S[q_0]}{\delta q}}$, where $\frac{\delta S}{\delta q}$ is the Euler Lagrange derivative (1.1)

$$\frac{\delta S}{\delta q} = \frac{d}{dt} \frac{\partial S}{\partial \dot{q}} - \frac{\partial S}{\partial q}$$

To make our intended result more obvious, we do a Wick rotation, taking $t \rightarrow it$, so $dt \rightarrow i dt$, and $S = \int dt \mathcal{L} \rightarrow i \int dt \mathcal{L} = iS$, and $e^{iS} \rightarrow e^{-S}$. Now, the path $q' = q_0 + \epsilon \delta q_0$ gets weight $e^{iS[q_0]} e^{-i\epsilon \delta q_0 \frac{\delta S[q_0]}{\delta q}}$.

So, if $\frac{\delta S}{\delta q}$ is very large, then the weight becomes exponentially small. In other words, the larger the variation of the action is, the less probable that path is.

So the *most* probable path is the one for the *smallest* value of $\frac{\delta S}{\delta t}$, or the path at which $\frac{\delta S}{\delta q} = 0$. And as we discussed in 1.1.1, this is the path of **Least Action**. Thus, we have recovered classical mechanics as the first order approximation of quantum mechanics.

So, the meaning of the path integral is that all imaginable paths are possible for the particle to travel in moving from one configuration to another. However, not all paths are equally probable. The likelihood of a given path is given by the action exponentiated, and therefore the most probable paths are the ones which minimize the action. This is the reason that, macroscopically, the world appears classical. The likelihood of every particle in, say, a baseball, simultaneously taking a path noticeably far from the path of least action is negligibly small.

We will find that path integral quantization provides an extremely powerful tool with which to create our relativistic quantum theory of interactions.

3.2.9 Expectation Values

Now that we have a way of finding $\langle q_2, t_2 | q_1, t_1 \rangle$, the natural question to ask next is how do we find expectation values like $\langle q_2, t_2 | Q(t') | q_1, t_1 \rangle$ or $\langle q_2, t_2 | P(t') | q_1, t_1 \rangle$. By doing a similar derivation as in the last section, it is easy to show that

$$\langle q_2, t_2 | Q(t') | q_1, t_1 \rangle = \cdots = \int \mathcal{D}q Q(t') e^{iS}$$

We will find that evaluating integrals of this form is simplified greatly through making use of **Functional Derivatives**. For some function $f(x)$, the functional derivative is defined by

$$\frac{\delta}{\delta f(y)} f(x) \equiv \delta(x - y)$$

Next, we modify our path integral by adding an **Auxiliary External Source** function, so that

$$\mathcal{L} \rightarrow \mathcal{L} + f(t)Q(t) + h(t)P(t)$$

So we now have

$$\langle q_2, t_2 | q_1, t_1 \rangle_{f,h} = \int \mathcal{D}q e^{\int dt (\mathcal{L} + fQ + hP)}$$

which allows us to write out expectation values in the simple form

$$\begin{aligned} \langle q_2, t_2 | Q(t') | q_1, t_1 \rangle &= \frac{1}{i} \frac{\delta}{\delta f(t')} \langle q_2, t_2 | q_1, t_1 \rangle_{f,h} \Big|_{f,h=0} = \int \mathcal{D}q Q(t') e^{iS + i \int dt (fQ + hP)} \Big|_{f,h=0} \\ &= \int \mathcal{D}q Q(t') e^{iS} \end{aligned}$$

or

$$\begin{aligned} \langle q_2, t_2 | P(t') | q_1, t_1 \rangle &= \frac{1}{i} \frac{\delta}{\delta h(t')} \langle q_2, t_2 | q_1, t_1 \rangle_{f,h} \Big|_{f,h=0} = \int \mathcal{D}q P(t') e^{iS + i \int dt (fQ + hP)} \Big|_{f,h=0} \\ &= \int \mathcal{D}q P(t') e^{iS} \end{aligned}$$

So, once we have $\langle q_2, t_2 | q_1, t_1 \rangle$, we can find any expectation value we want simply by taking successive functional derivatives.

3.2.10 Path Integrals with Fields

Because we can build whatever state we want by acting on the vacuum, the important quantity for us to work with will be the **Vacuum to Vacuum** expectation value, or VEV, $\langle 0|0\rangle$, and the various expectation values we can build through functional derivatives ($\langle 0|\phi\phi|0\rangle$, $\langle 0|\psi\phi\phi|0\rangle$, etc.).

For simplicity let's consider a scalar boson ϕ . The Lagrangian is given in equation (3.10). Using this, we can write the path integral

$$\langle 0|0\rangle = \int \mathcal{D}\phi e^{i \int d^4x \left[-\frac{1}{2} \partial^\mu \phi \partial_\mu \phi - \frac{1}{2} m^2 \phi^2 \right]} \equiv \int \mathcal{D}\phi e^{i \int d^4x \mathcal{L}_0}$$

We will eventually want to find expectation values, so we introduce the auxiliary field J , creating

$$\langle 0|0\rangle_J = \int \mathcal{D}\phi e^{i \int d^4x (\mathcal{L}_0 + J\phi)} \quad (3.38)$$

So, for example, $\langle 0|\phi|0\rangle = \frac{1}{i} \frac{\delta}{\delta J} \langle 0|0\rangle_J \Big|_{J=0}$.

Of course, we still have a path integral with an infinite number of integrals to evaluate. But, we are finally able to discuss how we can do the evaluation.

We define $Z_0(J) \equiv \langle 0|0\rangle_J$. Then, making use of the Fourier Transform of ϕ ,

$$\tilde{\phi}(k) = \int d^4x e^{-ikx} \phi(x) \quad \phi(x) = \int \frac{d^4k}{(2\pi)^4} e^{ikx} \tilde{\phi}(k)$$

we begin with the \mathcal{L}_0 part:

$$\begin{aligned} S_0 &= \int d^4x \mathcal{L}_0 = \int d^4x \left(-\frac{1}{2} \partial^\mu \phi \partial_\mu \phi - \frac{1}{2} m^2 \phi^2 \right) \\ &= \int d^4x \left[-\frac{1}{2} \partial^\mu \left(\int \frac{d^4k}{(2\pi)^4} e^{ik \cdot x} \tilde{\phi}(k) \right) \partial_\mu \left(\int \frac{d^4k'}{(2\pi)^4} e^{ik' \cdot x} \tilde{\phi}(k') \right) \right. \\ &\quad \left. - \frac{1}{2} m^2 \left(\int \frac{d^4k}{(2\pi)^4} e^{ik \cdot x} \tilde{\phi}(k) \right) \left(\int \frac{d^4k'}{(2\pi)^4} e^{ik' \cdot x} \tilde{\phi}(k') \right) \right] \\ &= \int d^4x \left[\frac{1}{2} \int \frac{d^4k d^4k'}{(2\pi)^8} e^{ik \cdot x} e^{ik' \cdot x} \tilde{\phi}(k) \tilde{\phi}(k') (k^\mu k'_\mu - m^2) \right] \\ &= \frac{1}{2} \int \frac{d^4k d^4k'}{(2\pi)^8} \tilde{\phi}(k) \tilde{\phi}(k') (k^\mu k'_\mu - m^2) \int d^4x e^{i(k+k') \cdot x} \\ &= \frac{1}{2} \int \frac{d^4k d^4k'}{(2\pi)^8} \tilde{\phi}(k) \tilde{\phi}(k') (k^\mu k'_\mu - m^2) (2\pi)^4 \delta^4(k+k') \\ &= -\frac{1}{2} \int \frac{d^4k}{(2\pi)^4} \tilde{\phi}(k) (k^2 + m^2) \tilde{\phi}(-k) \end{aligned}$$

Then, transforming the auxiliary field part,

$$\begin{aligned}
\int d^4x J(x)\phi(x) &= \int d^4x \left(\int \frac{d^4k}{(2\pi)^4} e^{ik \cdot x} \tilde{J}(k) \right) \left(\int \frac{d^4k'}{(2\pi)^4} e^{ik' \cdot x} \tilde{\phi}(k') \right) \\
&= \int \frac{d^4k d^4k'}{(2\pi)^8} \tilde{J}(k) \tilde{\phi}(k') \int d^4x e^{i(k+k') \cdot x} \\
&= \int \frac{d^4k d^4k'}{(2\pi)^8} \tilde{J}(k) \tilde{\phi}(k') (2\pi)^4 \delta^4(k+k') \\
&= \int \frac{d^4k}{(2\pi)^4} \tilde{J}(k) \tilde{\phi}(-k)
\end{aligned}$$

And because the integral is over all k^μ , we can rewrite this as

$$\int \frac{d^4k}{(2\pi)^4} \tilde{J}(k) \tilde{\phi}(-k) = \frac{1}{2} \int \frac{d^4k}{(2\pi)^4} (\tilde{J}(k) \tilde{\phi}(-k) + \tilde{J}(-k) \tilde{\phi}(k))$$

(we did this to get the factor of 1/2 out front in order to have the same coefficient as the \mathcal{L}_0 part from above).

So,

$$S = \frac{1}{2} \int \frac{d^4k}{(2\pi)^4} \left[-\tilde{\phi}(k)(k^2 + m^2)\tilde{\phi}(-k) + \tilde{J}(k)\tilde{\phi}(-k) + \tilde{J}(-k)\tilde{\phi}(k) \right]$$

Now, we make a change of variables,

$$\tilde{\chi}(k) \equiv \tilde{\phi}(k) - \frac{\tilde{J}(k)}{k^2 + m^2}$$

(Note that this leaves the measure of the path integral unchanged: $\mathcal{D}\phi \rightarrow \mathcal{D}\chi$.)

Plugging this in, we have,

$$\begin{aligned}
S &= \frac{1}{2} \int \frac{d^4k}{(2\pi)^4} \left[-\left(\tilde{\chi}(k) + \frac{\tilde{J}(k)}{k^2 + m^2} \right) (k^2 + m^2) \left(\tilde{\chi}(-k) + \frac{\tilde{J}(-k)}{k^2 + m^2} \right) \right. \\
&\quad \left. + \tilde{J}(k) \left(\tilde{\chi}(-k) + \frac{\tilde{J}(-k)}{k^2 + m^2} \right) + \tilde{J}(-k) \left(\tilde{\chi}(k) + \frac{\tilde{J}(k)}{k^2 + m^2} \right) \right] \\
&= \frac{1}{2} \int \frac{d^4k}{(2\pi)^4} \left[-\tilde{\chi}(k)(k^2 + m^2)\tilde{\chi}(-k) + \frac{\tilde{J}(k)\tilde{J}(-k)}{k^2 + m^2} \right]
\end{aligned}$$

(The point of all of this is that, in this form, we have all of the ϕ , or equivalently χ , dependence in the first term, with no ϕ or χ dependence on the second term.)

Finally, our path integral (3.38) is

$$\langle 0|0 \rangle_J = \int \mathcal{D}\chi e^{\frac{i}{2} \int \frac{d^4k}{(2\pi)^4} \left[-\tilde{\chi}(k)(k^2 + m^2)\tilde{\chi}(-k) + \frac{\tilde{J}(k)\tilde{J}(-k)}{k^2 + m^2} \right]}$$

Now, using some clever physical reasoning, we can see how to evaluate the infinite number of integrals in this expression. Notice that if we set $J = 0$, we have a free theory in which no interactions take place. This means that if we start with nothing (the vacuum), the probability of having nothing later is 100%. Or,

$$\langle 0|0\rangle_J|_{J=0} = 1 = \int \mathcal{D}\chi e^{\frac{i}{2} \int \frac{d^4 k}{(2\pi)^4} [-\tilde{\chi}(k)(k^2+m^2)\tilde{\chi}(-k)]}$$

And if that part is 1, then we have

$$\langle 0|0\rangle_J = \int \mathcal{D}\chi e^{\frac{i}{2} \int \frac{d^4 k}{(2\pi)^4} \frac{\tilde{J}(k)\tilde{J}(-k)}{k^2+m^2}}$$

And remarkably, the integrand *has no χ dependence!* Therefore, the infinite number of integrals over all possible paths becomes nothing more than a constant we can absorb into the normalization, leaving

$$\langle 0|0\rangle_J = e^{\frac{i}{2} \int \frac{d^4 k}{(2\pi)^4} \frac{\tilde{J}(k)\tilde{J}(-k)}{k^2+m^2}}$$

We can Fourier Transform back to coordinate space to get

$$Z_0(J) = \langle 0|0\rangle_J = e^{\frac{i}{2} \int d^4 x d^4 x' J(x) \Delta(x-x') J(x')} \quad (3.39)$$

where

$$\Delta(x-x') \equiv \int \frac{d^4 k}{(2\pi)^4} \frac{e^{ik \cdot (x-x')}}{k^2 + m^2}$$

is called the **Feynman Propagator** for the scalar field.

We can then find expectation values by operating on this with $\frac{1}{i} \frac{\delta}{\delta J}$ as described in section 3.2.9.

We can repeat everything we have just done for fermions, and while it is a great deal more complicated (and tedious), it is in essence the same calculation. We begin by adding the auxiliary function $\bar{\eta}\psi + \bar{\psi}\eta$, to get expectation values of $\bar{\psi}$ and ψ by using $\frac{1}{i} \frac{\delta}{\delta \eta}$ and $\frac{1}{i} \frac{\delta}{\delta \bar{\eta}}$, respectively.

We then Fourier Transform every term in the exponent and find that we can separate out the $\bar{\psi}$ and ψ dependence, allowing us to set the term which does depend on ψ and $\bar{\psi}$ equal to 1. Fourier Transforming back then gives

$$Z_0(\eta, \bar{\eta}) = e^{i \int d^4 x d^4 x' \bar{\eta}(x) S(x-x') \eta(x')} \quad (3.40)$$

where

$$S(x-x') = \int \frac{d^4 k}{(2\pi)^4} \frac{(-\gamma^\mu k_\mu + m) e^{ik \cdot (x-x')}}{k^2 + m^2}$$

is the Feynman propagator for fermion fields.

Recall that we are calling the auxiliary fields J , η , and $\bar{\eta}$ **Source Fields**. Comparing the form of the Lagrangian in equation (3.38) to (1.13) reveals why. J , η , and $\bar{\eta}$ behave mathematically as sources, giving rise to the field they are coupled to, in the same way that the electromagnetic source J^μ gives rise to the electromagnetic field A^μ . The meaning behind equations (3.39) (and (3.40)) is that J (or η and $\bar{\eta}$) act as sources for the fields, creating a ϕ (or ψ and $\bar{\psi}$) at spacetime point x , and absorbing it at point x' . The terms $\Delta(x - x')$ and $S(x - x')$ then represent the expression giving the probability amplitude $\langle 0|0 \rangle$ for that particular event to occur. In other words, the propagator is the statistical weight of a particle going from x to x' .

3.2.11 Interacting Scalar Fields and Feynman Diagrams

We can now consider how to incorporate interactions into our formalism, allowing us to finally have our relativistic quantum theory of interactions.

Beginning with the free scalar Lagrangian (3.10), we can add an interaction term \mathcal{L}_1 . At this point, we only have one type of particle, ϕ , so we can only have ϕ 's interacting with other ϕ 's. Terms proportional to ϕ or ϕ^2 are either constant or linear in the equations of motion, and therefore aren't valid candidates for interaction terms. So, the simplest expression we can have is

$$\mathcal{L}_1 = \frac{1}{3!} g \phi^3$$

where $\frac{1}{3!}$ is a conventional normalization, and g is a **Coupling Constant**. So our total Lagrangian is

$$\mathcal{L} = \mathcal{L}_0 + \mathcal{L}_1 = -\frac{1}{2} \partial^\mu \phi \partial_\mu \phi - \frac{1}{2} m^2 \phi^2 + \frac{1}{6} g \phi^3$$

and the path integral is

$$\begin{aligned} Z(J) &= \langle 0|0 \rangle_J \\ &= \int \mathcal{D}\phi e^{i \int d^4x [\mathcal{L}_0 + \mathcal{L}_1 + J\phi]} \\ &= \int \mathcal{D}\phi e^{i \int d^4x \mathcal{L}_1} e^{i \int d^4x [\mathcal{L}_0 + J\phi]} \\ &= \int \mathcal{D}\phi e^{i \int d^4x \mathcal{L}_1} Z_0(J) = \int \mathcal{D}\phi e^{i \int d^4x \mathcal{L}_1} \langle 0|0 \rangle_J \end{aligned}$$

But, recall that we can bring out a factor of ϕ from $\langle 0|0 \rangle_J$ using the functional derivative $\frac{1}{i} \frac{\delta}{\delta J}$. So, we can make the replacement

$$\mathcal{L}_1(\phi) \rightarrow \mathcal{L}_1\left(\frac{1}{i} \frac{\delta}{\delta J}\right) \Rightarrow \frac{1}{6} g \phi^3 \rightarrow \frac{g}{6} \left(\frac{1}{i} \frac{\delta}{\delta J}\right)^3$$

And notice that once this is done, there is no longer any ϕ dependence in $Z(J)$. So, with the free theory, we were able to remove the ϕ dependence, leading to (3.39). And here, we were able to remove it from the interaction term as well. So, once again, the infinite number of integrals in (3.37) will merely give a constant which we can absorb into the normalization.

This leaves the result

$$\begin{aligned} Z(J) &= e^{\frac{i}{6}g \int d^4x \left(\frac{1}{i} \frac{\delta}{\delta J(x)} \right)^3} Z_0(J) \\ &= e^{-\frac{1}{6}g \int d^4x \left(\frac{\delta}{\delta J(x)} \right)^3} e^{\frac{i}{2} \int d^4x d^4x' J(x) \Delta(x-x') J(x')} \end{aligned}$$

Now, we can do two separate Taylor expansions to these two exponentials,

$$Z(J) = \sum_{V=0}^{\infty} \frac{1}{V!} \left[-\frac{g}{6} \int d^4x \left(\frac{\delta}{\delta J(x)} \right)^3 \right]^V \times \sum_{P=0}^{\infty} \frac{1}{P!} \left[\frac{i}{2} \int d^4y d^4z J(y) \Delta(y-z) J(z) \right]^P \quad (3.41)$$

Now, recall that a functional derivative $\frac{1}{i} \frac{\delta}{\delta J}$, will remove a J term. Furthermore, after taking the functional derivatives, we will set $J = 0$ to get the physical result. So, for a term to survive, the $2P$ sources must all be exactly removed by the $3V$ functional derivatives.

So, using (3.41), we can expand in orders of g (the coupling constant), keeping only the terms which survive, and after removing the sources, evaluate the integrals over the propagators Δ . The value of the integral will then be the physical amplitude for a particular event.

In practice, a slightly different formalism is used to organize and keep track of each term in this expansion. Note that there will be P propagators Δ . We can represent each of these terms diagrammatically, by making each source a solid dot, each propagator a line, and let the g terms be vertices joining the lines together. There will be a total of V vertices, each joining 3 lines (matching the fact that we are looking at ϕ^3 theory; there would be 4 lines at each vertex for ϕ^4 theory, etc.).

For example, for $V = 0$ and $P = 1$,

$$Z(J) = \frac{i}{2} \int d^4y d^4z J(y) \Delta(y-z) J(z)$$

We have two sources, one located at z and the other located at y , so we draw two dots, corresponding to those locations. Then, the propagator $\Delta(y-z)$ connects them together, so we draw a line between the two dots. The diagram should look like this:

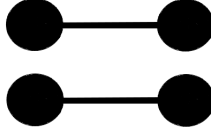


Of course, once we set $J = 0$, this will vanish because it contains two sources.

As another example, consider $V = 0$ and $P = 2$. Now,

$$Z(J) = \frac{1}{2!} \left(\frac{i}{2} \right)^2 \int d^4y d^4z d^4y' d^4z' (J(y) \Delta(y-z) J(z)) (J(y') \Delta(y'-z') J(z'))$$

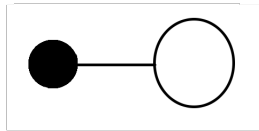
This corresponds to four sources, located at y, z, y' and z' , with propagator lines connecting y to z , and connecting y' to z' . But, there are no lines connecting an unprimed source to a primed source, so this results in two disconnected diagrams:



As another example, consider $V = 1$ and $P = 2$,

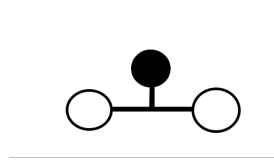
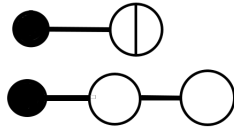
$$\begin{aligned} Z(J) &= -\frac{g}{6} \int d^4x \left(\frac{\delta}{\delta J(x)} \right)^3 \\ &\quad \times \frac{1}{2!} \left(\frac{i}{2} \right)^2 \int d^4y d^4z d^4y' d^4z' (J(y) \Delta(y-z) J(z)) (J(y') \Delta(y'-z') J(z')) \\ &= \frac{g}{48} \int d^4x d^4y d^4z d^4y' d^4z' \delta(y-x) \Delta(y-z) \delta(z-x) \delta(y'-x) \Delta(y'-z') J(z') \\ &= \frac{g}{48} \int d^4x d^4z' \Delta(x-x) \Delta(x-z') J(z') \end{aligned}$$

This will correspond to

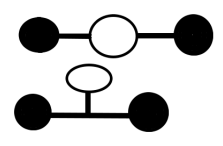


where the source J is located at the dot, and the vertex joining the line to the loop is at x .

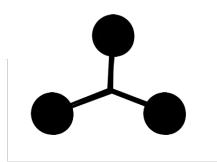
You can work out the following out, and see that there are multiple possible diagrams for $V = 3$, $P = 5$



And for $V = 2, P = 4$,



And for $V = 1, P = 3$,



and so on.

Through a series of combinatoric and physical arguments, it can be shown that only connected diagrams will contribute, and the $\frac{1}{P!}$ and $\frac{1}{V!}$ terms will always cancel exactly.

So, to calculate the amplitude for a particular interaction to happen (say N ϕ 's in and M ϕ 's out), draw every connected diagram that is topologically distinct, and has the correct number of in and out particles. Then, through a set of rules which you will learn formally in a QFT course, you can reconstruct the integrals which we started with in (3.41).

When you take a course on QFT, you will spend a tremendous amount of time learning how to evaluate these integrals for low order (they cannot be evaluated past about second order in most cases). While this is extremely important, it is not vital for the agenda of these notes, and we therefore do not discuss how they are evaluated.

The idea is that each diagram represents one of the possible paths the particle can take, along with the possible interactions it can be a part of. Because this is a quantum mechanical theory, we know it is actually in a superposition of all possible paths and interactions.

We don't make a measurement or observation until the particles leave the area in which they collide, so we have no idea about what is going on inside the accelerator. We know that if *this* goes in and *this* comes out, we can draw a particular set of diagrams which have the correct input and output, and the nature of the interaction terms (which determines what types of vertices you can have) tells us what types of interactions we can have inside the accelerator. Evaluating the integrals then tells us how much that particular event/diagram contributes towards the total probability amplitude. So, if you want to know how likely a certain incoming/outgoing set of particles is, write down all the diagrams, evaluate the corresponding integrals, and add them up.

And as we pointed out above, the classical behavior (which is more probable) is closer to the first order approximation of the quantum behavior. Therefore, even though in general we can't evaluate the integrals past about second order, the first few orders tell us to a reasonable (in fact, exceptional in most cases) degree of accuracy what the amplitude is. If we want more accuracy, we can seek to evaluate higher orders, but usually lower orders suffice for experiments at energy levels we can currently attain.

One of the difficulties encountered with evaluating these integrals is that you almost always find that they yield infinite amplitudes. Since an amplitude (which is a probability) should be between 0 and 1, this is obviously unacceptable. The process of finding the infinite parts and separating them from the finite parts of the amplitude is a very well defined mathematical construct called **Renormalization**. The basic idea is that any infinite term consists of a pure infinity and a finite part. For example (trust us for now) the infinite sum:

$$\sum_{n=1}^{\infty} n = \lim_{x \rightarrow 0} \frac{1}{x^2} - \frac{1}{12}$$

There is a part which is a pure infinity (the first term on the right hand side), and a term which is finite. While this may seem strange and extremely unfamiliar (and a bit like hand waving), it is actually a very rigorous and very well understood mathematical idea.

Much of what particle physicists attempt to do is find theories (and types of theories) that can be renormalized and theories that cannot. For example, the action which leads to General Relativity leads to a quantum theory which cannot be renormalized. Renormalization is a fascinating and deep topic, and will be covered in great depth in any standard QFT text or course. Unfortunately, we will not discuss it further.

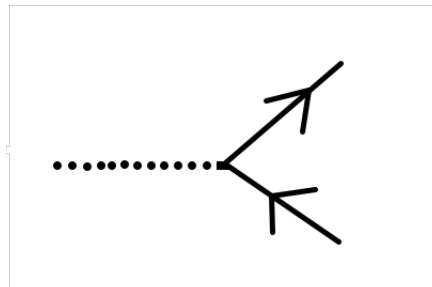
3.2.12 Interacting Fermion Fields

The analysis we performed above for scalar fields ϕ above is almost identical for fermions, and we therefore won't repeat it. The main difference is that the interaction terms will have a field ψ interacting with $\bar{\psi}$, and so the vertices will be slightly different. We won't bother with those details.

Finally, we can have a Lagrangian with *both* scalars and fermions. Then, naturally, you could have interaction terms where the scalars interact with fermions. While there are countless interaction terms of this type, the one that will be the most interesting to us is the **Yukawa** term,

$$\mathcal{L}_{Yuk} = g\phi\bar{\psi}\psi \quad (3.42)$$

If we represent ϕ by a dotted line, ψ by a line with an arrow in the forward time direction, and $\bar{\psi}$ with an arrow going backwards in time, this interaction term will show up in a Feynman diagram as



Once each diagram is drawn, there are well defined rules to write down an integral corresponding to each diagram.

3.3 Final Ingredients

The purpose of the previous section was merely to introduce the idea of **Feynman Diagrams** as a tool to calculate amplitudes for physical processes. In doing so, we have met the goal set out in section 3.1.1, a relativistic quantum mechanical theory of interactions. We achieve such a theory by finding a Lagrangian of a classical theory (both with and without interaction terms), and using equation (3.41) (and the analogous equation for fermions) to write down integrals which, when evaluated, give a contribution to a total amplitude. It is important to remember that we will eventually set all sources J to zero, and a functional derivative (as contained in the interaction term \mathcal{L}_1) will set any term without J 's to zero. So, the only non-zero terms will be the ones where all of the J 's are exactly removed by the functional derivatives.

A large portion of understanding QFT is learning how to set these integrals up in greater detail, and learning several methods to evaluate them. We will not delve into those details of **Perturbative Quantum Field Theory**, where amplitudes are studied order by order, here. The goal of these notes is merely to explain how, once given a Lagrangian, that Lagrangian can be turned into a physically measurable quantity.

With this done, we now set out to find the Lagrangian for the Standard Model of Particle Physics, the theory which seems to explain our universe (apart from gravity). Once this Lagrangian has been explained, we trust you have a general concept of what to do with it from the previous sections.

However, before we are able to explain the Standard Model Lagrangian, there are a few final concepts we need. They will be the subject of this section. Namely, we will be studying the ideas of **Spontaneous Symmetry Breaking** and **Gauge Theories**. In section 3.1.11, we discussed the simple $U(1)$ gauge theory, where we made a global $U(1)$ symmetry of the free Dirac Lagrangian a local $U(1)$ symmetry, or a gauged symmetry, and showed that consistency demanded the introduction of a gauge field A^μ , and consequently a kinetic term and a source term. Thus we recovered the entire electromagnetic force from nothing but $U(1)$. Later in this section, we generalize this to arbitrary Lie group. Because $U(1)$ is an Abelian group, we refer to the gauge theory of section 3.1.11 as an abelian gauge theory. For a more general, non-Abelian group, we refer to the theory resulting as a **Non-Abelian Gauge Theory**. Such theories introduce a great deal of complexity, and we therefore consider them in detail in this section before moving on to the Standard Model.

However, we begin with the idea of spontaneous symmetry breaking.

3.3.1 Spontaneous Symmetry Breaking

Consider a complex scalar boson ϕ and ϕ^\dagger . The Lagrangian will be

$$\mathcal{L} = -\frac{1}{2}\partial^\mu\phi^\dagger\partial_\mu\phi - \frac{1}{2}m^2\phi^\dagger\phi$$

Naturally we can write this as

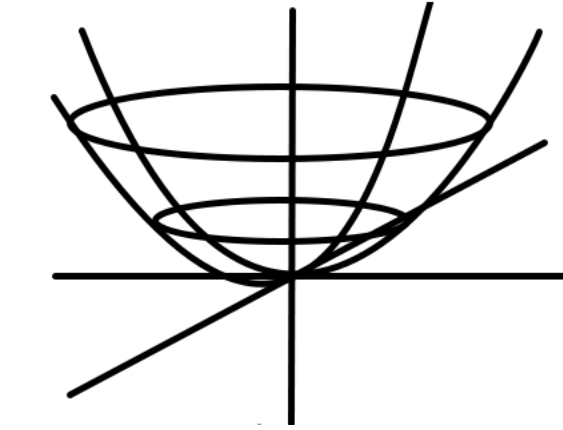
$$\mathcal{L} = -\frac{1}{2}\partial^\mu\phi^\dagger\partial_\mu\phi - V(\phi^\dagger, \phi) \tag{3.43}$$

where

$$V(\phi^\dagger, \phi) = \frac{1}{2}m^2\phi^\dagger\phi$$

This Lagrangian has the $U(1)$ symmetry we discussed in 3.1.11.

Also, notice that we can graph $V(\phi^\dagger, \phi)$, plotting V vs. $|\phi|$,

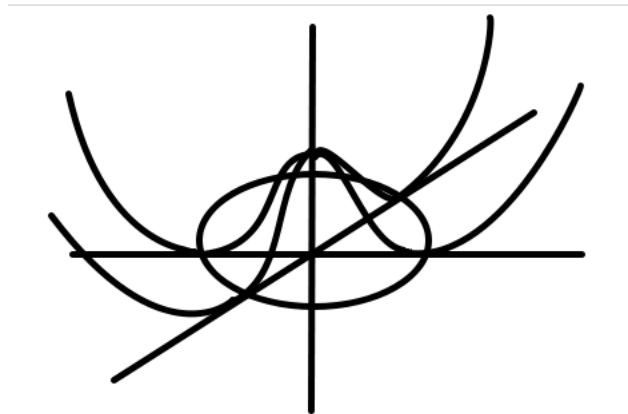


We see a “bowl” with V_{minimum} at $|\phi|^2 = 0$. The vacuum of any theory ends up being at the lowest potential point, and therefore the vacuum of this theory is at $\phi = 0$, as we would expect.

Now, let’s change the potential. Consider

$$V(\phi^\dagger, \phi) = \frac{1}{2} \lambda m^2 (\phi^\dagger \phi - \Phi^2)^2 \quad (3.44)$$

where λ and Φ are real constants. Notice that the Lagrangian will still have the global $U(1)$ symmetry from before. But, now if we graph V vs. $|\phi|$, we get



where now the vacuum V_{minimum} is represented by the circle at $|\phi| = \Phi$. In other words, there are an infinite number of vacuums in this theory. And because the circle drawn in the figure above represents a rotation through field space, this degenerate vacuum is parameterized by $e^{i\alpha}$, the global $U(1)$. There will be a vacuum for every value of α , located at $|\phi| = \Phi$.

In order to make sense of this theory, we must *choose* a vacuum by hand. Because the theory is completely invariant under the choice of the $U(1)$ $e^{i\alpha}$, we can choose any α and

define that as our true vacuum. So, we choose α to make our vacuum at $\phi = \Phi$, or where ϕ is real and equal to Φ . We have thus, in a sense, **Gauged Fixed** the symmetry in the Lagrangian, and the $U(1)$ symmetry is no longer manifest.

Now we need to rewrite this theory in terms of our new vacuum. We therefore expand around the constant vacuum value Φ to have the new field

$$\phi \equiv \Phi + \alpha + i\beta$$

where α and β are new real scalar fields (so $\phi^\dagger = \Phi + \alpha - i\beta$). We can now write out the Lagrangian as

$$\begin{aligned} \mathcal{L} &= -\frac{1}{2}\partial^\mu[\alpha - i\beta]\partial_\mu[\alpha + i\beta] - \frac{1}{2}\lambda m^2[(\Phi + \alpha - i\beta)(\Phi + \alpha + i\beta) - \Phi^2]^2 \\ &= \left[-\frac{1}{2}\partial^\mu\alpha\partial_\mu\alpha - \frac{1}{2}4\lambda m^2\Phi^2\alpha^2 - \frac{1}{2}\partial^\mu\beta\partial_\mu\beta \right] - \frac{1}{2}\lambda m^2 \left[4\Phi\alpha^3 + 4\Phi\alpha\beta^2 + \alpha^4 + \alpha^2\beta^2 + \beta^4 \right] \end{aligned} \quad (3.45)$$

This is now a theory of a *massive* real scalar field α (with mass $= \sqrt{4\lambda m^2\Phi^2}$), a *massless* real scalar field β , and five different types of interactions (one allowing three α 's to interact, the second allowing one α and two β 's, the third allowing four α 's, the fourth allowing two α 's and two β 's, and the last allowing four β 's.) In other words, there are five different types of vertices allowed in the Feynman diagrams for this theory.

Furthermore, notice that this theory has no obvious $U(1)$ symmetry. For this reason, writing the field in terms of fluctuations around the vacuum we choose is called “breaking” the symmetry. The symmetry is still there, but it can't be seen in this form.

Finally, notice that breaking the symmetry has resulted in the addition of the massless field β . It turns out that breaking global symmetries as we have done *always* results in a massless boson. Such particles are called **Goldstone Bosons**.

3.3.2 Breaking Local Symmetries

In the previous section, we broke a global $U(1)$ symmetry. In this section, we will break a local $U(1)$ and see what happens. We begin with the Lagrangian for a complex scalar field with a gauged $U(1)$:

$$\mathcal{L} = -\frac{1}{2}[(\partial^\mu - iqA^\mu)\phi^\dagger][(\partial_\mu + iqA_\mu)\phi] - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - V(\phi^\dagger, \phi)$$

where we have taken the external source $J^\mu = 0$. Let's once again assume $V(\phi^\dagger, \phi)$ has the form of equation (3.44), so the vacuum has the $U(1)$ degeneracy at $|\phi| = \Phi$.

Because our $U(1)$ is now local, we choose $\alpha(x)$ so that not only is the vacuum real, but also so that ϕ is always real. We therefore expand

$$\phi = \Phi + h \quad (3.46)$$

where h is a real scalar field representing fluctuations around the vacuum we chose.

Now,

$$\begin{aligned}
\mathcal{L} &= -\frac{1}{2}[(\partial^\mu - iqA^\mu)(\Phi + h)][(\partial_\mu + iqA_\mu)(\Phi + h)] - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \\
&\quad - \frac{1}{2}\lambda m^2[(\Phi + h)(\Phi + h) - \Phi^2]^2 \\
&= \dots \\
&= -\frac{1}{2}\partial^\mu h \partial_\mu h - \frac{1}{2}4\lambda m^2\Phi^2 h^2 - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} - \frac{1}{2}q^2\Phi^2 A^2 + \mathcal{L}_{interactions}
\end{aligned}$$

where the allowed interaction terms include a vertex connecting an h and two $A^{\mu'}$ s, four h 's, and three h 's.

So, before breaking, we had a complex scalar field ϕ and a *massless* vector field A^μ with two polarization states (because it is a photon). Now, we have a single real scalar h with mass $= \sqrt{4\lambda m^2\Phi^2}$ and a field A^μ with mass $= q\Phi$. In other words, our force-carrying particle A^μ has gained mass! We started with a theory with no mass, and by merely breaking the symmetry, we have introduced mass into our theory.

This mechanism for introducing mass into a theory, called the **Higgs Mechanism**, was first discovered by Peter Higgs, and the resulting field h is called the **Higgs Boson**.

So, whereas the consequence of global symmetry breaking is a massless boson called a Goldstone boson, the consequence of a local symmetry breaking is that the gauge field, which came about as a result of the symmetry being local, acquires mass.

3.3.3 Non-Abelian Gauge Theory

We are now ready to generalize what we did in section 3.1.11 to an arbitrary Lie group.

Consider a Lagrangian \mathcal{L} with N scalar (or spinor) fields ϕ_i ($i = 1, \dots, N$) that is invariant under a continuous $SO(N)$ or $SU(N)$ symmetry, $\phi_i \rightarrow U_{ij}\phi_j$, where U_{ij} is an $N \times N$ matrix of $SO(N)$ or $SU(N)$.

In section 3.1.11, we saw that if the group is $U(1)$, gauging it demands the introduction of the gauge field A^μ to preserve the symmetry, which shows up in the covariant derivative $D_\mu = \partial_\mu - ieA_\mu$. To say a field carried some sort of charge means that it has the corresponding term in its covariant derivative. We then added a kinetic term for A^μ as well as an external source J^μ . Then, higher order interaction terms can be included in whatever way is appropriate for the theory.

To generalize this, let's say for the sake of concreteness that our Lie group is $SU(N)$. An arbitrary element of $SU(N)$ is $e^{ig\theta^a(x)T^a}$, where g is a constant we have added for later convenience, θ^a are the $N^2 - 1$ parameters of the group (cf. section 2.2.15), and the T^a are

the generator matrices for the group. Notice that we have gauged the symmetry (in that $\theta(x)$ is a function of spacetime).

By definition, we know that the generators T^a will obey the commutation relations

$$[T^a, T^b] = if_{abc}T^c$$

(cf equation (2.9)), where f_{abc} are the structure constants of the group.

When gauging the $U(1)$ in section 3.1.11, the transformation of the gauge field was given by equation (3.16). For the more general transformation $\phi_i \rightarrow U_{ij}\phi_j$, the gauge field transforms according to

$$A^\mu \rightarrow U(x)A^\mu U^\dagger(x) + \frac{i}{g}U(x)\partial^\mu U^\dagger(x)$$

(where we have removed the indicial notation and it is understood that matrix multiplication is being discussed). If $U(x)$ is an element of $U(1)$ (so it is $e^{ig\theta(x)}$), then this transformation reduces to

$$A^\mu \rightarrow e^{ig\theta(x)}A^\mu e^{-ig\theta(x)} + \frac{i}{g}e^{ig\theta(x)}(-ig\partial^\mu\theta(x))e^{-ig\theta(x)} = A^\mu + \partial^\mu\theta(x)$$

which is exactly what we had in (1.15). For general $SU(N)$, however, the U 's are elements of a Non-Abelian group, and the A^μ 's are matrices of the same size.

Generalizing, we find that a general element of the $SU(N)$ is (changing notation slightly)

$$U(x)e^{-ig\Gamma^a(x)T^a}$$

with $N^2 - 1$ real parameters Γ^a . We then build the covariant derivative in the exact same way as in equation (3.17) by adding a term proportional to the gauge field

$$D_\mu = \mathbb{I}^{N \times N} \partial_\mu - igA_\mu$$

(Remember that each component of A_μ is an $N \times N$ matrix. They were scalars for $U(1)$ because $U(1)$ is a 1×1 matrix.) Or, acting on the fields, the covariant derivative is

$$(D_\mu\phi)_j = \partial_\mu\phi_j(x) - ig[A_\mu(x)]_{jk}\phi_k(x) \quad (3.47)$$

where k is understood to be summed on the last term. It will be understood from now on that the normal partial derivative term (the first term) has an $N \times N$ identity matrix multiplied by it.

Then, just as in (3.19), we have the field strength

$$F_{\mu\nu}(x) \equiv \frac{i}{g}[D_\mu, D_\nu] = \partial_\mu A_\nu - \partial_\nu A_\mu - ig[A_\mu, A_\nu] \quad (3.48)$$

where the commutator term doesn't vanish for arbitrary Lie group as it did for Abelian $U(1)$.

Recall from equation (1.16) that for $U(1)$, $F_{\mu\nu}$ is invariant under the gauge transformation (1.15) on its own, because the commutator term vanishes. In general, however, the commutator term does not vanish, and we must therefore be careful in writing down the correct kinetic term. It turns out that the correct choice is

$$\mathcal{L}_{Kin} = -\frac{1}{2} \text{Tr} (F_{\mu\nu} F^{\mu\nu}) \quad (3.49)$$

It may not be obvious, but this form is actually a consequence of (2.28). There is algebraic machinery working under the surface of this that, while extremely interesting, is unfortunately beyond the scope of what we are doing. We will discuss all of these ideas in much greater depth later in this series.

So, starting with a non-interacting Lagrangian that is invariant under the global $SU(N)$, we can gauge the $SU(N)$ to create a theory with a gauge field (or synonymously a “force carrying” field) A^μ , which is an $N \times N$ matrix. So, every Lie group gives rise to a particular gauge field (which is a force carrying particle, like the photon), and therefore a particular force.

For this reason, we discuss forces in terms of Lie groups, or synonymously **Gauge Groups**. Each group defines a force. As we said at the very end of section 2.2.11, $U(1)$ represents the electromagnetic force (as we have seen in section 3.1.11, while $SU(2)$ describes the weak force, and $SU(3)$ describes the strong color force.

3.3.4 Representations of Gauge Groups

As we discussed in section 2.2, given a set of structure constants f_{abc} , which define the Lie algebra of some Lie group, we can form a representation of that group, which we denote R . So, R will be a set of $D(R) \times D(R)$ matrices, where D is the dimension of the representation R . We then call the generators of the group (in the representation R) T_R^a , and they naturally obey $[T_R^a, T_R^b] = if_{abc} T_R^c$.

One representation which exists for any of the groups we have considered is the representation of $SO(N)$ or $SU(N)$ consisting of $N \times N$ matrices. We denote this the **Fundamental Representation** (also called **Defining Representation** in some books). Clearly, the fundamental representations of $SO(2)$, $SO(3)$, $SU(2)$, and $SU(3)$ are the 2×2 , 3×3 , 2×2 , and 3×3 matrix representations, respectively. We will denote the fundamental representation for a given group by writing the number in bold. So, the fundamental representation of $SU(2)$ will be denoted **2**, and the generators for $SU(2)$ in the fundamental representation will be denoted T_2^a . Obviously, the fundamental representation of $SU(3)$ will be **3** with generators T_3^a .

Furthermore, let's say we have some arbitrary representation generated by T_R^a , obeying $[T_R^a, T_R^b] = if_{abc} T_R^c$. We can take the complex conjugate of the commutation relations to get $[T_R^{*a}, T_R^{*b}] = -if_{abc} T_R^{*c}$. So, notice that if we define the *new* set of generators $T_R'^a \equiv -T_R^{*a}$,

then the $T_R'^a$ will obey the correct commutation relations, and will therefore form a representation of the group as well. If it turns out that $T_R'^a = -(T_R^a)^* = T_R^a$, or if there is some unitary similarity transformation $T_R^a \rightarrow U^{-1}T_R^aU$ such that $T_R'^a = -(T_R^a)^* = T_R^a$, then we call the representation **Real**, and the complex conjugate of the T_R^a 's is the same representation. However, if no such transformation exists, then we have a *new* representation, called the **Complex Conjugate** representation to R , or the **Anti- R** representation, which we denote \bar{R} .

For example, there is the fundamental representation of $SU(3)$, denoted $\mathbf{3}$, generated by T_3^a , and then there is the anti-fundamental representation $\bar{\mathbf{3}}$, generated by $T_{\bar{\mathbf{3}}}^a$.

The representations of a group which will be important to us are the fundamental, anti-fundamental, and adjoint.

3.3.5 Symmetry Breaking Revisited

As we said in section 3.3.3, given a field transforming in a particular representation R , the gauge fields A^μ will be $D(R) \times D(R)$ matrices.

Once we know what representation we are working in, and therefore know the generators T_R^a , it turns out that it is always possible to write the gauge fields in terms of the generators. Recall in sections 2.2.2 and 2.2.12, we encouraged you to think of the generators as basis vectors which span the parameter space for the group. Because the gauge fields live in the $N \times N$ space as well, we can write them in terms of the generators. That is, instead of the gauge fields being $N \times N$ matrices on their own, we will use the $N \times N$ matrix generators as basis vectors, and then the gauge fields can be written as scalar coefficients of each generator:

$$A^\mu = A_a^\mu T_R^a \quad (3.50)$$

where a is understood to be summed, and each A_a^μ is now a scalar function rather than a $D(R) \times D(R)$ matrix (the advantage of this is that we can continue to think of the gauge fields as scalars with an extra index, rather than as matrices). As a note, we haven't done anything particularly profound here. We are merely writing each component of the $D(R) \times D(R)$ matrix A^μ in terms of the $D(R) \times D(R)$ generators, allowing us to work with a *scalar* field A_a^μ rather than the matrix field A^μ . We now actually view each A_a^μ as a separate field. So, if a group has N generators, we say there are N gauge fields associated with it, each one having 4 spacetime components μ .

In matrix components, we will have

$$(A^\mu)_{ij} = (A_a^\mu T_R^a)_{ij}$$

Then, the covariant derivative in (3.47) will be

$$(D_\mu \phi)_j = \partial_\mu \phi_j(x) - ig[A_{\mu a}(x)T_R^a]_{jk}\phi_k(x) \quad (3.51)$$

We may assume that the field strength $F^{\mu\nu}$ can also be expressed in terms of the generators, so that we have

$$F^{\mu\nu} = F_a^{\mu\nu} T^a \quad (3.52)$$

or

$$(F^{\mu\nu})_{ij} = (F_a^{\mu\nu} T^a)_{ij} \quad (3.53)$$

Now, using (2.28) (and taking $\kappa = 1/2$ by convention), we can write (3.49) in terms of the new basis:

$$\begin{aligned} \mathcal{L}_{Kin} = -\frac{1}{2} \text{Tr} (F_{\mu\nu} F^{\mu\nu}) &= -\frac{1}{2} \text{Tr} (F_a^{\mu\nu} T^a F_{\mu\nu b} T^b) \\ &= -\frac{1}{2} F_a^{\mu\nu} F_{\mu\nu b} \text{Tr} (T^a T^b) \\ &= -\frac{1}{2} F_a^{\mu\nu} F_{\mu\nu b} \kappa \delta^{ab} \\ &= -\frac{1}{2} F_a^{\mu\nu} F_{\mu\nu}^a \kappa \\ &= -\frac{1}{4} F_a^{\mu\nu} F_{\mu\nu}^a \end{aligned} \quad (3.54)$$

(we have raised the index a on the second field strength term in the last two lines simply to explicitly imply the summation over it. The fact that it is raised doesn't change its value in this case; it is merely notational).

Furthermore, we can use (2.28) to invert (3.52):

$$\begin{aligned} F^{\mu\nu} = F_a^{\mu\nu} T^a &\Rightarrow F^{\mu\nu} T^b = F_a^{\mu\nu} T^a T^b \\ &\Rightarrow \text{Tr} (F^{\mu\nu} T^b) = F_a^{\mu\nu} \text{Tr} (T^a T^b) \\ &\Rightarrow \text{Tr} (F^{\mu\nu} T^b) = F_a^{\mu\nu} \kappa \delta^{ab} \\ &\Rightarrow \text{Tr} (F^{\mu\nu} T^b) = \frac{1}{2} F_b^{\mu\nu} \\ &\Rightarrow F_a^{\mu\nu} = 2 \text{Tr} (F^{\mu\nu} T^a) \end{aligned} \quad (3.55)$$

In sections 3.3.1 and 3.3.2, we broke the $U(1)$ symmetry, which only had one generator. However, if we break larger groups we may only break part of it. For example, we will see that $SU(3)$ has an $SU(2)$ subgroup. It is actually possible to break only the $SU(2)$ part of the $SU(3)$. So, three of the $SU(3)$ generators are broken (the three corresponding to the $SU(2)$ subgroup/subalgebra), and the other five are unbroken. Because we are now writing our gauge fields using the generators as a basis, this means that three of the gauge fields are broken, while five of the gauge fields are not.

Finally, recall from section 3.3.2 that breaking a local symmetry results in a gauge field gaining mass. We seek now to elucidate the relationship between breaking a symmetry

and a field gaining mass. First, we can summarize as follows: *Gauge fields corresponding to broken generators get mass, while those corresponding to unbroken generators do not. The unbroken generators form a new gauge group that is smaller than the original group that was broken.*

In 3.3.2, we saw that breaking a symmetry gave the gauge field mass. Now, we see that giving a gauge field mass will break the symmetry.

To make this clearer, we begin with a very simple example, then move on to a more complicated example.

3.3.6 Simple Examples of Symmetry Breaking

Consider a theory with three real massless scalar fields ϕ_i ($i = 1, 2, 3$) and with Lagrangian

$$\mathcal{L} = -\frac{1}{2}\partial^\mu\phi_i\partial_\mu\phi^i$$

which is clearly invariant under the global $SO(3)$ rotation

$$\phi_i \rightarrow R^{ij}\phi_j$$

where R^{ij} is an element of $SO(3)$, because the Lagrangian is merely a dot product in field space, and we know that dot products are invariant under $SO(3)$.

Now, let's say that one of the fields, say ϕ_1 , gains mass. The new Lagrangian will then be

$$\mathcal{L} = -\frac{1}{2}\partial^\mu\phi_i\partial_\mu\phi^i - \frac{1}{2}m^2\phi_1^2$$

So this Lagrangian is no longer invariant under the full $SO(3)$ group, which mixes any two of the three fields. Rather, it is only invariant under rotations in field space that mix ϕ_2 and ϕ_3 or $SO(2)$. In other words, giving one field mass broke $SO(3)$ to the smaller $SO(2)$.

As another simple example, if we started with five massless complex scalar fields ϕ_i , with Lagrangian

$$\mathcal{L} = -\frac{1}{2}\partial^\mu\phi_i^\dagger\partial_\mu\phi^i$$

This will be invariant under any $SU(5)$ transformation.

Then let's say we give two of the fields, ϕ_1 and ϕ_2 (equal) mass. The new Lagrangian will be

$$\mathcal{L} = -\frac{1}{2}\partial^\mu\phi_i^\dagger\partial_\mu\phi^i - \frac{1}{2}m(\phi_1^\dagger\phi_1 + \phi_2^\dagger\phi_2)$$

So now, we no longer have the full $SU(5)$ symmetry, but we do have the special unitary transformations mixing ϕ_3 , ϕ_4 , and ϕ_5 . This is an $SU(3)$ subgroup. Furthermore, we can do a special unitary transformation mixing ϕ_1 and ϕ_2 . This is an $SU(2)$ subgroup. So, we have broken $SU(5) \rightarrow SU(3) \otimes SU(2)$.

Before considering a more complicated example of this, we further discuss the connection between symmetry breaking and fields gaining mass.

When we introduced spontaneous symmetry breaking in section 3.3.1, recall that we shifted the potential minimum from V_{minimum} at $\phi = 0$ to V_{minimum} at $|\phi| = \Phi$. But we were discussing this in very classical language. We can interpret all of this in a more “quantum” way in terms of VEV’s. As we said, the vacuum of a theory is defined as the minimum potential field configuration. For the V_{minimum} at $\phi = 0$ potential, the VEV of the field ϕ was at 0, or

$$\langle 0|\phi|0\rangle = 0$$

However, for the V_{minimum} at $|\phi| = \Phi$ potential, we have

$$\langle 0|\phi|0\rangle = \Phi$$

So, in quantum mechanical language, symmetry breaking occurs when a field, or some components of a field, take on a non-zero VEV.

This seems to be what is happening in nature. At higher energies, there is some “Master Theory” with some gauge group defining the physics, and all of the fields involved have 0 VEV’s. At lower energies, for whatever reason (the reason for this is not well understood at the time of this writing), some of the fields take on non-zero VEV’s, which break the symmetry into smaller groups, giving mass to certain fields through the Higgs Mechanism discussed in section 3.3.2. We call the theory with the unbroken gauge symmetry at higher energies the more fundamental theory (analogous to equation (3.43)), and the Lagrangian which results from breaking the symmetry (analogous to (3.45)) the **Low Energy Effective Theory**.

And this is how mass is introduced into the Standard Model. It turns out that if a theory is renormalizable one can prove that any lower energy effective theory that results from breaking the original theory’s symmetry is also renormalizable, even if it doesn’t appear to be. And, because the actions that appear to describe the universe at the energy level we live at (and the levels attainable by current experiment) are not renormalizable when they have mass terms, we work with a larger theory which has no massive particles but can be renormalized, and use the Higgs Mechanism to give various particles mass. So, whereas the physics we see at low energies may not appear renormalizable, if we can find a renormalizable theory which breaks down to our physics, we are safe.

Now, we consider a slightly more complicated (and realistic) example of symmetry breaking.

3.3.7 A More Complicated Example of Symmetry Breaking

Consider the gauge group $SU(N)$, acting on N complex scalar fields ϕ_i ($i = 1, \dots, N$) in the fundamental representation \mathbf{N} . Recall that in section 3.3.2, in order to get equation (3.46), we made use of the $U(1)$ symmetry to make the vacuum, or the VEV, real. We can now do something similar: we make use of the $SU(N)$ to not only make the VEV real, but also to rotate it to a single component of the field, ϕ_N . In other words, we do an $SU(N)$ rotation so that

$$\begin{aligned}\langle 0|\phi_i|0\rangle &= 0 & \text{for } i = 1, \dots, N-1 \\ \langle 0|\phi_N|0\rangle &= \Phi\end{aligned}$$

So, we expand ϕ_N around this new vacuum:

$$\begin{aligned}\phi_i &= \phi_i & \text{for } i = 1, \dots, N-1 \\ \phi_N &= \Phi + \chi\end{aligned}$$

This means that, in the vacuum configuration, the fields will have the form

$$\begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_N \end{pmatrix}_{\text{vacuum}} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ \Phi \end{pmatrix}$$

So, how will the action of $SU(N)$ be affected by this VEV? If we consider a general element of $SU(N)$ acting on this,

$$\begin{pmatrix} U_{11} & U_{12} & \cdots & U_{1N} \\ U_{21} & U_{22} & \cdots & U_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ U_{N1} & U_{N2} & \vdots & U_{NN} \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ \vdots \\ \Phi \end{pmatrix} = \begin{pmatrix} U_{1N} \\ U_{2N} \\ \vdots \\ U_{NN} \end{pmatrix}$$

So, only elements of $SU(N)$ with non-zero elements in the last column will be affected by this VEV. But the other $N-1$ elements' rows and columns are unaffected. This means that we have an $SU(N-1)$ symmetry left. Or in other words, we have broken $SU(N) \rightarrow SU(N-1)$ with this VEV.

Let's consider a specific example of this. Consider $SU(3)$. The generators are written out in (2.46). Notice that exactly three of them have all zeros in the last column; λ^1 , λ^2 , and λ^3 . We expect these three to give an $SU(3-1) = SU(2)$ subgroup. And looking at the upper left 2×2 boxes in those three generators, we can see that they are the Pauli matrices, the generators of $SU(2)$. So, if we give a non-zero VEV to the fields transforming under $SU(3)$, we see that they do indeed break the $SU(3)$ to $SU(2)$. The other five generators of $SU(3)$ will be affected by the VEV, and consequently the corresponding fields will acquire mass.

3.4 Particle Physics

3.4.1 Introduction to the Standard Model

We are finally ready to study the **Standard Model of Particle Physics**, which (except for gravity) appears to be the theory which explains our universe. To state the Standard Model in the simplest possible terms, it is

A Yang-Mills (Gauge) Theory with Gauge Group

$$SU(3) \otimes SU(2) \otimes U(1)$$

with left-handed Weyl fields in three copies of the representation

$$(1, 2, -1/2) \oplus (1, 1, 1) \oplus (3, 2, 1/6) \oplus (\bar{3}, 1, -2/3) \oplus (\bar{3}, 1, 1/3)$$

(where the last entry specifies the value of the $U(1)$ hypercharge),

and a single copy of a complex scalar field in the representation

$$(1, 2, -1/2)$$

Admittedly, our exposition will be somewhat cursory. This is largely because every concept and tool we use in this section has been discussed in detail in the previous sections. The purpose of these notes is to provide an introduction to the primary concepts and mathematical tools used in Particle Physics, not to give the details of the theory. We will cover the main points of the Standard Model, but there is tremendous detail we are skipping over. A second reason the following section is cursory is that we will not be working out every step in detail, as we have been doing. For nearly all calculations being done in this section, we have worked out a similar tedious calculation previously. We will therefore frequently refer to previous sections/equations. It will be worthwhile to go back and carefully study the parts which we refer to.

Because this section is slightly more experimental, or at least phenomenological, than the rest, and because the general purpose of these notes is to develop the mathematical tools and framework of particle physics (especially gauge theory), undue attention should not be given to this section. The purpose is merely to show, as briefly as possible, where everything we have done so far lines up with experiment. It will be useful to read through this section, but do not spend too much time bogged down in the details.

Before diving into this in detail, look over the general structure of the Standard Model on page 139.

3.4.2 The Gauge and Higgs Sector

We begin our exposition with the **Electroweak** part of the Standard Model gauge group, the $SU(2) \otimes U(1)$ part, as well as the Higgs.

Beginning with the Higgs, a scalar field in the $(2, -1/2)$ representation of $SU(2) \otimes U(1)$, the first step is to write down the covariant derivative as in (3.51). We denote the generators of the **2** representation of $SU(2)$ as $T_2^a = \frac{1}{2}\sigma^a$ (the Pauli matrices) and the gauge fields as A_μ^a . The generator of $U(1)$ is $Y = C \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ where C is the hypercharge ($-1/2$ in this case), and the $U(1)$ gauge field is B_μ . So, the covariant derivative is

$$(D_\mu \phi)_i = \partial_\mu \phi_i - i[g_2 A_\mu^a T_2^a + g_1 B_\mu Y]_{ij} \phi_j \quad (3.56)$$

where g_1 and g_2 are coupling constants for the $U(1)$ part and the $SU(2)$ part, respectively. If the reason we wrote it down this way isn't clear, compare this expression to equation (3.51), and remember that we are saying the field carries *two* charges; one for $SU(2)$ and one for $U(1)$. Therefore, it has two terms in its covariant derivative. And, as usual, μ is a spacetime index.

Knowing that the generators of $SU(2)$ are the Pauli matrices, we can expand the second part of the covariant derivative in matrix form,

$$\begin{aligned} g_2 A_\mu^a T_2^a + g_1 B_\mu Y &= \frac{g_2}{2} (A_\mu^1 \sigma^1 + A_\mu^2 \sigma^2 + A_\mu^3 \sigma^3) - \frac{g_1}{2} B_\mu \mathbb{I}^{2 \times 2} \\ &= \frac{1}{2} \begin{pmatrix} g_2 A_\mu^3 - g_1 B_\mu & g_2 (A_\mu^1 - i A_\mu^2) \\ g_2 (A_\mu^1 + i A_\mu^2) & -g_2 A_\mu^3 - g_1 B_\mu \end{pmatrix} \end{aligned}$$

So, the full covariant derivative is

$$(D_\mu \phi)_i \doteq \begin{pmatrix} D_\mu \phi_1 \\ D_\mu \phi_2 \end{pmatrix} = \begin{pmatrix} \partial_\mu \phi_1 + \frac{i}{2} (g_2 A_\mu^3 - g_1 B_\mu) \phi_1 + \frac{ig_2}{2} (A_\mu^1 - i A_\mu^2) \phi_2 \\ \partial_\mu \phi_2 + \frac{ig_2}{2} (A_\mu^1 + i A_\mu^2) \phi_1 - \frac{i}{2} (g_2 A_\mu^3 + g_1 B_\mu) \phi_2 \end{pmatrix} \quad (3.57)$$

Now, we know that the Lagrangian will have the kinetic term and some potential:

$$\mathcal{L} = -\frac{1}{2} D_\mu \phi_i^\dagger D^\mu \phi_i - V(\phi^\dagger, \phi) \quad (3.58)$$

Let's assume that the potential has a similar form as equation (3.44) (we add the factors of one-half here for the sake of convention; they don't amount to anything other than a rescaling of λ and Φ),

$$V(\phi^\dagger, \phi) = \frac{1}{4} \lambda \left(\phi^\dagger \phi - \frac{1}{2} \Phi^2 \right)^2 \quad (3.59)$$

Clearly the minimum field configuration is not at $\phi = 0$, but at $|\phi| = \frac{v}{\sqrt{2}}$. So, following what we did in section 3.3.7, we make a global $SU(2)$ transformation to put the entire

VEV on the *first* component of ϕ , and then make a global $U(1)$ transformation to make the field real. So,

$$\langle 0|\phi|0\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} v \\ 0 \end{pmatrix} \quad (3.60)$$

and we expand ϕ around this new vacuum:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} v + h(x) \\ 0 \end{pmatrix} \quad (3.61)$$

Remember that we have chosen our $SU(2)$ to keep the second component 0 and our $U(1)$ to keep the first component real. So, $h(x)$ is a real scalar field.

Clearly, plugging this into the covariant derivative (3.57) will give the exact same expression as before, but with ϕ_1 replaced by $\frac{1}{\sqrt{2}}h(x)$ and ϕ_2 replaced by 0, plus an extra term for v . When we plug this extra term into the kinetic term in the Lagrangian (3.58), we get that it is

$$-\frac{1}{8}v^2 \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} g_2 A_\mu^3 - g_1 B_\mu & g_2(A_\mu^1 - iA_\mu^2) \\ g_2(A_\mu^1 + iA_\mu^2) & -g_2 A_\mu^3 - g_1 B_\mu \end{pmatrix} \begin{pmatrix} g_2 A^{3\mu} - g_1 B^\mu & g_2(A^{1\mu} - iA^{2\mu}) \\ g_2(A^{1\mu} + iA^{2\mu}) & -g_2 A^{3\mu} - g_1 B^\mu \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad (3.62)$$

Before multiplying this out, we employ a trick. Define the **Weak Mixing Angle**

$$\theta_w \equiv \tan^{-1} \left(\frac{g_1}{g_2} \right)$$

and the shorthand notation

$$s_w \equiv \sin \theta_w \quad \text{and} \quad c_w \equiv \cos \theta_w$$

And finally, we can define four new gauge fields as linear combinations of the four we have been using:

$$W_\mu^+ \equiv \frac{1}{\sqrt{2}}(A_\mu^1 - iA_\mu^2) \quad (3.63)$$

$$W_\mu^- \equiv \frac{1}{\sqrt{2}}(A_\mu^1 + iA_\mu^2) \quad (3.64)$$

$$Z_\mu \equiv c_w A_\mu^3 - s_w B_\mu \quad (3.65)$$

$$A_\mu \equiv s_w A_\mu^3 + c_w B_\mu \quad (3.66)$$

These can easily be inverted to give the old fields in terms of the new fields,

$$A_\mu^1 = \frac{1}{\sqrt{2}}(W_\mu^+ + W_\mu^-) \quad (3.67)$$

$$A_\mu^2 = \frac{i}{\sqrt{2}}(W_\mu^+ - W_\mu^-) \quad (3.68)$$

$$A_\mu^3 = c_w Z_\mu + s_w A_\mu \quad (3.69)$$

$$B_\mu = -s_w Z_\mu + c_w A_\mu \quad (3.70)$$

We make a few observations about these fields before moving on. First of all, they are merely linear combinations of the gauge fields introduced in equation (3.56). Second, notice that the two fields W_μ^\pm are both linear combinations of fields corresponding to non-Cartan generators of $SU(2)$, whereas Z_μ and A_μ are both linear combinations of fields corresponding to Cartan generators of $SU(2)$ and $U(1)$. So, according to our discussion in section 2.2.16, we expect that Z_μ and A_μ will interact but not change the charge, and that W_μ^\pm will interact and change the charge. Incidentally, notice that W_μ^\pm has the exact form of the raising and lowering operators defined in (2.18).

With these fields defined, we can now rewrite (3.62) as

$$-\frac{1}{8}g_2^2v^2 \begin{pmatrix} 1 & 0 \\ \frac{1}{\sqrt{2}}Z_\mu & \star \end{pmatrix} \begin{pmatrix} \sqrt{2}W_\mu^+ \\ \sqrt{2}W_\mu^- \end{pmatrix}^2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = -M_w^2 W^{+\mu} W_\mu^- - \frac{1}{2}M_Z^2 Z^\mu Z_\mu$$

(the \star is there because that matrix element will always be multiplied by 0, so we don't bother writing it), where we have defined

$$M_w = \frac{g_2 v}{2} \quad \text{and} \quad M_Z = \frac{M_w}{c_w} = \frac{g_2 v}{2c_w}$$

So, we see that, by symmetry breaking, we have given mass to the W_μ^+ , the W_μ^- , and the Z_μ fields. However, the A_μ has not gained mass.

These particles are the W and Z vector bosons, which are the force carrying particles of the **Weak Force**. Each of these particles has an extremely large mass ($M_W \approx 80.4$ GeV, and $M_Z \approx 91.2$ GeV), which explains why they only act over a very short range ($\approx 10^{-18}$ meters).

Also note that the A_μ remains massless, implying that it did not acquire a VEV, and because it is a single generator, we see that a single $U(1)$ remains unbroken. This $U(1)$ and A_μ are the gauge group and field of **Electromagnetism**, as discussed in section 3.1.11.

The idea of all of this is that at very high energies (above the breaking of the $SU(2) \otimes U(1)$), we have only a Higgs complex scalar field, along with four identical massless vector boson gauge fields ($A_\mu^1, A_\mu^2, A_\mu^3, B_\mu$), each of which behave basically like a photon. At low energies, however, the $SU(2) \otimes U(1)$ symmetry of the Higgs is broken, and the low energy effective theory consists of a linear combination of the original four fields. Three of those linear combinations have gained mass, and one remains massless, retaining the photon-like properties from before symmetry breaking. The theory above the symmetry breaking scale is called the **Electroweak Theory** (with four photon-like force carrying particles), whereas below the breaking scale they become two separate forces; the **Weak** and the **Electromagnetic**. This is the first and most basic example of unification we have in our universe. At low energies, the electromagnetic and weak forces are separate. At high energies, they unify into a single theory that is described by $SU(2) \otimes U(1)$.

We can express the new fields as simple Euler rotations of the old fields:

$$\begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} A_\mu^3 \cos \theta_w - B_\mu \sin \theta_w \\ A_\mu^3 \sin \theta_w + B_\mu \cos \theta_w \end{pmatrix} \Rightarrow \begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = R(\theta_w) \begin{pmatrix} A_\mu^3 \\ B_\mu \end{pmatrix}$$

So, the Z_μ is a massive linear combination of the A_μ^3 and B_μ , while the photon A_μ is a massless linear combination of the two.

We can do the same type of analysis for the W_μ^\pm , where they are both massive linear combinations of A_μ^1 and A_μ^2 . The Z_μ and A_μ are both made up of a mixture of the $SU(2)$ and $U(1)$ gauge groups, whereas the W_μ^\pm come solely from the $SU(2)$ part.

Before moving on to include leptons (and then hadrons), we first write out the full Lagrangian for the effective field theory for $h(x)$ and the gauge fields.

We start with the complete Lagrangian term for $h(x)$. We have written the original field ϕ as in equation (3.61). So, our potential in equation (3.59) is now

$$V(\phi^\dagger, \phi) = \frac{1}{4}\lambda(\phi^\dagger\phi - \frac{1}{2}v^2)^2 = \dots = \frac{1}{4}\lambda v^2 h^2 + \frac{1}{4}\lambda v h^3 + \frac{1}{16}\lambda h^4$$

The first term on the right hand side is clearly a mass term giving the mass of the Higgs ($= \sqrt{\frac{\lambda}{2}}v$), and the second two terms are interaction vertices. The kinetic term for the Higgs will be the usual $-\frac{1}{2}\partial_\mu h \partial^\mu h$.

Now, following loosely what we did in section 3.1.11, we want to find kinetic terms for the gauge fields. We start by finding them for the original gauge fields before symmetry breaking ($A_\mu^1, A_\mu^2, A_\mu^3$ and B_μ). Using (3.55), (3.48), and (3.50), and the $SU(2)$ structure constants given in equation (2.17), we have

$$\begin{aligned} F_{\mu\nu}^1 &= 2 \text{Tr} (F_{\mu\nu} T^1) \\ &= 2 \text{Tr} ((\partial_\mu A_\nu - \partial_\nu A_\mu - ig_2[A_\mu, A_\nu])T^1) \\ &= 2 \text{Tr} ((\partial_\mu A_\nu^a T^a - \partial_\nu A_\mu^a T^a - ig_2 A_\mu^a A_\nu^b [T^a, T^b])T^1) \\ &= 2 \text{Tr} (\partial_\mu A_\nu^a T^a T^1 - \partial_\nu A_\mu^a T^a T^1 - ig_2 A_\mu^a A_\nu^b i f^{abd} T^c T^1) \\ &= \partial_\mu A_\nu^a \delta^{a1} - \partial_\nu A_\mu^a \delta^{a1} + g_2 A_\mu^a A_\nu^b f^{abc} \delta^{c1} \\ &= \partial_\mu A_\nu^1 - \partial_\nu A_\mu^1 + g_2 A_\mu^a A_\nu^b f^{ab1} \\ &= \partial_\mu A_\nu^1 - \partial_\nu A_\mu^1 + g_2 A_\mu^a A_\nu^b \epsilon^{ab1} \\ &= \partial_\mu A_\nu^1 - \partial_\nu A_\mu^1 + g_2 (A_\mu^2 A_\nu^3 - A_\nu^2 A_\mu^3) \end{aligned}$$

And similarly,

$$\begin{aligned} F_{\mu\nu}^2 &= \partial_\mu A_\nu^2 - \partial_\nu A_\mu^2 + g_2 (A_\mu^3 A_\nu^1 - A_\nu^3 A_\mu^1) \\ F_{\mu\nu}^3 &= \partial_\mu A_\nu^3 - \partial_\nu A_\mu^3 + g_2 (A_\mu^1 A_\nu^2 - A_\nu^1 A_\mu^2) \end{aligned}$$

And the gauge field corresponding to the $U(1)$ will be defined as in (3.20):

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$$

So, we can now write the kinetic term for our fields according to equation (3.54):

$$\mathcal{L}_{Kin} = -\frac{1}{4}F_a^{\mu\nu}F_{\mu\nu}^a - \frac{1}{4}B^{\mu\nu}B_{\mu\nu}$$

We can then use (3.67–3.70) to translate these kinetic terms into the new fields. We will spare the extremely tedious detail and skip right to the Lagrangian:

$$\begin{aligned}
\mathcal{L}_{eff} &= -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} - \frac{1}{4}Z^{\mu\nu}Z_{\mu\nu} - D^{\dagger\mu}W^{-\nu}D_{\mu}W_{\nu}^{+} + D^{\dagger\mu}W^{-\nu}D_{\nu}W_{\mu}^{+} \\
&\quad + ie(F^{\mu\nu} + \cot\theta_w Z^{\mu\nu})W_{\mu}^{+}W_{\nu}^{-} \\
&\quad - \frac{1}{2}\left(\frac{e^2}{\sin^2\theta_w}\right)(W^{+\mu}W_{\mu}^{-}W^{+\nu}W_{\nu}^{-} - W^{+\mu}W_{\mu}^{+}W^{-\nu}W_{\nu}^{-}) \\
&\quad - (M_W^2 W^{+\mu}W_{\mu}^{-} + \frac{1}{2}M_Z^2 Z^{\mu}Z_{\mu})\left(1 + \frac{h}{v}\right)^2 \\
&= -\frac{1}{2}\partial^{\mu}h\partial_{\mu}h - \frac{1}{2}m_h^2 h^2 - \frac{1}{2}\frac{m_h^2}{v}h^3 - \frac{1}{8}\frac{m_h^2}{v^2}h^4
\end{aligned}$$

where we have chosen the following definitions:

$$\begin{aligned}
F_{\mu\nu} &= \partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu} && \text{(Electromagnetic Field Strength)} \\
Z_{\mu\nu} &= \partial_{\mu}Z_{\nu} - \partial_{\nu}Z_{\mu} && \text{(Kinetic term for } Z_{\mu}) \\
D_{\mu} &= \partial_{\mu} - ie(A_{\mu} + \cot\theta_w Z_{\mu})
\end{aligned}$$

and the rest of the terms were defined previously in this section.

3.4.3 The Lepton Sector

We now turn to the lepton sector (which is still in the $SU(2) \otimes U(1)$ part of the Standard Model gauge group). A **Lepton** is a spin-1/2 particle that does *not* interact with the $SU(3)$ color group (the strong force). There are six **Flavors** of leptons arranged into three **Families**, or **Generations**. The table on page 139 explains this. The first generation consists of the electron (e) and the electron neutrino (ν_e), the second generation the muon (μ) and the muon neutrino (ν_{μ}), and the third the tau (τ) and tau neutrino (ν_{τ}). Each family behaves exactly the same way, so we will only discuss one generation in this section (e and ν_e). To incorporate the physics of the other families, merely change the e to either a μ or τ , and the ν_e to a ν_{μ} or ν_{τ} in the following notes.

What we will see is that, in a sense, the neutrinos don't really interact with anything on their own (which is why they are incredibly difficult to detect). For this reason, neutrinos don't have their own place in a representation of $SU(3) \otimes SU(2) \otimes U(1)$ (see table on 139). Electrons on the other hand, do interact with other things on their own, and we therefore see them in the $(1, 1)$ representation.

However, the neutrino *does* interact with other things *as part of an* $SU(2)$ doublet *with* the electron,

$$l = \begin{pmatrix} \nu_e \\ e \end{pmatrix} \tag{3.71}$$

This is why it is arranged as it is on page 139 with the electron under the $(\mathbf{2}, -1/2)$ representation of $SU(2) \otimes U(1)$.

This may seem confusing, but we hope the following will make it clear. We will proceed in what we believe is the clearest way to see this (primarily following [32]). We start with 2 fields, \bar{e} and l , where \bar{e} is a single left-handed Weyl field (see section 3.2.4), and l is defined in (3.71). As we have said, l is in the $(\mathbf{2}, -1/2)$ representation, \bar{e} is in the $(\mathbf{1}, 1)$ representation, and ν_e has no representation of its own.

So, mimicking what we did in equation (3.56) in the previous section, we can write down the covariant derivative for each field,

$$(D_\mu l)_i = \partial_\mu l_i - ig_2 A_\mu^a (T^a)_{ij} l_j - ig_1 B_\mu Y_l l_i \quad (3.72)$$

$$D_\mu \bar{e} = \partial_\mu \bar{e} - ig_1 B_\mu Y_{\bar{e}} \bar{e} \quad (3.73)$$

The field \bar{e} has no $SU(2)$ term in its covariant derivative because the $\mathbf{1}$ representation of $SU(2)$ is the trivial representation - this means it doesn't carry $SU(2)$ charge. Also, we know that

$$Y_l = -\frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (3.74)$$

and

$$Y_{\bar{e}} = (1) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (3.75)$$

Following the Lagrangian for the spin-1/2 fields we wrote out in equation (3.11), we can write out the kinetic term for both (massless) fields:

$$\mathcal{L}_{Kin} = i l^\dagger \bar{\sigma}^\mu (D_\mu l)_i + i \bar{e}^\dagger \bar{\sigma}^\mu D_\mu \bar{e} \quad (3.76)$$

At the end of section 3.2.2 and of section 3.2.11, we briefly discussed the idea of renormalization. We said that certain theories can be renormalized and others cannot. It turns out (for reasons beyond the scope of these notes) that while the theory we have outlined so far is renormalizable, if we try to add mass terms for and l and \bar{e} fields, the theory breaks down. Therefore we cannot add a mass term. But, we know experimentally that electrons and neutrinos have mass, so obviously something is wrong. We must incorporate mass into the theory, but in a more subtle way than merely adding a mass term. It turns out that we can use the Higgs mechanism as follows.

While adding mass terms renders the theory inconsistent, we can add a Yukawa term (cf. equation (3.42)),

$$\mathcal{L}_{Yuk} = -y \epsilon^{ij} \phi_i l_j \bar{e} + \text{h.c.}$$

where y is another coupling constant, ϵ^{ij} is the totally antisymmetric tensor, and h.c. is the Hermitian Conjugate of the first term.

Now that we have added \mathcal{L}_{Yuk} to the Lagrangian, we want to break the symmetry exactly as we did in the previous section. First, we replace ϕ_1 with $\frac{1}{\sqrt{2}}(v + h(x))$ and ϕ_2 with 0, exactly as we did in equation (3.61). So,

$$\begin{aligned}
\mathcal{L}_{Yuk} &= -y\epsilon^{ij}\phi_i l_j \bar{e} + \text{h.c.} \\
&= -y(\phi_1 l_2 - \phi_2 l_1) \bar{e} + \text{h.c.} \\
&= -\frac{1}{\sqrt{2}}y(v + h)l_2 \bar{e} + \text{h.c.} \\
&= -\frac{1}{\sqrt{2}}y(v + h)e\bar{e} - \frac{1}{\sqrt{2}}y(v + h)\bar{e}e \\
&= -\frac{1}{\sqrt{2}}y(v + h)\bar{\mathcal{E}}\mathcal{E}
\end{aligned} \tag{3.77}$$

where $\mathcal{E} = \begin{pmatrix} e \\ \bar{e}^\dagger \end{pmatrix}$ is the Dirac field for the electron (e is the electron and \bar{e}^\dagger is the anti-electron, or positron). Comparing (3.77) with (3.12), we see that it is a mass term for the electron and positron.

Now we want a kinetic term for the neutrino. It is believed that neutrinos are described by Majorana fields (see section 3.2.4), so we begin with the field $\mathcal{N}' = \begin{pmatrix} \nu_e \\ \nu_e^\dagger \end{pmatrix}$. Now, we employ a trick. Referring back to equations (3.33) and (3.34), the kinetic term for Majorana fields has only one term (because Majorana fields have only one Weyl spinor), whereas the Dirac field sums over both Weyl spinors composing it. So, instead of working with the Majorana field \mathcal{N}' , we can instead work with the Dirac field

$$\mathcal{N} = \begin{pmatrix} \nu_e \\ 0 \end{pmatrix}$$

So, the Dirac kinetic term $i\bar{\mathcal{N}}\gamma^\mu\partial_\mu\mathcal{N}$ will clearly result in the correct kinetic term from (3.76), or $i\nu^\dagger\bar{\sigma}^\mu\partial_\mu\nu$.

Now, continuing with the symmetry breaking, we want to write the covariant derivative (3.72) and (3.73) in terms of our low energy gauge fields (3.63–3.66).

We said in the previous section (which echoed our discussion in section 2.2.16) that the gauge fields corresponding to Cartan generators (A_μ and Z_μ) act as force carrying particles, but do not change the charge of the particles they interact with. On the other hand, the non-Cartan generators' gauge fields (W_μ^\pm) are force carrying particles which *do* change the charge of the particle they interact with. Therefore, to make calculations simpler, we will break the covariant derivative up into the non-Cartan part and the Cartan part.

The non-Cartan part of the covariant derivative (3.72) is

$$\begin{aligned}
g_2(A_\mu^1 T^1 + A_\mu^2 T^2) &= \frac{1}{2} g_2 \left(A_\mu^1 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + A_\mu^2 \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \right) \\
&= \frac{1}{2} g_2 \begin{pmatrix} 0 & A_\mu^1 - i A_\mu^2 \\ A_\mu^1 + i A_\mu^2 & 0 \end{pmatrix} \\
&= \frac{g_2}{\sqrt{2}} \begin{pmatrix} 0 & W_\mu^+ \\ W_\mu^- & 0 \end{pmatrix}
\end{aligned}$$

and the Cartan part is

$$\begin{aligned}
g_2 A_\mu^3 T^3 + g_1 B_\mu Y &= \frac{e}{s_w} (s_w A_\mu + c_w Z_\mu) T^3 + \frac{e}{c_w} (c_w A_\mu - s_w Z_\mu) Y \\
&= e (A_\mu + \cot \theta_w Z_\mu) T^3 + e (A_\mu - \tan \theta_w Z_\mu) Y \\
&= e (T^3 + Y) A_\mu + e (\cot \theta_w T^3 - \tan \theta_w Y) Z_\mu
\end{aligned}$$

We have noted before that A_μ is the photon, or the electromagnetic field, and e is the electromagnetic charge. Therefore, the linear combination $T^3 + Y$ must be the generator of electric charge. Notice that the electromagnetic generator is in a linear combination of the two Cartan generators of $SU(2) \otimes U(1)$.

We know that $T^3 = \frac{1}{2} \sigma^3$, and Y_l and $Y_{\bar{e}}$ are defined in equations (3.74) and (3.75), so we can write

$$\begin{aligned}
T^3 l &= \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \nu_e \\ e \end{pmatrix} = \frac{1}{2} \begin{pmatrix} \nu_e \\ -e \end{pmatrix} \\
Y_l l &= -\frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \nu_e \\ e \end{pmatrix} = -\frac{1}{2} \begin{pmatrix} \nu_e \\ e \end{pmatrix}
\end{aligned}$$

And we know that \bar{e} carries no T^3 charge, so its T^3 eigenvalue is 0, while $Y_{\bar{e}}$ is +1. So, summarizing all of this,

$$\begin{aligned}
T^3 \nu_e &= +\frac{1}{2} \nu_e & T^3 e &= -\frac{1}{2} e & T^3 \bar{e} &= 0 \\
Y \nu_e &= -\frac{1}{2} \nu_e & Y e &= -\frac{1}{2} e & Y \bar{e} &= +\bar{e}
\end{aligned}$$

Then defining the generator of electric charge to be $Q \equiv T^3 + Y$, we have

$$Q \nu_e = 0 \quad Q e = -e \quad Q \bar{e} = +\bar{e}$$

So the neutrino ν_e has no electric charge, the electron e has negative electric charge, and the antielectron, or positron, has plus one electric charge—all exactly what we would expect.

We can now take all of the terms we have discussed so far and write out a complete Lagrangian. However, doing so is both tedious and unnecessary for our purposes.

The primary idea is that electrons/positrons and neutrinos all interact with the $SU(2) \otimes U(1)$ gauge particles, the W^\pm , Z_μ , and A_μ . The Z_μ and A_μ (the Cartan gauge particles) interact but do not affect the charge. On the other hand, the W^\pm act as $SU(2)$ raising and lowering operators (as can easily be seen by comparing (3.63) and (3.64) to equation (2.18)). The $SU(2)$ doublet state acted on by these raising and lowering operators is the doublet in equation (3.71). The W^+ interacts with a left-handed electron, raising its electric charge from minus one to zero, turning it into a neutrino. However W^+ does not interact with left-handed neutrinos. On the other hand, W^- will lower the electric charge of a neutrino, making it an electron. But W^- will not interact with an electron.³

3.4.4 The Quark Sector

A Quark is a spin-1/2 particle that interacts with the $SU(3)$ color force. Just as with leptons, there are six flavors of quarks, arranged in three families or generations (see the table on page 139).

Following very closely what we did with the leptons, we work with only one generation. Extending to the other generators is then trivial. To begin, define three fields: q , \bar{u} , and \bar{d} , in the representations $(\mathbf{3}, \mathbf{2}, +1/6)$, $(\bar{\mathbf{3}}, 1, -2/3)$, and $(\bar{\mathbf{3}}, 1, +1/2)$ of $SU(3) \otimes SU(2) \otimes U(1)$. The field q will be the $SU(2)$ doublet

$$q = \begin{pmatrix} u \\ d \end{pmatrix} \quad (3.78)$$

This is exactly analogous to equation (3.71).

Again, following what we did with the leptons, we can write out the covariant derivative for all three fields:

$$(D_\mu q)_{\alpha i} = \partial_\mu q_{\alpha i} - ig_3 A_\mu^a (T_2^a)_{\alpha}^{\beta} q_{\beta i} - ig_2 A_\mu^a (T_2^a)_i^j q_{\beta j} - ig_1 \left(\frac{1}{6} \right) B_\mu q_{\alpha i} \quad (3.79)$$

$$(D_\mu \bar{u})^\alpha = \partial_\mu \bar{u}^\alpha - ig_3 A_\mu^a (T_3^a)_{\beta}^{\alpha} \bar{u}^\beta - ig_1 \left(-\frac{2}{3} \right) B_\mu \bar{u}^\alpha \quad (3.80)$$

$$(D_\mu \bar{d})^\alpha = \partial_\mu \bar{d}^\alpha - ig_3 A_\mu^a (T_3^a)_{\beta}^{\alpha} \bar{d}^\beta - ig_1 \left(\frac{1}{3} \right) B_\mu \bar{d}^\alpha \quad (3.81)$$

where i is an $SU(2)$ index and α is an $SU(3)$ index. The $SU(3)$ index is lowered for the $\mathbf{3}$ representation and raised for the $\bar{\mathbf{3}}$ representation.

Just as with leptons, we cannot write down a mass term for these particles, but we can include a Yukawa term coupling these fields to the Higgs:

$$\mathcal{L}_{Yuk} = -y' \epsilon^{ij} \phi_i q_{\alpha j} \bar{d}^\alpha - y'' \phi^\dagger_i q_{\alpha i} \bar{u}^\alpha + \text{h.c.}$$

³This does not mean that no vertex in the Feynman diagrams will include a W^- and an electron field, but rather that if you collide an electron and a W^- , there will be no interaction

As with the leptons, we can break the symmetry according to equation (3.61), and writing out this Yukawa term, we get

$$\begin{aligned}\mathcal{L}_{Yuk} &= -\frac{1}{\sqrt{2}}y'(v+h)(d_\alpha\bar{d}^\alpha + \bar{d}_\alpha^\dagger d^{\dagger\alpha}) - \frac{1}{\sqrt{2}}y''(v+h)(u_\alpha\bar{u}^\alpha + \bar{u}_\alpha^\dagger u^{\dagger\alpha}) \\ &= -\frac{1}{\sqrt{2}}y'(v+h)\bar{\mathcal{D}}^\alpha\mathcal{D}_\alpha - \frac{1}{\sqrt{2}}y''(v+h)\bar{\mathcal{U}}^\alpha\mathcal{U}_\alpha\end{aligned}$$

where we have defined the Dirac fields for the up and down quarks:

$$\mathcal{D}_\alpha \equiv \begin{pmatrix} d_\alpha \\ \bar{d}_\alpha^\dagger \end{pmatrix} \quad \mathcal{U}_\alpha \equiv \begin{pmatrix} u_\alpha \\ \bar{u}_\alpha^\dagger \end{pmatrix}$$

Notice that, whereas both the up and down quarks were massless before breaking, they have now acquired masses

$$m_d = \frac{y'v}{\sqrt{2}} \quad m_u = \frac{y''v}{\sqrt{2}}$$

Writing out the non-Cartan and Cartan parts of the covariant derivatives in terms of the lower energy $SU(2) \otimes U(1)$ gauge fields, we get

$$\begin{aligned}g_2 A_\mu^1 T^1 + g_2 A_\mu^2 T^2 &= \frac{g_2}{\sqrt{2}} \begin{pmatrix} 0 & W_\mu^+ \\ W_\mu^- & 0 \end{pmatrix} \\ g_2 A_\mu^3 T^3 + g_1 B_\mu Y &= eQ A_\mu + \frac{e}{s_w c_w} (T^3 - s_w^2 Q) Z_\mu\end{aligned}$$

And it is again straightforward to find the electric charge eigenvalue for each field:

$$Qu = +\frac{2}{3}u \quad Qd = -\frac{1}{3}d \quad Q\bar{u} = -\frac{2}{3}\bar{u} \quad Q\bar{d} = +\frac{1}{3}\bar{d}$$

Again, we can collect all of these terms and write out a complete Lagrangian. But, doing so is extremely tedious and unnecessary for our purposes.

The primary idea to take away is that the $SU(2)$ doublet (3.78) behaves exactly as the lepton doublet in (3.71) when interacting with the “raising” and “lowering” gauge particles W^\pm . This is why the u and d are arranged in the $SU(2)$ doublet q in (3.78), and why q carries the $SU(2)$ index i in the covariant derivative (3.79), whereas \bar{u} and \bar{d} carry only the $SU(3)$ index.

The $SU(3)$ index runs from 1 to 3, and the 3 values are conventionally denoted *red*, *green*, and *blue* (r, g, b). These obviously are merely labels and have nothing to do with the colors in the visible spectrum.

The eight gauge fields associated with the eight $SU(3)$ generators are called **Gluons**, and they are represented by the matrices in (2.48). We label each gluon as follows:

$$g_{\alpha}^{\beta} \doteq \begin{pmatrix} r\bar{r} & r\bar{g} & r\bar{b} \\ g\bar{r} & g\bar{g} & g\bar{b} \\ b\bar{r} & b\bar{g} & b\bar{b} \end{pmatrix}$$

so that the upper index is the anti-color index, and denotes the column of the matrix, and the lower index is the color index denoting the row of the matrix. Then, from (2.48), consider the gluon

$$g_r^{\bar{g}} \propto \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

and the quarks

$$q_r = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad q_g = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad q_b = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

It is easy to see that this gluon will interact as

$$g_r^{\bar{g}} q_r = 0 \quad g_r^{\bar{g}} q_g = q_r \quad g_r^{\bar{g}} q_b = 0$$

Or in other words, the gluon with the anti-green index will only interact with a green quark. There will be no interaction with the other quarks. Multiplying this out, and looking more closely at the behavior of the $SU(3)$ generators and eigentstates as discussed in sections 2.2.12–2.2.15, you can work out all of the interaction rules between quarks and gluons. You will see that they behave exactly according to the root space of $SU(3)$.

3.5 References and Further Reading

The primary source for these notes is [32], which is an exceptionally clear introduction to Quantum Field Theory. We also used a great deal of material from [3], [24], [29], and [35], all of which are outstanding QFT texts. The derivation of the Dirac equation came from [21], which is written mostly above the scope of these notes, but is an excellent survey of some of the mathematical ideas of Non-Perturbative QFT and Gauge Theory.

The sections on the Standard Model come almost entirely from [32] with little change, in that Srednicki's exposition could hardly be improved upon for the scope of these notes.

For further reading, we also recommend [1], [11], [23], [26], and [27].

4 The Standard Model — A Summary

4.1 How Does All of This Relate to Real Life?

In the fifth century B.C., a Greek named Empedocles took the ideas of several others before him and combined them to say that matter is made up of earth, wind, fire, and water, and that there are two forces, Love and Strife, that govern the way they grow and act. More scientifically, he was saying that matter is made of smaller substances that interact with each other through repulsion and attraction. Democritus, a contemporary of Empedocles, went a step further to say that all matter is made of fundamental particles that are indestructible. He called these particles atoms, meaning “indivisible”⁴ [14].

The field of particle physics seeks to continue studying these same concepts. Are there fundamental, indivisible particles and if so, what are they? How do they behave? How do they group together to form the matter that we see? How do they interact with each other?

The current answer to these questions is called the Standard Model, the theory we spent this paper developing. We have now spent more than one hundred pages expositing a series of mathematical tricks for various types of “fields”. In doing so, we talked about “massless scalars with $U(1)$ charge”, and about things “in a $j = \frac{1}{2}$ representation of $SU(2)$ ”. But one could easily be left wondering how exactly this relates to the things we see in nature. We only discussed 25 particles in the previous section and in the table on page 139 (particles and antiparticles), but you are likely aware that there are hundreds of particles in nature. What about those? How does the mathematical framework detailed so far form the building blocks for the universe?

While we wish to reiterate that the primary purpose of these notes is to provide the mathematical tools with which particle physics is done, and not to outline the phenomenological details of the theory, we are physicists still—not mathematicians. Therefore, before concluding this paper, we will take a brief hiatus from the mathematical rigor and look at a qualitative summary of particle physics.

Throughout this section, the footnotes will provide brief explanations of the analogous mathematical ideas from above. This section⁵ can be read with or without paying attention to them. We provide them merely for those curious.

⁴Of course, our modern use of the word is different. At their discovery, it was thought that different elements were the indivisible particles sought for, so the name atom seemed appropriate

⁵Nearly everything in this section is adapted from [4], including the tables on page 134

4.2 The Fundamental Forces

The two forces most familiar to people are **Gravity** and **Electromagnetism**. Just the act of standing on the ground or sitting in a chair makes use of both, and every “Physics I” student has drawn a free body diagram with a gravitational force going down and a normal force (caused by the electromagnetic repulsion between the two objects) going up. However, these two are only half of the four fundamental forces in our universe (that we know of).

We can think about the third by first considering a compact nucleus which we know to be made of protons and neutrons. From electromagnetism we know that the protons should repel each other because of their like charge. But the nuclei of atoms somehow hold together, which is evidence for some stronger force that causes these particles to attract. This force, which overcomes the electromagnetic repulsions and allows atomic nuclei to remain stable, is called the **Strong** force.⁶ Just as electrically-charged particles are subject to the electromagnetic force, some particles have a property similar to charge, called **Color**, and are subject to the strong force. The field theory that describes this is called **Quantum Chromodynamics** (QCD)⁷ and was first proposed in 1965 by Han, Nambu, and Greenberg [20]. This theory predicts the existence of the gluon, which is the mediator of the strong force between two matter particles.

The fourth force is the one we have the least familiarity with. It is responsible for certain types of radioactive decays; for example, permitting a proton to turn into a neutron and vice versa. It is called the **Weak** force.⁸

In the 1960’s, Sheldon Glashow, Abdus Salam, and Steven Weinberg independently developed a gauge-invariant theory that unified the electromagnetic and weak force [20]. At sufficiently high energies it is observed that the difference between these two separate forces is negligible and that they instead act together as the **Electroweak** force.⁹ For processes at lower energy scales, the symmetry between the electromagnetic and the weak force is broken and we observe two different forces with different properties. Similar to QCD, electroweak theory predicts four force-carrier particles¹⁰ that mediate the force between matter particles. The mediating particle for electromagnetism is the neutral photon, and those for the weak force are the W^+ (with +1 electron charge), W^- (with -1 electron charge) and Z^0 (neutral) bosons.

The electromagnetic, weak, and strong forces described above form what is called the **Standard Model of Particle Physics**. The Standard Model is an incomplete theory in the sense that it fails to describe gravitation, the force that acts on matter. Physicists

⁶The $SU(3)$ color force

⁷Again, the study of the $SU(3)$ color force

⁸The $SU(2)$ part that is left over when $SU(2) \otimes U(1)$ is broken

⁹The unbroken $SU(2) \otimes U(1)$ force

¹⁰Corresponding to the 3 generators of $SU(2)$ and the 1 in $U(1)$. Two of them are Cartan and are, therefore, uncharged, while two are non-Cartan and therefore carry charge

continue to work towards a theory that describes all four fundamental forces, with String Theory currently the most promising. The papers later in this series will discuss these ideas. For the rest of the sections in this review, however, it should be assumed that we are talking about physics under the Standard Model only¹¹ which, despite the shortcoming of not explaining gravity, has tremendous experimental support.

4.3 Categorizing Particles

In the last century, experimenters were surprised as they discovered new particle after new particle. It seemed disorganized and overwhelming that there could be so many elementary objects. Eventually, however, the properties of these particles became better understood and it was found that there really is just a small, finite set of fundamental particles, some of which can be grouped together to make up larger objects. In the next two sections, we will introduce the elementary particles and then will discuss the types of composite particles.

One property of the "zoo" of discovered particles that helps in our organizing them is their intrinsic spin.¹² Any particle, elementary or composite, that is of half-integer spin¹³ is a **Fermion**. Those with integer spin are **Bosons**.¹⁴ The spins govern the statistics of a set of such particles, so fermions and bosons may also be defined according to the statistics they obey.

Namely, fermions obey Fermi-Dirac statistics and therefore also obey the Pauli Exclusion Principle. This means that no two identical fermions can be found in the same quantum state at the same time. Furthermore, to accurately display this behavior it is found that the wave function of a system with fermions must be antisymmetric; swapping any two like fermions causes a change in sign of the overall wave function.

Bosons on the other hand obey Bose-Einstein statistics; any number of the same type of particle can be in the same state at the same time. In contrast to fermions, the wavefunction of a system of bosons is symmetric.¹⁵

The Venn diagrams on page 138, the table provided on page 139, and the table below should be referenced as you read through what follows.

¹¹We are not assuming Supersymmetry in this paper, though we will consider Supersymmetry in a later paper

¹²Or in other words, which representation of $SU(2)$ they sit in

¹³Is in the $j = \frac{1}{2}$ or $j = \frac{3}{2}$ representation of $SU(2)$

¹⁴Is in the $j = 0$ or $j = 1$ representation of $SU(2)$

¹⁵Cf. section 3.2.3

4.4 Elementary Particles

The elementary particles are those that are considered fundamental, or in other words, are not composed of smaller particles.¹⁶ They can be divided into two groups: matter particles and non-matter particles.

The elementary matter particles all have half-integer spin (so are fermions) and the elementary non-matter particles all have integer spin (so are bosons). We can then observe that an equivalent grouping is made if we divide the elementary particles instead by their intrinsic spin, which is commonly done. Then an elementary matter particle is the same thing as an elementary fermion, and similarly for the bosons. The two terms are used interchangeably in the discussion below.

4.4.1 Elementary Fermions

The elementary fermions are the building blocks of all other matter. For example, the proton and neutron are made up of different combinations of three elementary quarks. Electrons, which are also elementary, cloud around the protons and neutrons, and when all three group together in a particular way, an atom is formed. Less familiar examples include those that are unstable, such as the muon, which decay into something else fairly quickly.

For every elementary (and sometimes composite) matter particle, there is also a corresponding particle with the same mass but of different charge and magnetic moment.¹⁷ Generally the name of such a particle is the same as the corresponding “normal” matter particle, but with the prefix “anti” in front of it (e.g. antiquark, antilepton, etc.). In this paper, whenever we discuss matter and its properties, it is implied that the antimatter counterparts have similar properties.

Now we further divide the elementary fermions into two groups, quarks and leptons. A convenient way to distinguish these two sets is by whether or not they interact via the strong force: quarks may interact via the strong force, while leptons do not.

Quarks

Experiments involving high energy collisions of electrons and protons led Murray Gell-Mann to suggest in 1964 [25] that protons and neutrons are actually composite particles, made of three point-like, spin-1/2 particles whose charges are either $-1/3$ or $+2/3$ units of electron charge. He called these particles **Quarks**. Through further experiments it has been found that there are six flavors of quarks total, grouped into three generations with

¹⁶These are the ones that are in some representation of $SU(3) \otimes SU(2) \otimes U(1)$ on the table on page 139

¹⁷Cf. material on spin-1/2 particles in section 3.1

the first generation containing the up and down quarks, the second generation containing the more massive charm and strange quarks, and the third generation containing the even more massive top and bottom quarks.

As electrically charged particles are subject to the electromagnetic force, quarks have a property similar to charge, called color, and any colored particle is subject to the strong force. It is found that there are three different types of colors: (defined as) red, green, and blue (plus three more for antiquarks: antired, antigreen, and antiblue). Quarks are grouped together to make composite particles that are colorless (the color charges cancel out), which is why the concept of color was only discovered after quarks themselves were found. The addition of color to the quark model also ensures that any quarks contained in a composite particle will not violate the exclusion principle since each has a different color. Again, QCD is the field theory that describes these properties.

Another interesting feature of quarks is that they are never found alone, but rather always inside of a composite particle. This phenomenon is called **Confinement**.¹⁸ It is more a property of the strong force, which increases in strength as two colored particles are pulled away from each other, just as would happen when the ends of a piece of elastic are pulled apart. We can consider reaching a distance between the two quarks where there is sufficient potential energy built up that it can be converted to matter, creating a quark-antiquark pair. The pair will separate and the resulting particles will recombine with the original quarks. As this process repeats, and more quark-antiquark pairs are created, the end result in the whole process is a multiplication of the number of quarks and of the number of composite particles. In the opposite extreme, as two quarks get closer together, the strong force between them becomes weaker until the quarks move around freely and more independently. This is called **Asymptotic Freedom**.

Quarks also interact with other particles via the weak force, which is the only force that can cause a change of flavor (changing an up into a down, for example). When this happens, a quark either turns into a heavier quark by absorbing a W boson, or it emits a W boson and then decays to a lighter quark. Beta decay, a common radioactive process, is caused by this mechanism. Instead of just thinking of beta decay as a neutron in the nucleus of an atom decaying, or splitting, into a proton, electron, and antineutrino, we can go a step further with our understanding of quarks subject to the weak force. We add that, really, it is one of the down quarks in the neutron that emits a W^- boson and then decays to the lighter up quark, keeping charge conserved in the process.¹⁹ The neutron, which used to have one up and two down quarks, now has one down and two up quarks, which is the composition of a proton. The electron and antineutrino are created from the decay of the W^- boson.

¹⁸We did not discuss confinement in the main body of this paper, though it can be derived from what we did discuss

¹⁹Other conserved quantities are momentum, energy, quark number, lepton number, and (approximately) lepton generation number

Leptons

Leptons interact with other matter via the electromagnetic, the weak, and gravitational forces, but not through the strong force.²⁰ There are three charged leptons, grouped, like the quarks, into three different generations based on their masses.²¹ The electron is the lightest of the charged leptons, then the muon, and the tau. There are also three neutral leptons, called neutrinos (“little neutral one”), one type for each of the charged leptons: the electron neutrino, the muon neutrino, and the tau neutrino.

Some quantities in lepton events are found to be conserved.²² If we define lepton number as the number of leptons minus the number of antileptons, then lepton number is constant in all interactions. Additionally, the lepton number *within each generation* is also approximately conserved. For example, the number of electrons and electron neutrinos minus the number of antielectrons and electron antineutrinos is found to be constant in most particle reactions.

An interesting exception is in neutrino oscillations, where a neutrino changes lepton flavors as it travels. For example, we can take a measurement and observe an electron neutrino, even though it was known to have been created as a muon neutrino. These oscillations of flavor only occur if neutrinos have mass (even just very small mass), so the fact that the Standard Model currently predicts them to be massless demonstrates that there are some parameters in the theory that need to be adjusted.

4.4.2 Elementary Bosons

Throughout the development of the Standard Model it was found that some elementary particles play a different role than the ordinary matter particles that make up the stuff of the universe. Both the gauge bosons and the Higgs boson fall into this group.

Gauge Bosons

In the mathematical formulation of quantum field theory, the Lagrangian can be made invariant under a local gauge transformation by the addition of a vector field called a gauge field. As with the more familiar example of an electron, the quanta of the gauge field is a type of particle, which in this case is called a **Gauge Boson**. There are three types of gauge bosons described by the Standard Model.²³ They are the photon, which

²⁰This means that leptons carry $SU(2) \otimes U(1)$ terms in their covariant derivatives, but not $SU(3)$ terms

²¹This is equivalent to the statement above that there are three copies of the Standard Model Gauge Group - Cf. page 115

²²These conservation laws can all be derived from the rules we discussed above, though they are typically treated separately because they are extremely useful when talking about specific interactions

²³This is equivalent to saying that there are three gauge groups, each with their own set of generators

carries the electromagnetic force, the W and Z bosons, which carry the weak force, and the gluons which carry the strong force. Each of these bosons have been experimentally detected.

Evidence for the neutral photon first came in 1905 when Einstein proposed an explanation of the photoelectric effect, that light was quantized into energy packets [7]. Confirmation of the W^+ , W^- , and Z^0 bosons came in 1983 through proton-proton collisions at the European Organization for Nuclear Research (CERN) [5].

The gluons were first experimentally observed in 1979 in the electron-positron collider at the German Electron Synchrotron (DESY) in Hamburg [5]. Further experiments have demonstrated that the gluons have eight different color states and that, because they interact via the strong force, they have properties similar to quarks, such as confinement.

Taking into account their possible charge or color, we find that there are 12 gauge bosons in all, one for the electromagnetic force, three for the weak force, and eight for the strong force.

The Higgs Boson

The Higgs boson is the only Standard Model particle that has not yet been observed. It is also the only elementary boson that is not a gauge boson. Rather, it is the carrier particle of the scalar Higgs field from which other particles acquire mass. The existence of the Higgs would explain why some particles have mass and others do not. For example, the W and Z bosons are very massive, whereas the photon is massless. One of the main goals of the Large Hadron Collider (LHC), located at CERN in Switzerland, is to provide evidence for the Higgs. It is expected to be in full operation in 2009.

4.5 Composite Particles

Examples of composite particles include hadrons, nuclei, atoms, and molecules. The latter three are well known and will not be described here.

Hadrons are made up of bound quarks and interact via the strong force. They can be either fermions or bosons, depending on the number of quarks that make them up. An odd number of bound quarks create a spin-1/2 or spin-3/2 hadron, which is called a baryon, and an even number of quarks create spin-0 or spin-1 hadrons, called mesons. Experimentally, only combinations of three quarks or two quarks have been found, so the terms baryon and meson often just refer to three or two bound quarks, respectively.

You can understand why mesons and baryons have the spin that they do by considering how many spin-1/2 quarks compose them. A meson has two quarks, and therefore the total spin of a meson is the sum of an even number of half-integer spin particles, which

will be integer spin. And because there are only two of them, it is either spin 0 or 1. Baryons, on the other hand, will have a linear combination of *three* particles with half-integer spin, which will of course be half-integer: 1/2 or 3/2.

The most well known examples of baryons are protons and neutrons. Protons are made of two up quarks and one down quark, or $|uud\rangle$, and neutrons are made of two down and one up, or $|udd\rangle$. The baryons are made of “normal” quarks only and their antimatter counterparts are made of the corresponding antiquarks.

The mesons are made of a quark and an antiquark pair, though not necessarily of the same generation. Examples include the π^+ $|u\bar{d}\rangle$ and K^+ $|u\bar{s}\rangle$.

One of the reasons for the zoo of particles discovered in the past century is because of the numerous possible combinations of six quarks put into a three-quark or two-quark hadron. Additionally, each of these combinations can be in different quantum mechanical states, thereby displaying different properties. For example, a rho meson ρ has the same combination of quarks as a pion π , but the ρ is spin-1 whereas the pion is spin-0.

4.6 Visualizing It All

Finally, we provide a few tables which should help you see all of this more clearly.

Interactions	Acts On	Strength	Range
Strong	Hadrons	1	10^{-15} m
Electromagnetism	Electric Charges	10^{-2}	$\propto (1/r^2)$
Weak	Leptons and Hadrons	10^{-5}	10^{-18} m
Gravity	Mass	10^{-39}	$\propto (1/r^2)$

where the relative strengths have been normalized to unity for the strong force.

Also, the four classes of force-carrying gauge bosons are shown below;²⁴

Interaction	Gauge Boson	Spin	Acts On
Strong	Gluon	1	Hadrons
Electromagnetism	Photon	1	Electric Charges
Weak	W^\pm, Z^\pm	1	Leptons and Hadrons
Gravity	Graviton	2	Mass

²⁴The graviton is a the hypothetical carrying particle for the gravitational force; it is not described by the Standard Model

5 A Look Ahead

Now that we have completed our introduction to basic particle theory, we can begin our uphill climb towards more fundamental concepts. As a preview, notice that everything we have done so far has been an exposition of how gauge theories work. Our investigation into gauge theories has been purely algebraic (working entirely from group theory, as Part II demonstrates). As gauge theory seems to be the correct approach to understanding our universe, everything we do for the remainder of this series will be focused on a more fundamental understanding of gauge theory, culminating in String Theory.

As we just stated, we have been treating gauge theory as a purely algebraic construct. However String Theory, if true, must obviously be able to reproduce the same general framework we have seen so far. But, String Theory is fundamentally a geometric construct. As we will see, String Theory will reproduce literally everything we have seen about gauge theory, but from a geometric framework.

This should not be entirely foreign, though. Recall that, for electromagnetism, the gauge group is $U(1)$. We can “draw” this geometrically as a circle in the complex plane. The Weak force is represented by the gauge group $SU(2)$, which we have seen is parameterized by three numbers, and therefore has three generators. As we discussed in these notes, we should think of these spaces as vector spaces and the generators as basis vectors spanning the entire space. The same is true of $SU(3)$, though it is an eight-dimensional space. So, because there is a space associated with each of these groups, it should be somewhat obvious that there is a natural geometric picture associated with a Lie group.

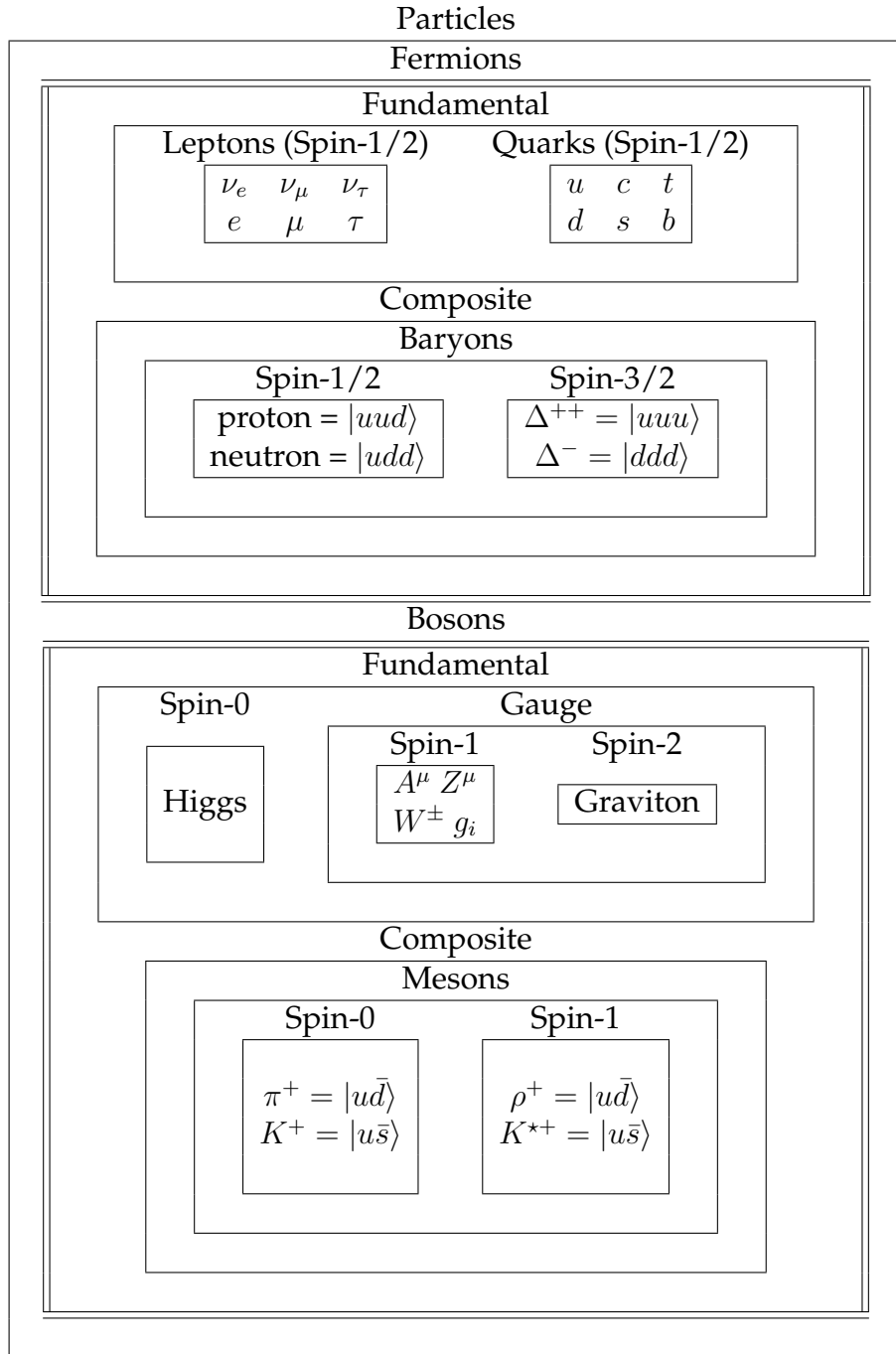
While the idea of the parameter space of a Lie group having a geometric picture associated with it may seem straightforward, the geometry undergirding gauge theory can be extremely complicated, and we therefore must spend a significant amount of time investigating it. Therefore, the next paper in this series will be an introduction to the geometric structure of gauge theory. Just as we have built gauge theory from algebra, we will in a sense start over and rebuild it using geometry. However, because we have already covered a great deal of detail in the physics and mathematics of gauge theory and particle physics in general, we will move much more quickly to avoid being repetitive.

When we finally get to String Theory (later in this series), we will see that the geometric and algebraic pictures come together beautifully, and that a thorough understanding of both will be necessary to understand what may be the “ultimate” theory of our universe.

References

- [1] D. Bailin and A. Love, "Introduction to Gauge Field Theory", Taylor and Francis (1993)
- [2] R. Cahn, "Semi-Simple Lie Groups and Their Representations", Dover (2006)
- [3] W. N. Cottingham and D. A. Greenwood, "Introduction to the Standard Model", Cambridge University Press (2007)
- [4] R. Dunlap, "An Introduction to the Physics of Nuclei and Particles", Brooks Cole (2003)
- [5] V. Ezhela and B. Armstrong, "Particle Physics: One Hundred Years of Discoveries: An Annotated Chronological Bibliography", Springer (1996)
- [6] R. P. Feynman, R. B. Leighton, and M. Sands, "The Feynman Lectures Vols. I-III", Addison Wesley (2005)
- [7] K. Ford, "The Quantum World", Harvard University Press (2005)
- [8] J. B. Fraleigh, "A First Course in Abstract Algebra", Addison Wesley (2002)
- [9] H. Georgi, "Lie Algebras and Particle Physics", Westview Press (1999)
- [10] R. Gilmore, "Lie Groups, Lie Algebras, and Some of Their Applications", Dover (2006)
- [11] R. Gilmore, "Lie Groups, Physics, and Geometry", Cambridge University Press (2008)
- [12] H. Goldstein, C. P. Poole, and J. L. Safko, "Classical Mechanics", Addison Wesley (2001)
- [13] D. J. Griffiths, "Introduction to Classical Electrodynamics", Benjamin Cummings (1999)
- [14] J. Hakim, "The Story of Science: Aristotle Leads the Way", Smithsonian Books (2004)
- [15] B. C. Hall, "Lie Groups, Lie Algebras, and Representations", Springer (2004)
- [16] J. E. Humphreys, "Introduction to Lie Algebras and Representation Theory", Springer (1994)
- [17] T. W. Hungerford, "Algebra", Springer (2003)
- [18] J. D. Jackson, "Classical Electrodynamics", Wiley (1998)
- [19] J. V. Jose and E. J. Saletan, "Classical Dynamics: A Contemporary Approach", Cambridge University Press (1998)

- [20] J. Mehra and H. Rechenberg, "The Historical Development of Quantum Theory; v.6", Springer (2001)
- [21] G. Naber, "Topology, Geometry, and Gauge Fields: Interactions", Springer (2000)
- [22] G. Naber, "The Geometry of Minkowski Spacetime", Dover (2003)
- [23] M. Nakahara, "Geometry, Topology, and Physics", Taylor and Francis (2003)
- [24] M. E. Peskin and D. V. Schroeder, "An Introduction to Quantum Field Theory", Westview Press (1995)
- [25] A. Pickering, "Constructing Quarks: A Sociological History of Particle Physics", University of Chicago Press (1984)
- [26] P. Ramond, "Field Theory: A Modern Primer", Westview Press (2001)
- [27] P. Ramond, "Journeys Beyond the Standard Model", Westview Press (2003)
- [28] J. Rotman, "An Introduction to the Theory of Groups", Springer (1999)
- [29] L. H. Ryder, "Quantum Field Theory", Cambridge University Press (1996)
- [30] B. Sagan, "The Symmetric Group", Springer (2001)
- [31] J. J. Sakurai, "Modern Quantum Mechanics", Addison Wesley (1993)
- [32] M. Srednicki, "Quantum Field Theory", Cambridge University Press (2007)
- [33] J. Schwinger, "Classical Electrodynamics", Westview Press (1998)
- [34] N. M. J. Woodhouse, "Special Relativity", Springer (2007)
- [35] A. Zee, "Quantum Field Theory in a Nutshell", Princeton University Press (2003)



	Leptons			Hadrons				Higgs
	$(1, \mathbf{2}, -1/2)$	$(1, 1, 1)$		$(\mathbf{3}, \mathbf{2}, 1/6)$	$(\mathbf{3}, 1, -2/3)$	$(\mathbf{3}, 1, 1/3)$		$(1, \mathbf{2}, -1/2)$
Generation 1	$\begin{pmatrix} \text{electron neutrino} \\ \text{electron} \end{pmatrix}$	electron		$\begin{pmatrix} \text{up} \\ \text{down} \end{pmatrix}$	up	down		1 Generation Only
Generation 2	$\begin{pmatrix} \text{muon neutrino} \\ \text{muon} \end{pmatrix}$	muon		$\begin{pmatrix} \text{charm} \\ \text{strange} \end{pmatrix}$	charm	strange		
Generation 3	$\begin{pmatrix} \text{tau neutrino} \\ \text{tau} \end{pmatrix}$	tau		$\begin{pmatrix} \text{top} \\ \text{bottom} \end{pmatrix}$	top	bottom		