

An Introduction to Gauge Field Theories: From Electrons to Vector Dark Matter

Henry Atwater^{}, Yuri Rostovtsev^{}

Center for Nonlinear Sciences, Department of Physics, University of North Texas, Denton, TX, USA
Email: hankatwater@my.unt.edu, Yuri.Rostovtsev@unt.edu

How to cite this paper: Atwater, H. and Rostovtsev, Y. (2025) An Introduction to Gauge Field Theories: From Electrons to Vector Dark Matter. *Journal of Applied Mathematics and Physics*, 13, 2465-2495. <https://doi.org/10.4236/jamp.2025.137141>

Received: June 21, 2025

Accepted: July 28, 2025

Published: July 31, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0). <http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The various quantum field theories now recognized collectively as the standard model (SM) are the most precise and predictive models of the universe ever proposed. There is much to explore in each theory and each of the fundamental particles and interactions possess their own unique qualities and conundrums. However, if we zoom out, we see the mathematics posited by nature in each of these theories are one and the same in structure. This most enchanting structure is colloquially known amongst physicists as *Yang-Mills theory* or, more generally, *gauge field theory*. This article is meant to serve as an introduction to the world of gauge theories for the physics student interested in particle physics and adjacent fields. First is a brief discussion of groups followed by an optional discussion of smooth manifolds. A basic introduction to Lie groups and algebras is provided intuitively, without relying on the context of representation theory. Several globally invariant systems are then presented, followed by a discussion on the generalization to local invariance. A formal yet approachable predication of a general gauge theory is then given, along with a variety of examples. The second to last section introduces an extension of the previously discussed version of gauge theory to theories currently being investigated as models for vector dark matter. In the final section we present and briefly discuss a theory invariant under action by the quaternionic sphere. It is assumed the reader has taken advanced undergraduate courses in classical mechanics, electromagnetic theory, and quantum mechanics. Therefore, the reader should have a firm grasp on concepts from linear algebra/vector space theory (including abstract vector spaces), calculus (including that of functions of a complex variable), and has a *passive understanding* of basic notions from differential geometry—such as the algebra of smooth functions, tangent spaces and their dual/cotangent spaces, metric tensors, etc. Here, a “passive understanding” is meant as little more than the equivalent to having an operational knowledge of tensor indices. Even though there is a section dedicated to defining a smooth manifold, a background in topology is not

required. Beyond said section, concepts from topology are mentioned sparsely. Formal knowledge of group representations is not necessary; however, the reader should acknowledge that many of the groups relevant to this paper have *representations* in each of an infinite number of dimensions—a fact that may be confronted if the literature on this general topic is pursued further.

Keywords

Gauge Theory, Standard Model, Lie Groups, Lie Algebras, Dark Matter, Quaternion

1. Introduction

Before the 20th century, the first basic mathematical consequence of what would be later understood as a gauge theory was noticed in Maxwell's theory of electromagnetism when the electric and magnetic fields were generalized to a duo of vector and scalar potential functions. It was noticed is that the resulting electric and magnetic fields described by Maxwell's equations remain unaltered by the addition and subtraction (or vice versa) of an arbitrary smooth function's time derivative and spatial derivatives to the scalar and vector potential respectively. Every year, undergraduate physics students take courses in electromagnetism and inevitably encounter a peculiar two-word description of this mysterious mathematical consequence of Maxwell's theory: *gauge invariance*. The trouble is this idea is often discussed to a degree that is deemed less than satisfactory by the student. There is good reason for this, as truly appreciating this topic requires a firm grip on some rather heavy mathematical machinery—a large component of this machinery is an intricate cocktail of algebra, geometry, and topology called Lie theory. However, gauge theory is principally mentioned in undergraduate electromagnetism because of its colossal utility in modern theoretical physics. Therefore, it's a shame that students are usually left to their own devices in navigating the literature surrounding this topic to learn more—a task that can be demoralizing without a certain amount of mathematical training. The main motivation for this article is to create a resource accessible to undergraduates that outlines the mathematical framework of gauge theory and conveys the unparalleled beauty and utility it possesses when applied to the particle physics of the standard model and beyond. The hope is that this work aids in nourishing the minds of those curious students who are still dissatisfied with their knowledge of this concept and provide guidance in the pursuit of its understanding.

In the early 20th century, the primary focus of the physics community departed from the classical realm to embrace a new, quantum mechanical approach to describing the universe. Ever since this pivotal period, which is often referred to as the quantum revolution, many of the core ideas that have crept forth and shaped the frontier of physics have become increasingly abstract and further removed from our classical intuition. As a result, these new ideas confronted by physicists

and certain mathematicians over the years have become increasingly difficult to communicate to the public. This is a tragedy, because public interest surrounding these ideas has only trended upward as the subject has become more puzzling and bizarre. Gauge invariance is a concept so central to the modern understanding of physics that physicists simply can't help but mention it when they're tasked with informing the public. Because of this, it has unintentionally earned its place among buzz-word phenomena like quantum entanglement and time dilation in the popular science sphere. It's been commented on in books written by prominent science communicators like Lawrence Krauss and Michio Kaku and made several appearances in YouTube videos created by science channels like PBS Spacetime. Not unlike the other two phenomena mentioned, attempts to convey the nature of gauge invariance to the lay community has been largely done through story and analogy. These attempts usually don't succeed fully in capturing much of the brilliance in the subject matter. A secondary motivation for writing this article is for the members of the public who are hungry for a resource on the gauge theoretic aspect of fundamental physics that is more friendly than a textbook but goes into much more rigorous detail than the abstruse analogies often provided by popular science literature and media.

The quantum field theoretic notion of gauge theory is one of the most fruitful and resilient ideas resulting from the quantum revolution. It's application in these theories is the reason gauge theory has survived and held its ground as a concept fundamental to our reality since it took its fully ripened form as Yang-Mills theory around the middle 20th century. Nearly every significant development in particle physics that has been observed since then was previously theorized or understood subsequently by way of a gauge theory. The Aharonov-Bohm effect, the Higgs mechanism, and quarks, all are observed phenomena whose mathematical descriptions are hinged on the principles of gauge theory. In fact, every elementary particle and three of four fundamental interactions we know to exist appear to be, for the most part, unified and well-modeled within the tripartite conglomeration of gauge theories known as the standard model of particle physics.

What makes gauge theory so powerful in the realm of particle physics is what makes it so difficult to convey meaningfully without mathematics. This is the fact that its defining property is a collection of smooth symmetry transformations (Lie groups) under which the dynamical descriptions of particles (Lagrangians/equations of motion) are unaffected. These symmetries correspond to the quantum numbers (spin, electric charge, color charge, etc.) possessed by fermionic particles and seamlessly merge them with the bosonic particles that mediate their interactions. In the standard model and more generally in classical/quantum field theory, particles are fields (most of the time spinor or vector fields) that roam around on spacetime yet live up above spacetime in "internal spaces". The term "internal spaces" in the context of gauge theory should be understood conceptually as the spaces where the smooth symmetries associated with given interactions exist and transform things (as opposed to the Hilbert or Fock space of particle states). The

interactions that fermions are allowed to take part in directly corresponds to the symmetries that naturally transform their internal field spaces.

The demonstrated effectiveness of gauge theory in well-modeling every particle we've observed makes theories of this type a primary choice in the search for models of particles we haven't observed yet but have reason to believe exist. Dark matter is, of course, the most well-known of the hypothesized yet unobserved particles. The reason it's a primary choice is because if we understand gauge theory, and we know that every fundamental interaction corresponds to a symmetry, it gives us some footing in our speculation. For instance, dark matter got its name because it doesn't seem to interact very much—or at all—with light (electromagnetism). If we assume dark matter has a particle-like description (a field over spacetime with internal structure), and we know that the electromagnetic interaction between two particles only occurs if a one-dimensional smooth symmetry called $U(1)$ is present, we can hypothesize that, whatever the particle description of dark matter is, it probably isn't invariant under transformations by the $U(1)$ symmetry. This is slightly ironic to mention, because the toy model dubbed a theory of vector dark matter in this article is a theory invariant under the symmetry associated with electromagnetism.

We have put a fair amount of stress on gauge theory being a body of ideas that is both tremendously useful from the perspective of particle physics and breathtakingly beautiful from that of mathematics. However, grandiose comments on these aspects of gauge theory are akin to the popular science descriptions previously complained about, and they don't really shed any light on how it works. Despite earlier complaints about using stories and analogies to convey ideas in theoretical physics, the author humbly invites the reader to entertain the following flawed analogy to convey how gauge theory works in the context of the standard model.

Imagine a custom box of tinker toys, specifically spools and rods, and a play area which is just a flat open region of floor. Pretend that we look at each spool and see that not all of them have the same number of holes around the outer edge or through the center. We notice that some of them are more like wheels and have only one hole through the center. Other spools have different numbers of holes equidistant from each other around the outer edge and are not necessarily in possession of the center hole. The number of holes each spool has determines how many other spools it can be connected to via a rod. We also notice that not all the rods have the same length and cross-section, so that they are only able to connect two spools in possession of holes with the corresponding shape. The spools in this analogy are matter particles (electrons, quarks, etc.) whose internal spaces are characterized by their respective number of holes. The rods are the interaction particles (photons, gluons, etc.) and when they connect two spools, represent an interaction. Both the spools and rods are left unaltered by specific symmetry transformations (rotations) with dimension equal to the total number of holes and rods of the same cross-section. The play area represents the topologically non-threat-

ening flat manifold (usually Minkowski spacetime) we almost always choose to play with field theories on. We can move spools around to different points on the floor (which are different points on the manifold) and when we move two spools with like holes close enough together, they can be connected using a rod specific to the holes they have in common. We could say a single circular hole through the center represents the spool's possession of electric charge and its ability to interact via the photon rod. If a spool has three triangular holes around the outer edge, it might indicate that it's able to interact with other spools through three different rods (W^+ , W^- , Z^0) associated with the weak interaction. A spool with only one circular center hole could be a right chiral electron, which only interacts with other matter via electromagnetism. A spool with both the center hole and three triangular holes might be a left chiral electron, which can participate in both electromagnetic and weak interactions. If a spool doesn't have the circular center hole but has the three triangular ones, it might be a left chiral electron neutrino. If we place an electron spool near an electron neutrino spool, they cannot interact via the photon rod. If a spool has a central circular hole and eight octagonal holes in superposition with three triangular holes around the outer edge (the symmetries must be considered independently), it is certainly a flavor of quark and able to participate in the strong interaction via gluon rods.

The analogy could potentially be taken slightly further by color coding spools and rods and so on, but we can get a somewhat insightful glimpse of the standard model varieties of particles, interactions and the symmetries that correspond to them in the standard model as it is. To extend this analogy to gauge theories beyond the standard model, we'd likely need to be more creative—there are an infinite number of symmetries available to construct an infinite number of gauge theories around. Regardless of how creative we are, there's only so much an analogy can do. The time has come to embark on our quest to understand what gauge theory really is.

2. Group Theory and Matrix Groups

2.1. Group Theory

Definition: A group is two things: a set, G , and a binary operation, $*$: $G \times G \rightarrow G$, often written as a pair $(G, *)$, that together satisfy four requirements:

- 1) **Closure:** For any $g, h \in G$, $(g * h) \in G$;
- 2) **Associativity:** For any $g, h, k \in G$, $(g * h) * k = g * (h * k)$;
- 3) **Identity:** There exists $e \in G$ such that $e * g = g * e = g$ for all $g \in G$ — the identity is provably unique (you can't have more than one);
- 4) **Inverse:** For every $g \in G$ there is a $g^{-1} \in G$ so that $g * g^{-1} = g^{-1} * g = e$.

Any combination of set and operation that satisfy these four conditions constitute a group. This general abstract definition is very far reaching and is not in any way limited to sets of numbers. However, groups with sets containing abstract elements not akin to numbers (or arrays of numbers) are beyond the scope of this article and will not be discussed. We will restrict our examples to non-abstract

sets and operations which a physics student is familiar with. For instance, $(\mathbb{R}, +)$ —the set of real numbers with operation addition—forms a group with identity $e = 0$ and $g^{-1} = -g$ for all $g \in \mathbb{R}$. However, (\mathbb{R}, \cdot) , where the dot operation is multiplication, is not a group unless we exclude the $\{0\}$ element from \mathbb{R} . In this case, $(\mathbb{R} \setminus \{0\}, \cdot)$ forms a group with identity $e = 1$ and $g^{-1} = \frac{1}{g}$ for all g . Conditions of closure and associativity are trivially checked for these examples; it's obvious that any real number added or multiplied by another is again a real number and both operations are associative. Similar conclusions can be made for other sets like the complex numbers, \mathbb{C} , the rational numbers, \mathbb{Q} , the integers, \mathbb{Z} , etc. The groups above have elements that can be combined under their operation in any order (they commute). These commutative groups are a special case of general groups and therefore deserve a special name. We call a group whose elements commute an Abelian group.

A map $\rho: G \rightarrow H$ between two groups $(G, *)$ and (H, \diamond) is called a homomorphism if for every $g_1, g_2 \in G$, $\rho(g_1 * g_2) = \rho(g_1) \diamond \rho(g_2)$. In words, two groups are homomorphic if there exists a map between them that preserves group operations.

An example of a homomorphism is a map $\varphi: (\mathbb{R}^2, +) \rightarrow (\mathbb{C}, +)$, sending coordinate pairs $(x, y) \in \mathbb{R}^2$ to numbers $x + iy \in \mathbb{C}$ and sending the operation of vector addition on \mathbb{R}^2 to the usual addition on \mathbb{C} . This relation clearly preserves each group's respective operations. Another example is the exponential map $\rho: (\mathbb{R}, +) \rightarrow (\mathbb{R}^+, \cdot)$, where \mathbb{R}^+ is all real numbers greater than zero. The dot operation is once again multiplication. We can check the group operations are preserved: for $\rho(x) \equiv e^x$, we have $\rho(a + b) = e^{a+b} = \rho(a) \cdot \rho(b) = e^a \cdot e^b$ for any $a, b \in \mathbb{R}$ and $\rho^{-1}(a) = e^{-a}$.

Another relevant definition to include in this section is that of an *algebra*. We choose to define an algebra as nothing more than an additive group, $(G, +)$, which is given an additional binary operation $\circ: G \times G \rightarrow G$.

The set of vectors in \mathbb{R}^n with addition is an additive group (and together with scalar multiplication, a vector space). An example of an algebra would be the $(\mathbb{R}^3, +, \times)$ where \times is the cross product of two vectors.

2.2. Matrix Groups

The kinds of groups most relevant to particle physics are matrix groups—simply sets whose elements are matrices and whose operation is matrix multiplication, that together satisfy the four requirements of group-hood. It should be immediately deduced from the latter part of the previous sentence the identity element of these groups is always the identity matrix. Due to this choice of operation, group elements in general do not commute with each other. We call general groups whose elements don't commute non-Abelian. An important example of a matrix group is the group of all invertible linear transformations on a vector space V , usually \mathbb{C}^n or \mathbb{R}^n for some positive integer n , written $GL(V)$. The matrix

groups in particle physics are most often communicated as subgroups of $GL_n(\mathbb{C})$ for some n . These subgroups can be simply thought of as $GL_n(\mathbb{C})$ with extra requirements. We almost always choose to work on \mathbb{C}^n for technical ease. This is because $\mathbb{C}^n \cong \mathbb{R}^{2n}$ for all n .

Important Examples:

- The set of all $n \times n$ matrices, $Mat_n \mathbb{C}$, does not form a group under matrix multiplication since not all its elements have inverse. However, it forms a group under addition (more importantly a vector space) that we will come to know in the context of Lie theory.

- The general linear group on \mathbb{C}^n is

$$GL_n(\mathbb{C}) \equiv \{ \text{all } A \in Mat_n \mathbb{C} \text{ such that } \det|A| \neq 0 \}$$

- The special linear group on \mathbb{C}^n is

$$SL_n(\mathbb{C}) \equiv \{ \text{all } A \in GL_n(\mathbb{C}) \text{ such that } \det|A| = 1 \}.$$

- The unitary group on \mathbb{C}^n

$$U_n(\mathbb{C}) \equiv \{ \text{all } A \in GL_n(\mathbb{C}) \text{ such that } A^\dagger = A^{-1} \}.$$

- The special unitary group on \mathbb{C}^n is

$$SU_n(\mathbb{C}) = SU(n) \equiv \{ \text{all } A \in GL_n(\mathbb{C}) \text{ such that } \det|A| = 1 \text{ and } A^\dagger = A^{-1} \},$$

or equivalently $\{ \text{all } A \in SL_n(\mathbb{C}) \text{ and } A^\dagger = A^{-1} \}$, or $SL_n(\mathbb{C}) \cap U_n(\mathbb{C})$.

- Similarly, the special orthogonal group on \mathbb{R}^n is

$$SO_n(\mathbb{R}) = SO(n) \equiv \{ \text{all } A \in SL_n(\mathbb{R}) \text{ such that } A^T = A^{-1} \}.$$

If $n=1$, there is no notion of determinant, so $SU_1(\mathbb{C}) = U_1(\mathbb{C}) = U(1)$. This is the one-dimensional group of rotations in the complex plane, and the symmetry/gauge group for the electromagnetic interaction.

3. Smooth Manifolds

To fulfill the previous promise that a strong background in topology and other regimes of higher mathematics need not be required to read this article, we limit this section's development to the requirements a topological space must meet to be called a smooth manifold. Even further, this section is optional for the reader who has no knowledge of topology. This is for the sake of accessibility, as most physics undergraduates have had limited exposure to such ideas. For this reason, a full comprehensive development of the machinery for calculus on manifolds is sufficiently cumbersome to detract from the purpose of this article. But, since requiring a set to have smooth manifold structure is half the definition of a Lie group (and Lie groups are gauge groups), it would be in bad taste to not include a definition in this article on gauge theory. Also, for the most part, gaining a workable understanding and appreciation for the structure of gauge theories from an undergraduate physicist's perspective doesn't warrant a need for such definitions. Therefore, we give the reader two options based on their level of comfort with topology:

1) The reader has no experience with topology:

If the reader has no experience with topology, they should feel free to skip this section for now, taking with them the tractable idea that a smooth manifold, is a just space equipped with the properties that allow us to parameterize various functions on it whose partial derivatives exist everywhere. The requirements of the smooth manifold ensure we can traverse between its points smoothly. We need calculus and linear algebra to do physics, and a smooth manifold is pretty much the cathedral for the symphony.

2) The reader knows some topology:

In the case of the reader who has some footing in topology, they should have little trouble hashing out the ideas presented in this section. The topological structure we wish to define is commonly referred to as a smooth manifold. A smooth manifold is just a topological manifold with (potentially a choice of) smooth structures definable on it, so it makes sense to first provide a definition of the latter.

3.1. Manifolds

Definition: An **n-dimensional manifold**, M , is a topological space satisfying:

1) **The Hausdorff condition:** For any two distinct points $p, q \in M$ there exists open subsets $U, V \subseteq M$ such that $p \in U, q \in V$, and $U \cap V = \emptyset$.

2) **Paracompact-ness:** Every *open cover*—which is a (not necessarily finite) collection of open sets $\{U_i\}$ where each $U_i \subseteq M$ such that the union of all U_i 's contains M —has an *open refinement* that is *locally finite*. An open refinement is locally finite if for every point $p \in M$ there is an open subset $V \subseteq M$ containing the point p such that V intersects finitely many of the U_i 's in the cover.

- The real line is paracompact. However, it does not heed to the stricter requirement of a *compactum*; a topological space is compact if any open cover has a *finite subcover*. A finite subcover is a finite subset of elements from the cover whose union is equivalent to the union of the cover. Obviously, all compact spaces are paracompact.

3) **Local \mathbb{R}^n -ness:** Every point $p \in M$ is contained in an open set $U_p \subseteq M$ that is homeomorphic to an open subset of some \mathbb{R}^n .

If the homeomorphism to some \mathbb{R}^n is not only local but global, meaning the manifold itself admits this homeomorphism, then our lives are made easy since any function or field that lives on M is essentially just a function or field on some \mathbb{R}^n . One example of this is $M = \mathbb{C}^d \cong \mathbb{R}^{2d}$. Another example is Minkowski space, which is identical to \mathbb{R}^4 apart from the notion of lengths defined on it (the metric is *split signature*). Unfortunately, this is not the case for most manifolds. Just because a space looks locally like \mathbb{R}^n doesn't mean the same is true globally. Which is why we need the following definitions of *charts* and *atlases*.

3.2. Charts, Atlases, and Smooth Manifolds

A chart (U, σ) is an open $U \subseteq M$ and a homeomorphism $\sigma: U \rightarrow N$ where $N \subseteq \mathbb{R}^n$. A collection of charts $A = \{(U_i, \sigma_i)\}$ is called an **atlas** if the collection

of U_i 's cover M . As stated, it's necessary to consider atlases because if our manifold is globally unlike \mathbb{R}^n , for instance if $M = S^2$ (the surface of a sphere in 3-dimensions) which is not globally like \mathbb{R}^n , we want to always be able to stitch overlapping charts together (especially functions on them) in a way that is *compatible on transitions*. In lieu of the following definition, it's easy to see the usefulness in \mathcal{A} being defined as a cover of M .

Definition: Two charts (U_a, σ_a) and (U_b, σ_b) where $U_a \cap U_b \neq \emptyset$ are compatible on transitions if there exists the composition of chart maps

$$\tau_{ab} \equiv \sigma_b \circ \sigma_a^{-1} : \sigma_a(U_a \cap U_b) \rightarrow \sigma_b(U_a \cap U_b)$$

called a transition map.

A composition of homeomorphisms is itself a homeomorphism. It's hopefully clear from the definition why the domain of the map must be restricted to $\sigma_a(U_a \cap U_b)$ rather than, say, $\sigma_a(U_a)$.

A composition of smooth functions is not necessarily smooth. However, in the delightful case that we possess an atlas with smooth transition functions, we call it a smooth atlas. If M is a manifold equipped with a smooth atlas, then M is a smooth manifold. As stated at the beginning of this section, defining all this machinery on sets to turn them into smooth manifolds is carried out because it, along with a bit more structure, eventually allows us to use the techniques of Riemannian/Pseudo-Riemannian geometry and do calculus on it. If we want to talk meaningfully about smooth structures beyond the basic definitions covered here, we will need to spend some time carefully developing the theory of calculus on manifolds. However, one can easily learn the techniques necessary to do dynamical calculations on flat manifolds (like Minkowski spacetime) without the machinery of bundles, pushforwards and pullbacks, etc. To perform any calculation in this article, all one must know is how to raise/lower and contract tensor indices with a flat space metric.

4. Lie Groups and Algebras

A *Lie group* is a set given both group and smooth manifold structure at the same time. We can convey the symbiosis of these structures with a slightly more revealing definition:

A Lie group G is a group, where actions of the group on itself, usually done by multiplication, are smooth maps, and the inversion of any group element is also a smooth map.

It's a simple statement that you may have heard before, however the question may remain as to precisely how these simultaneously algebraic and topological spaces are constructed and studied. To understand this, we must first understand the matrix exponential.

4.1. The Exponential Map

Given a matrix $A \in Mat_n \mathbb{C}$, we can understand the result of the exponentiation of A by considering the Taylor expansion

$$e^{tA} = \sum_{k=0}^{\infty} \frac{1}{k!} t^k A^k = I_{n \times n} + tA + \frac{1}{2!} t^2 A^2 + \dots,$$

which is itself a matrix. The A^k terms are just the matrix multiplied by itself k -times.

If a given matrix $A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} \in Mat_n \mathbb{C}$ is diagonal or triangular, we have $e^A = \begin{pmatrix} e^{a_{11}} & \dots & * \\ \vdots & \ddots & \vdots \\ * & \dots & e^{a_{nn}} \end{pmatrix}$.

If A is diagonalizable: $A = MDM^{-1}$, which equivalently means $M^{-1}AM = D$ for an M invertible and D diagonal, we have $e^A = Me^DM^{-1}$ or $M^{-1}e^AM = e^D$. The exponential of any (not necessarily invertible) square matrix is invertible since $(e^A)^{-1} = e^{-A}$ for any $A \in Mat_n \mathbb{C}$.

Some other properties we should notice:

- 1) If A is diagonal, then $\det(e^A) = e^{trA}$.
- 2) The transpose commutes with the exponential map: $(e^A)^T = e^{A^T}$.
- 3) The conjugate transpose commutes: $(e^A)^\dagger = e^{A^\dagger}$, which is true by the previous assertion combined with the trivial fact $(e^A)^* = e^{A^*}$ where $*$ denotes the complex conjugate.
- 4) $e^A e^B = e^{A+B}$ is true for two matrices A, B if they commute ($AB = BA$).
- 5) For $t \in \mathbb{R}$, the map $t \rightarrow e^{tA}$ is a homomorphism $(\mathbb{R}, +) \rightarrow GL_n \mathbb{C}$, since $e^{tA} e^{t'A} = e^{(t+t')A}$ and $e^{-tA} = (e^{tA})^{-1}$.

In exponentiating matrices, we often refer to the exponential map as a *lifting* of an object in $Mat_n \mathbb{C}$ to an object in $GL_n \mathbb{C}$.

4.2. Lie Groups and Algebras

The exponential map, $exp: Mat_n \mathbb{C} \rightarrow GL_n \mathbb{C}$, is a map from the vector space formed by all $n \times n$ matrices to the group of all invertible matrices. We also know that the exponential is a smooth function (it has a Taylor series), and thus the multiplication and inversion of group elements are smooth operations. We have just realized that $GL_n \mathbb{C}$ is a Lie group—whose *Lie algebra* is $Mat_n \mathbb{C}$. If we consider a specific subset of all matrix exponentials e^{tA} for $A \in Mat_n \mathbb{C}$ and $t \in \mathbb{R}$ that itself forms a group (a subgroup), this means, by definition, the subgroup is, itself, a Lie group.

Suppose $G \subseteq GL_n \mathbb{C}$ is a subgroup (the word subgroup here is crucial, as it states the subset G itself forms a group). Then we define the Lie algebra $T_e G$ of the Lie group G as follows.

$$T_e G = \{x \in Mat_n \mathbb{C} : e^{tx} \in G \text{ for any } t \in \mathbb{R}\}$$

In words, the Lie algebra of a Lie group $G \subseteq GL_n \mathbb{C}$ is just the subset of matrices from $Mat_n \mathbb{C}$ that when multiplied by any real number, under the exponential map all end up as points in the Lie group G .

There is a common convention worth mentioning. As stated, the elements of the Lie algebra of $GL_n\mathbb{C}$ are the elements of $Mat_n\mathbb{C}$, and if we are talking about the general space of these matrices we call it exactly that—either $Mat_n\mathbb{C}$ or $End(\mathbb{C}^n)$. The notation $End(\mathbb{C}^n)$ is short for the *endomorphisms* of \mathbb{C}^n , which is the set of all transformations from \mathbb{C}^n to itself. However, if we are referring to $Mat_n\mathbb{C}$ as the Lie algebra of $GL_n\mathbb{C}$, we write $\mathfrak{gl}_n\mathbb{C}$.

Again, we say the elements of the Lie algebra *lift* via exponential map to elements of the Lie group. Perhaps this isn't obvious yet, but all the structure of a (*connected*) Lie group is encoded in the structure of its Lie algebra. For this reason, the study of any Lie group reduces to studying its Lie algebra. We are happy about this, because the group is also a smooth manifold and therefore has infinitely many elements and could have other topological complications. However, the algebra is (at least for the cases we are interested in) a vector space of finite dimension—meaning we can express any element as a finite linear combination of basis vectors. Something that should be relatively obvious is that this finite number of elements that span the algebra is precisely equal to the dimension of the manifold, and the elements of the Lie algebra live in and form the basis for the *tangent space at the identity* of the group. The vector space $\mathfrak{gl}_n\mathbb{C}$ has complex dimension n^2 , and real dimension $2n^2$. This in turn means that $GL_n\mathbb{C}$ is a real manifold of dimension $2n^2$.

Now that we've seen how Lie groups are “grown out of” their Lie algebras, we should discuss how we might extract the Lie algebra that corresponds to a given Lie subgroup of $GL_n\mathbb{C}$ from its group elements. Earlier in this section, when we realized $GL_n\mathbb{C}$ is a Lie group, we mentioned the first key to doing this: the exponential map is smooth. The second key we need to unlock the door to a Lie group's Lie algebra is another previously mentioned fact: the Lie algebra is a (real) vector space—which, by definition, means we can freely multiply the vectors (which are not necessarily invertible matrices) by real scalars without affecting the Lie group to which they are mapped under exponentiation. This notion is not just fascinating, but also convenient, because it gives us the means to take derivatives of arbitrary group elements with respect to the real parameter.

To extract from an element of a Lie group the Lie algebra element that generated it, take an arbitrary element X of the Lie group G as $X = e^{tx}$ for $t \in \mathbb{R}$ and some element of the Lie algebra $x \in T_e G$. If we take a derivative with respect to t we get $D_t X = D_t(e^{tx}) = xe^{tx}$. If we evaluate our derivative at $t = 0$, $D_t X|_{t=0} = xI = x$ where I is the identity matrix, directly gives $x \in T_e G$. The derivative evaluated at the identity matrix is a map to the Lie algebra. This is the reason we say the Lie algebra is the tangent space at the identity. In this same spirit we sometimes interpret the Lie algebra as the infinitesimal behavior of the group—the first order term in the Taylor expansion of an arbitrary group element is the Lie algebra element that generated it. Fair warning, in the bracket section we will sometimes state simply “derivative” to mean “derivative evaluated at the identity”—it is implied by the context that evaluation at the identity is part of the plan.

4.3. The Bracket

In defining the Lie algebra of a Lie subgroup of $GL_n\mathbb{C}$, we are confronted with a theorem. The theorem is this: the Lie algebra, T_eG , of a Lie group, $G \subseteq GL_n\mathbb{C}$, is closed under the bracket, which is the specific operation that promotes its status as a vector space (a group with addition) to that of an algebra. We do not prove this theorem here but instead opt to do our best to motivate the existence of the bracket operation—for a multitude of reasons, being comfortable with the bracket is essential. This operation can seem quite mysterious at first, and motivating it involves some abstract ideas. In developing the bracket, what we are looking for is an operation that will encapsulate all the information necessary to completely characterize a given Lie algebra—which ultimately means that we can use it to compare Lie algebras (discern isomorphisms). Keep in mind that sometimes there is no reason to pursue a notion other than for the reason it does something useful. However, it is often the case that notions which are first mysterious and abstract, when viewed through the right lens, reveal themselves as very natural and tractable ideas. This is certainly the case here.

As stated earlier, the Lie algebra is a vector space. We need to put this fact to use, so we should state it in more concrete terms. The Lie algebra T_eG of any closed subgroup $G \subseteq GL_n\mathbb{C}$ is a real vector space. The term “closed” in the last sentence refers to topologically closed. We say this explicitly because the only topologically open subgroup of $GL_n\mathbb{C}$ is itself. We say the Lie algebra is a real vector space because we usually want to think of the corresponding Lie group as a real manifold.

Define the conjugation of $h \in G$ by $g \in G$ as $C_g(h) = ghg^{-1}$. This arcane operative construction is ubiquitous in the general study of algebra, as it serves as a means of identifying important subsets of algebraic structures. Actions of a group on itself by conjugation are *automorphisms* of the group. An automorphism is an invertible endomorphism. For Abelian groups, conjugation is uneventful:

If G is Abelian, then $C_g(h) = ghg^{-1} = gg^{-1}h = hgg^{-1} = h$ for all $g, h \in G$.

In the general case, where G is not necessarily Abelian, we can only say for sure that one point is left unscathed by conjugation. For any group G , we have that $C_g(e) = geg^{-1} = e$ for $e \in G$ the identity. Assuming the group is a (*connected*) Lie group (which means $C_g : G \rightarrow G$ is smooth), the fact that the identity is preserved by the conjugation automorphism makes it meaningful to take the differential of the automorphism and evaluate it there. The result of this differential is an automorphism of the Lie algebra $dC_g|_e : T_eG \rightarrow T_eG$. We call this most important automorphism the adjoint action of the Lie group G on its Lie algebra T_eG , and denote it $Ad_G : T_eG \rightarrow T_eG$. As is obvious from the map’s definition and not too difficult to justify, is that T_eG is Ad_G -invariant (T_eG is closed under Ad_G). The adjoint action is just a special name for the conjugation by a Lie group on its algebra, and just as conjugation preserved the identity element of the

group, adjoint action preserves $T_e G$ as a vector space. Given some $x \in T_e G$ and $g \in G$, the adjoint action of g on x is written as $Ad_g(x) \equiv gxg^{-1} \in T_e G$.

The adjoint action should be thought of as a very natural way for a Lie group to act on its Lie algebra, and it will come up later when we utilize Lie algebras to achieve gauge symmetry—but more on that later. We set out to define the bracket, which we now show can be readily thought of as the infinitesimal behavior of the adjoint action.

If we again consider an arbitrary element g of a Lie group G as $g = e^{ty} \in G$ for $t \in \mathbb{R}$, $y \in T_e G$ acting on $x \in T_e G$, we have that

$$Ad_g(x) = Ad_{e^{ty}}(x) = e^{ty} x e^{-ty} \in T_e G.$$

Differentiating gives:

$$\begin{aligned} D_t \left[Ad_{e^{ty}}(x) \right] \Big|_{t=0} &= D_t \left[e^{ty} x e^{-ty} \right] \Big|_{t=0} = \left[y e^{ty} x e^{-ty} - e^{ty} x y e^{-ty} \right] \Big|_{t=0} \\ &= yx - xy \equiv [y, x] \end{aligned}$$

We have just identified the bracket, sometimes called $ad_{T_e G}$, which harkens back to the way we defined it—by evaluating Ad_G at the Lie algebra. The algebra is closed under the bracket, or *bracket closed*. The bracket relations of any Lie algebra completely define it as a Lie algebra. Any two Lie algebras with the same bracket relations are isomorphic as Lie algebras.

If you'd like to understand this better, some representation theory is necessary— Ad_G 's real name is the *adjoint representation* of G on the tangent space at its identity because $Ad_G : G \rightarrow T_e G$ is a homomorphism. The real reason we care about the bracket is because it takes an equivalent form no matter which representation we work with, and, slightly more generally, any two isomorphic Lie algebras share an equivalent bracket (although the Lie groups they lift to may not necessarily be isomorphic). An example of isomorphic Lie algebras whose corresponding Lie groups are not isomorphic is $\mathfrak{su}(2) \cong \mathfrak{so}(3)$. We will make another comment on this in section IV. See [1] for a more complete development of the bracket in the context of representation theory.

A more general definition for a Lie algebra is any vector space with any anti-symmetric bilinear product that satisfies a *Jacobi identity*. Refer to [2] for a fully comprehensive description of the representation theory of finite and Lie groups.

4.4. $SU(2)$ and $\mathfrak{su}(2)$

So far, we've covered all the ideas of this section without the helpful context of a familiar closed subgroup $G \subseteq GL_n \mathbb{C}$. We forfeit generality now, as we use the concepts just defined to analyze the specific Lie group $SU(2)$ and its Lie algebra $\mathfrak{su}(2)$ as a subgroup of $GL_2(\mathbb{C})$ and a subspace of $\mathfrak{gl}_2 \mathbb{C}$ respectively.

A general element g of $GL_2(\mathbb{C})$ is any matrix of the form

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \text{ for } a, b, c, d \in \mathbb{C} \text{ such that } \det |g| = ad - bc \neq 0.$$

Now, recall our definition of the special unitary group from section I and write

$$SU(2) \equiv \{ \text{all } \Omega \in GL_2(\mathbb{C}) : \det|\Omega| = 1 \text{ and } \Omega^\dagger = \Omega^{-1} \}.$$

Using this definition, we can write the general form of a 2×2 complex matrix Ω and state the requirements of $SU(2)$ that must be satisfied by said matrix in terms of its entries.

$$\Omega = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \text{ any } a, b, c, d \in \mathbb{C}, \text{ satisfying the determinant condition } ad - bc = 1,$$

and the unitary condition, $\Omega^\dagger = \begin{pmatrix} a^* & c^* \\ b^* & d^* \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1}.$

Multiplying together $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} a^* & c^* \\ b^* & d^* \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ gives 4 equations:

$$aa^* + bb^* = 1, ca^* + db^* = 0, ac^* + bd^* = 0, cc^* + dd^* = 1.$$

These equations can be satisfied if $a = d^*$ and $b = -c^*$, which restricts the eight choices of real numbers we had for elements of $GL_2(\mathbb{C})$ with any $a, b, c, d \in \mathbb{C}$ to just four choices. We currently have $a = d^* = \alpha + i\beta$ and $b = -c^* = \gamma + i\lambda$ for some $\alpha, \beta, \gamma, \lambda \in \mathbb{R}$, along with the still-to-be-verified determinant condition that says $\alpha^2 + \beta^2 + \gamma^2 + \lambda^2 = 1$. The determinant condition restricts the four possibilities to just three. Since if we write it as $\alpha^2 = 1 - \beta^2 - \gamma^2 - \lambda^2$, it's clear three of the real parameters completely define the fourth. This is how we know that $SU(2)$ is, as a manifold, real 3-dimensional.

Now we define the Lie algebra $\mathfrak{su}(2)$ of $SU(2)$. If we assert that a general element of a Lie group has determinant equal to one, then a general element of its Lie algebra is traceless. To see this we use the property, mentioned previously, that if A is a diagonal matrix, $\det(e^A) = e^{tr(A)}$.

For some element $g = e^{tx} \in SU(2)$ generated by $x \in \mathfrak{su}(2)$,

$$\det(g) = \det(e^{tx}) = e^{t \operatorname{tr}(x)} = 1.$$

If we want to express this condition for the group $SU(2)$ as a condition for $\mathfrak{su}(2)$, we simply map it to the Lie algebra in the usual way:

$$D_t [\det|g|=1]_{t=0} \rightarrow D_t [e^{t \operatorname{tr}(x)}]_{t=0} = D_t [1]_{t=0} \rightarrow \operatorname{tr}(x) = 0.$$

We can now readily define $\mathfrak{su}(2)$ as

$$\mathfrak{su}(2) \equiv \{ x \in \mathfrak{gl}_2(\mathbb{C}) : x^\dagger = -x \text{ and } \operatorname{tr}(x) = 0 \}.$$

Now, in a similar manner to how we analyzed the general form for the elements of the group $SU(2)$, we can look for a general form of $\mathfrak{su}(2)$'s elements by explicitly constraining general elements of $\mathfrak{gl}_2(\mathbb{C})$.

For $x \in \mathfrak{su}(2)$, $x = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ where $a, d \in \mathbb{R}\{i\}$ and $b, c \in \mathbb{C}$ where

$a + d = 0$ and $b = -c^*$. The notation $\mathbb{R}\{i\}$ denoting the set to which the diagonal entries a, d belong means *pure complex*, which is to say they are any real multiple of the pure complex number i . If we combine all these conditions:

$x = \begin{pmatrix} i\gamma & -\alpha + i\beta \\ \alpha + i\beta & -i\gamma \end{pmatrix} = \begin{pmatrix} i\gamma & -z^* \\ z & -i\gamma \end{pmatrix}$ where $\gamma \in \mathbb{R}$ and $z \in \mathbb{C}$. Which means we can write any element of $\mathfrak{su}(2)$ as a real linear combination of three (vectors) matrices $\sigma_i \in \mathfrak{gl}_2(\mathbb{C})$ $i=1,2,3$.

As stated before, the number of real parameters you can freely pick simultaneously in the entries of general group elements is equal to the dimension of the real manifold. This means that the tangent space at the identity of $SU(2)$ should be spanned by exactly three linearly independent vectors. Not to go into this too much, but when we pick the specific vectors to do this spanning (more accurately, we pick the vector space that they transform), we pick the specific *representation* of the algebra we choose to work with (and since the group is generated by the algebra, the representation of the group is dictated by this choice). However, we already know very well the most common representation. Often called the *fundamental representation* of $\mathfrak{su}(2)$ by physicists, the Pauli matrices are a real vector space which is one valid way of representing the tangent space at the identity of the real smooth manifold $SU(2)$.

The Pauli matrices are $\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$, $\sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

Define $\mathfrak{su}(2) = \left\{ T_x = \frac{1}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, T_y = \frac{1}{2} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, T_z = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \right\}$.

We stick a factor of $\frac{1}{2}$ on them simply because it makes the bracket relation nicer, *i.e.*, $[\sigma_x, \sigma_y] = 2i\sigma_z$, $[\sigma_z, \sigma_y] = -2i\sigma_x$ and other combinations become like $[T_x, T_y] = iT_z$.

We write the general relation for this representation of $\mathfrak{su}(2)$ as $[T_i, T_j] = i\epsilon_{ijk}T_k$. The Levi-Civita symbol ϵ_{ijk} (which you've probably seen before) is anti-symmetric in all three indices (0 if two or more indices are the same and either 1 or -1 depending on the order of the indices).

One might notice that the Pauli matrices do not satisfy the definition of the Lie algebra $\mathfrak{su}(2)$ that we identified. In fact, they behave as $-i$ times the matrices in the definition. There are two explanations for this, one more intricate than the other, but both involve realizing that we defined Lie algebras as real vector spaces and therefore multiplying by $-i$ technically isn't allowed. The less intricate explanation is simply that physicists prefer a representation of $\mathfrak{su}(2)$ that is both Hermitian and unitary—also, it's convenient to use a representation of $\mathfrak{su}(2)$ that must be multiplied by i explicitly when exponentiated to an element of $SU(2)$ (we will see why in the next section). The more intricate explanation is that if we *complexify* $\mathfrak{su}(2)$, which is to take $\mathbb{C} \otimes_{\mathbb{R}} \mathfrak{su}(2) = \mathfrak{su}(2) + i\mathfrak{su}(2) \equiv \mathfrak{sl}(2, \mathbb{C})$, the “pure complex” part of $\mathfrak{sl}(2, \mathbb{C})$ is the negation of the Pauli matrices and is of course still isomorphic to $\mathfrak{su}(2)$. Taking the pure complex part of $\mathfrak{sl}(2, \mathbb{C})$ as our representation of $\mathfrak{su}(2)$ is totally legal provided we multiply by $-i$ before exponentiation (otherwise we end up in $SL(2, \mathbb{C})$). For our purposes, this still just amounts to the first, less intricate explanation that we do it for convenience. However, *complexification* is worth mentioning as it's a powerful idea that comes

up often in studying algebraic structures relevant to particle physics and we will briefly comment on it again in the last section.

As stated before, the bracket completely defines the Lie algebra. Any other Lie algebra that has a general bracket equivalent to $\mathfrak{su}(2)$'s, is isomorphic to $\mathfrak{su}(2)$ as a Lie algebra. It's implied by asserting the bracket be the same that the two algebras have the same number of elements. This concludes our study of Lie theory as we have covered as much as is needed to write down a gauge theory using any Lie group whose Lie algebra we are given explicitly.

5. Gauge Symmetry

Without getting too ahead of ourselves, a gauge theory is, in some sense, just a theory that's unchanged by the action of a Lie group. In fact, once we've developed a working definition of a gauge theory, we will use the term *gauge group* to mean nothing other than the specific Lie group that a specific theory is invariant under. However, that's not quite the whole story, as we are interested in a specific action called *local action*, but the previously stated idea is a good starting point. This section primarily serves to unite the ideas discussed in the sections on matrix groups and Lie theory with the machinery of particle physics.

5.1. Global Symmetry

For the remainder of this article, when it is relevant, we work with the Minkowski metric $\eta_{\mu\nu}$ with signature $(+1, -1, -1, -1)$. We follow the convention that Greek indices, such as $\mu = 0, 1, 2, 3$ are spacetime indices. To index other sets or spaces, we use Latin indices.

Consider the following Lagrangian (density) for a massive complex scalar field of the form $\psi(x) = \varphi_1(x) + i\varphi_2(x)$ on an open subset U of Minkowski space \mathcal{M} , which is a map $\psi : U \rightarrow \mathbb{C}$ parameterized by two real scalar φ_i 's. Each real scalar field is of course a map $\varphi_i : U \rightarrow \mathbb{R}$.

$$\mathcal{L} = \partial_\mu \psi^* \partial^\mu \psi - m\psi^* \psi$$

Here, ψ^* is the adjoint (complex conjugate) of ψ . It's easy to see that if we act on the field with what are called *global* elements of the Lie group $U(1)$, which are just elements of the form $e^{i\alpha} \in U(1)$ where $\alpha \in \mathbb{R}$ (action on ψ^* is with the conjugate transformation), the Lagrangian and the equations of motion are unchanged. Written explicitly:

$$\begin{aligned} \text{If } \psi \rightarrow \psi' = e^{i\alpha} \psi \text{ and } \psi^* \rightarrow \psi'^* = \psi^* e^{-i\alpha}, \text{ then} \\ \mathcal{L} \rightarrow \mathcal{L}' = \partial_\mu (\psi^* e^{-i\alpha}) \partial^\mu (e^{i\alpha} \psi) - m\psi^* e^{-i\alpha} e^{i\alpha} \psi \\ = \partial_\mu \psi^* (e^{-i\alpha} e^{i\alpha}) \partial^\mu \psi - m\psi^* (e^{-i\alpha} e^{i\alpha}) \psi = \mathcal{L}. \end{aligned}$$

Since the group is one-dimensional and unitary, the conjugates of elements are their inverses. Since the transformations are global (which just means the phase of the transformation $\alpha \in \mathbb{R}$ is a constant) they commute through the derivatives and $e^{-i\alpha} e^{i\alpha} = 1$ is the identity of the group $U(1)$. This means the Lagrangian

is unchanged. If the Lagrangian is unchanged by some symmetry transformations, so are the equations of motion. We might call a Lagrangian theory that's unchanged by a global U(1) symmetry globally U(1)-invariant or, for an arbitrary group G, globally G-invariant. It's important to realize the necessity in ψ being a complex one component vector. Without the notion of a complex conjugate of ψ , there would be no conjugate (inverse) transformations carried out and the Lagrangian wouldn't be invariant.

Let's look at a theory with a non-Abelian global symmetry. Consider the following theory for a two-component spinor $\psi_i(x) = \begin{pmatrix} \psi_1(x) \\ \psi_2(x) \end{pmatrix}$ (a vector in \mathbb{C}^2 with each component a complex scalar field) which is globally SU(2)-invariant.

$$\mathcal{L} = \partial_\mu \psi^\dagger_i \partial^\mu \psi^i - m \psi^\dagger_i \psi^i \quad \text{for } i=1,2 \text{ the components of the spinor.}$$

If we transform the field with the global $\Omega^i_j = (e^{i\alpha_a T_a})^i_j \in \text{SU}(2)$ where $T_a = \frac{1}{2} \sigma_a$ $a=1,2,3$ are $\frac{1}{2}$ the Pauli matrices and $\alpha_a \in \mathbb{R}$ are constants, the same sequence of events as unfolded with the global U(1) theory unfolds here as well.

$$\begin{aligned} \psi^i &\rightarrow \Omega^i_j \psi^j \quad \text{and} \quad \psi^\dagger_i \rightarrow \psi^\dagger_j [\Omega^i_j]^{-1} = \psi^\dagger_j [\Omega^j_i]^* \\ \mathcal{L} &\rightarrow \mathcal{L}' = \partial_\mu \left(\psi^\dagger_j [\Omega^j_i]^* \right) \partial^\mu \left(\Omega^i_j \psi^j \right) - m \psi^\dagger_j [\Omega^j_i]^* \Omega^i_j \psi^j = \mathcal{L} \end{aligned}$$

Notice that at the particle level the only change we made to the U(1) theory to make it an SU(2) theory was to give the field another complex component to make it an element of the vector space acted on by the previously defined two-dimensional representation of SU(2).

We can write a globally SO(3) invariant theory for a vector $\varphi^i = \begin{pmatrix} \varphi_1 \\ \varphi_2 \\ \varphi_3 \end{pmatrix} \in \mathbb{R}^3$

whose components are real scalar fields. This vector (often called a triplet since it's not a spacetime vector) transforms under the real 3×3 rotation matrices whose transpose is their inverse (orthogonal). This is, in the grand scheme of things, just one 3-dimensional real representation of this group.

$$\mathcal{L} = \partial_\mu \varphi_i \partial^\mu \varphi^i - m \varphi_i \varphi^i$$

We again choose the most familiar representation of SO(3)'s Lie algebra. Often called generators of the 3-dimensional rotation group, the three Lie algebra elements are

$$T_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad T_2 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \quad T_3 = \begin{pmatrix} 0 & i & 0 \\ -i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

A neat fact is that SU(2) double covers SO(3) and their algebras are isomorphic. All that's necessary to check this, is to check the bracket relations are isomorphic.

The reader is encouraged to do this. If we define a general element of $SO(3)$ as $R = e^{i\alpha_a T_a}$ for $\alpha_a \in \mathbb{R}$ $a = 1, 2, 3$, the same scheme of invariance manifests itself for this Lagrangian of three real scalar fields.

A simple but potentially illuminating exercise is to show that a globally $SO(2)$ -invariant theory for the two component real vector $\varphi^i = \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix}$ is completely equivalent to the above complex scalar field theory for $\psi(x) = \varphi_1(x) + i\varphi_2(x)$ that's invariant under $U(1)$ transformations. In other words, the one-dimensional complex representation of $U(1)$ on \mathbb{C} is an isomorphic structure to the two-dimensional real representation of $SO(2)$ on \mathbb{R}^2 . This is a direct consequence of the repeatedly mentioned isomorphism: $\mathbb{C} \cong \mathbb{R}^2$.

It's important to recognize that spacetime in all these theories is just the underlying or *base* manifold. The fields take in a domain of spacetime coordinates but their images live in different vector spaces *above* spacetime that relevant group actions are allowed to transform. This will be cleared up soon.

5.2. Local Symmetry

After spending some time with theories of fields exhibiting invariance under global action of Lie groups, one may naturally arrive at the following question. What happens if the parameters α_a are no longer constant but are instead allowed to vary at different points in spacetime? This question is natural because the fields in the Lagrangian are functions of spacetime—so, why not the symmetry transformations of these fields as well?

Consider the complex scalar field again.

$$\mathcal{L} = \partial_\mu \psi^* \partial^\mu \psi - m\psi^* \psi$$

Define a local $U(1)$ transformation as $e^{i\alpha(x)} \in U(1)$ where $\alpha(x)$ is now a real scalar function of spacetime. If we transform the complex field in the usual way,

$$\psi \rightarrow \psi' = e^{i\alpha} \psi \text{ and } \psi^* \rightarrow \psi'^* = \psi^* e^{-i\alpha}$$

the transformed Lagrangian is,

$$\begin{aligned} \mathcal{L} \rightarrow \mathcal{L}' &= \partial_\mu (\psi^* e^{-i\alpha}) \partial^\mu (e^{i\alpha} \psi) - m\psi^* e^{-i\alpha} e^{i\alpha} \psi \\ &= \left[(\partial_\mu \psi^*) e^{-i\alpha} + \psi^* (-i\partial_\mu \alpha) e^{-i\alpha} \right] \left[(e^{i\alpha}) \partial^\mu \psi + (i\partial^\mu \alpha) e^{i\alpha} \psi \right] - m\psi^* \psi \neq \mathcal{L}. \end{aligned}$$

We find the theory not to be invariant under $U(1)$ transformations that are local on spacetime. However, we are not without hope. There is a surprising strategy we can use to achieve local invariance. The strategy is to add a new field into the Lagrangian. This new field is a vector field over spacetime, call it A_μ , and enters the Lagrangian as part of a new derivative we will swap with the old derivative. We replace the derivative ∂_μ with what's called a gauge covariant derivative (GCD) of the form $D_\mu = \partial_\mu + iqA_\mu$. Here, q is a real number serving as a coupling constant, which in physics is often interpreted as the strength of the interaction between the field A_μ and the complex scalar. We assert that A_μ transforms under the adjoint action of the $U(1)$ group as

$$A_\mu \rightarrow A'_\mu = e^{i\alpha} A_\mu e^{-i\alpha} + \frac{i}{q} (\partial_\mu e^{i\alpha}) e^{-i\alpha} = A_\mu - \frac{1}{q} \partial_\mu \alpha.$$

We can make the above (right-most equals sign) simplification because U(1) is Abelian. This simplification should look familiar, as it is nothing but the spacetime analogue of the gauge freedom often mentioned when defining the magnetic field as the curl of a vector potential in a standard undergraduate course on electromagnetism. Replacing derivatives in the Lagrangian with GCD's, we have

$$\mathcal{L} = (D_\mu \psi)^* (D^\mu \psi) - m \psi^* \psi = (\partial_\mu - iqA_\mu) \psi^* (\partial^\mu + iqA^\mu) \psi - m \psi^* \psi.$$

To demonstrate the effectiveness of this configuration, all we must do is locally transform every field in the Lagrangian. The mass term is trivially invariant. We only need to check the kinetic term. Transforming and then evaluating each derivative term individually, we get

$$\begin{aligned} (D_\mu \psi)^* &= (\partial_\mu - iqA'_\mu) \psi^{*'} = \left[\partial_\mu - iq \left(A_\mu - \frac{1}{q} \partial_\mu \alpha \right) \right] \psi^* e^{-i\alpha} \\ &= \left[(\partial_\mu \psi^*) e^{-i\alpha} + \psi^* (-i \partial_\mu \alpha) e^{-i\alpha} \right] - iqA_\mu \psi^* e^{-i\alpha} + i (\partial_\mu \alpha) \psi^* e^{-i\alpha} \\ &= (D_\mu \psi)^* e^{-i\alpha} \end{aligned}$$

$$\begin{aligned} (D^\mu \psi)' &= (\partial^\mu + iqA'^\mu) \psi' = \left[\partial^\mu + iq \left(A^\mu - \frac{1}{q} \partial^\mu \alpha \right) \right] e^{i\alpha} \psi \\ &= \left[(e^{i\alpha}) \partial^\mu \psi + (i \partial^\mu \alpha) e^{i\alpha} \psi \right] + iqA^\mu e^{i\alpha} \psi - i (\partial^\mu \alpha) e^{i\alpha} \psi \\ &= e^{i\alpha} (D^\mu \psi). \end{aligned}$$

$$\text{Then } \mathcal{L} \rightarrow \mathcal{L}' = (D_\mu \psi)^* e^{-i\alpha(x)} e^{i\alpha(x)} (D^\mu \psi) - m \psi^* \psi = (D_\mu \psi)^* (D^\mu \psi) - m \psi^* \psi = \mathcal{L}.$$

By trading the derivative out for the U(1) gauge covariant version, we have successfully made our theory invariant under local U(1) transformations. To clarify this once more, the terminology of global vs. local symmetry refers to whether the actions of the Lie group are the same at every point in spacetime or differ from point to point. When we talk about a gauge theory with a gauge symmetry, we are always talking about a local symmetry.

Now, the theory of a complex scalar field is fun to play with, but it doesn't show up in the standard model. Let's look for a U(1) symmetry in a physical theory.

Consider the Dirac theory of the electron

$$\mathcal{L}_{Dirac} = i \bar{\psi} \gamma^\mu \partial_\mu \psi - m \bar{\psi} \psi.$$

It is trivially checked that this theory is globally U(1)-invariant. Here, ψ is a *Dirac spinor* and $\bar{\psi}$ is its *Dirac adjoint* (not quite as simple as the Hermitian conjugate, but it transforms the same) and m is the mass of these spinor fields $\bar{\psi}, \psi$. The four matrices γ^μ are the Dirac matrices where $\mu = 0, 1, 2, 3$ indexes spacetime.

The Dirac matrices are the 4×4 complex matrices $\gamma^0 = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}$ and

$\gamma^j = \begin{pmatrix} 0 & \sigma_i \\ -\sigma_i & 0 \end{pmatrix}$ for $i = 1, 2, 3$ where σ_i are the Pauli matrices and I is the 2×2 identity.

Perhaps you've seen this Lagrangian and the Dirac matrices before. If so, what we have written is not too mysterious. If not, there is a beautiful connection to be understood here regarding the Lorentz algebra, $\mathfrak{so}(1,3)$, and two non-interacting $\mathfrak{su}(2)$ subalgebras. This relationship will not be discussed in this article on gauge theories, but here's a hint for the ambitious reader:

$$\mathfrak{so}(1,3) \cong \mathfrak{so}(4, \mathbb{R}) \cong \mathfrak{sl}(2, \mathbb{C}) \cong \mathfrak{su}(2) \oplus \mathfrak{su}(2).$$

We don't need to discuss this relationship because the only thing forcing us to acknowledge the Dirac spinor as anything other than a complex scalar is the contraction of the derivative with the Dirac matrices. From the gauge theory perspective, if we want to make this theory U(1) invariant, we need only think of the Dirac spinor as a single component complex vector in the U(1) transformation space. In other words, through the eyes of the U(1) gauge group, the Dirac field is just a complex number. Knowing this, the whole point is to achieve local symmetry, so we replace the derivative with a U(1)—gauge covariant one.

$$\mathcal{L}_{\text{Dirac-Gauged}} = i\bar{\psi}\gamma^\mu D_\mu\psi - m\bar{\psi}\psi = i\bar{\psi}\gamma^\mu (\partial_\mu + iqA_\mu)\psi - m\bar{\psi}\psi$$

So far everything fits together mathematically, but let's think about the physical implications of our actions. We've taken the relativistic theory of the electron, which happened to be globally U(1) invariant, and made it locally invariant through substitution of the gauge covariant derivative. In doing this, however, we added a new field to the Lagrangian. The presence of this new field should be able to be interpreted physically. One aspect of this is that if the Dirac field were suddenly "turned off", the Lagrangian should collapse down into that of the free field A_μ . As we have it now, turning off the Dirac field gives $\mathcal{L} = 0$. What we need is a kinetic term for A_μ that is gauge invariant. By way that A_μ transforms, we don't have many options and it's easy to see the correct structure for the momentum should be a differential two-form/anti-symmetric two-tensor field. This is a tensor field of the form $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. The conventional kinetic term for A_μ is $\mathcal{L}_{\text{Kinetic}}^A = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$.

We call the tensor $F_{\mu\nu}$ the field strength. An alternative way to arrive at this form for the correct derivative of A_μ is via the bracket of gauge covariant derivatives:

$$[D_\mu, D_\nu] = iqF_{\mu\nu}.$$

Adding this term to the gauged Dirac theory gives immediately the theory of *Quantum Electrodynamics*, or QED.

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

Now that we have the context of QED, the fields are quantum, and we can refer

to the vector field A_μ by its physics name—the photon. Like magic, requiring that the Dirac theory be not just globally but locally $U(1)$ invariant, led us directly to writing down essentially the entire electromagnetic sector of the SM and $\frac{1}{4}$ of known interactions. The real star of the show here is the interaction particle; this is the vector field we added to the system by adding the GCD. This vector field, and the fact that it transforms adjointly, is the catalyst for the local invariance which we will now refer to as gauge invariance.

6. Yang Mills and General Gauge Theories

Now that we've seen the process of making a global $U(1)$ theory into a local one, you're probably wondering what this looks like for general Lie groups. $U(1)$ was indeed a very special case because the group is Abelian. The generalization to non-Abelian groups isn't exactly trivial, which is why C.N. Yang and Robert Mills famously have their names attached to doing so. However, there is a very beautiful pattern here and we are going to do our best to lay it out clearly. To do so, it should first be defined exactly what we are doing when we posit a gauge theory. This is, in all ways, the climax of this article. Everything we have defined thus far has built up to this moment. So, without further ado...

6.1. The Recipe

The recipe for a standard gauge theory on a flat Minkowski spacetime is as follows. Take an open subset $U \subseteq \mathcal{M}$, with spacetime coordinates denoted $x \in U$. Let G be a Lie group, $T_e G$ its Lie algebra, and V a real or complex vector space on which a representation of G and $T_e G$ can act. Define a “scalar function” on spacetime as a map $\Phi: U \rightarrow V$. Under a local gauge transformation, $\gamma(x): U \rightarrow G$, Φ is transformed as $\Phi' \rightarrow \rho(\gamma(x))\Phi$, where $\rho: G \rightarrow GL(V)$ is the relevant *representation* on V [3].

Here, scalar function is in quotations because, by the definition of the map, Φ is an honest scalar function on spacetime, but it's not in the regular way (like say a map to \mathbb{R}), since it's a vector valued function in the space that we defined to be acted on by gauge transformations.

A comment on the representation map (ρ): A representation $\rho: G \rightarrow GL(V)$ is a homomorphism from an abstract group to invertible linear transformations of some vector space V . The vector space the group elements transform is defined to be precisely the same vector space in which our particle(s) given by $\Phi: U \rightarrow V$ live. We hinted at this throughout section five of the article. For instance, in choosing the complex one-dimensional representation of the $U(1)$ gauge group, which is to say $\rho: U(1) \rightarrow GL_1\mathbb{C}$, we acknowledged that our particle must minimally be a map $\Phi: U \rightarrow \mathbb{C}$ (a complex scalar field). In the global $SO(3)$ theory, we chose a real three-dimensional representation $\rho: SO(3) \rightarrow GL_3\mathbb{R}$ and thus, necessarily the particles were given by $\Phi: U \rightarrow \mathbb{R}^3$ (a triplet of real scalar fields). In the case of matrix Lie groups, where group elements are defined as the exponential of a matrix in the Lie algebra, the dimensions of group representations are defined by

(equal to) the dimension of the representation of the corresponding algebra. As we've stated before, Lie groups/algebras have many representations. In this article so far, every Lie algebra we have explicitly written has been written in what is usually called the fundamental representation—which are the convention/standard of most physics' literature. The fundamental representation of a Lie algebra can roughly be defined as a representation on a vector space of *natural* dimension and is such that one must multiply by the complex number $i = \sqrt{-1}$ when exponentiating to an element of the group. It is hard to provide a general, logical description of what dimension is considered “natural”, but for the groups we have considered in this article—namely the (special) unitary and orthogonal groups—we claim the natural dimension is N complex dimensions for any $SU(N)$ and N real dimensions for $SO(N)$. To clarify once more, the dimension of a representation is the dimension of the space it acts on (as opposed to the dimension of the manifold/tangent spaces on the manifold)—which for $SU(2)$ in the fundamental representation is two complex dimensions.

Now that this stage has been set and our mission made clearer, let us localize the globally $SU(2)$ -invariant theory of a two-component spinor discussed previously. The Lagrangian of the theory was

$$\mathcal{L} = \partial_\mu \psi^\dagger_i \partial^\mu \psi^i - m \psi^\dagger_i \psi^i \quad \text{for } i = 1, 2.$$

As you might've noticed before, we've written the Latin indices as covariant/contravariant. This is purely for convenience as we will soon have many indices and maybe will want to change their script to tidy up our terms. If we imagine these indices as running over components of vectors in some space with Euclidean metric (which is here dimension 2), the regular summation convention is applied and there is no fault.

If we transform the spinor fields with local gauge transformations of the form

$$\Omega^i_j = \left(e^{i\alpha_a T_a} \right)^i_j \in SU(2) \quad \text{where } T_a = \frac{1}{2} \sigma_a \quad \text{and } \alpha_a \equiv \alpha_a(x) \quad \text{for } a = 1, 2, 3$$

we run into the same problem as before. Again, the invariance of the mass term is trivial, so we only need to consider the kinetic term.

$$\begin{aligned} \partial^\mu \left(\Omega^i_j \psi^j \right) &= \Omega^i_j \left(\partial^\mu \psi^j \right) + \left(\partial^\mu \Omega^i_j \right) \psi^j \\ \partial_\mu \left(\psi^\dagger_j \Omega^{j*}_i \right) &= \left(\partial_\mu \psi^\dagger_j \right) \Omega^{j*}_i + \psi^\dagger_j \left(\partial_\mu \Omega^{j*}_i \right) \\ \mathcal{L}_{Kinetic} &\rightarrow \mathcal{L}'_{Kinetic} = \left[\left(\partial_\mu \psi^\dagger_j \right) \Omega^{j*}_i + \psi^\dagger_j \left(\partial_\mu \Omega^{j*}_i \right) \right] \left[\Omega^i_j \left(\partial^\mu \psi^j \right) + \left(\partial^\mu \Omega^i_j \right) \psi^j \right] \end{aligned}$$

We already have a general idea of the way to fix this problem. We need to find a gauge covariant derivative that includes something like a vector field that transforms adjointly in such a way as to cancel the factors of the form $\left(\partial^\mu \Omega^i_j \right) \psi^j$. The key realization here is that to accomplish this we need more than one vector field. In fact, we need exactly as many vector fields as there are basis elements of the Lie algebra. As we already know, a Lie group has infinitely many elements, however

its manifold dimension can be finite—which is usually the only dimension we ever refer too when talking about a Lie group. The dimension of the manifold is equivalent to how many orthogonal tangent vectors it has at every point. The number of orthogonal tangent vectors at every point is equal to the dimension of the tangent spaces at every point. The Lie algebra forms the tangent space at one of these points (the identity). All we are saying is that the Lie group is a structure of continuous symmetries, and we can identify the orthogonal basis of a given tangent space of the group with linearly independent directions of continuous symmetries. To account for this in general, we need as many vector fields as there are orthogonal basis vectors in the tangent spaces on whatever group we’re referring to. SU(n), for example, requires $n^2 - 1$ vector fields. For a more technical but directly related reason (see literature on *principal fiber bundles* and *connection one-forms*), the vector fields are often called *Lie algebra-valued*, or in the context of a gauge theory, they are simply called gauge fields. Let’s see how we do this.

Suppose we have in our possession a field theory that is globally invariant under an m -dimensional representation (the matrices that span the Lie algebra are $m \times m$) of some n -dimensional Lie group (n matrices span the Lie algebra). To achieve local invariance, define vector fields X_μ^r to be Lie algebra valued and τ_r for $r = 1, \dots, n$ the elements of the Lie algebra. The Lie algebra valued index r can be contracted and summed over. The general gauge covariant derivative we need is this:

$$D_\mu^{ij} = \delta^{ij} \partial_\mu + ig \tau_r^{ij} X_\mu^r = \delta^{ij} \partial_\mu + ig (\tau_1^{ij} X_\mu^1 + \tau_2^{ij} X_\mu^2 + \dots + \tau_n^{ij} X_\mu^n)$$

The indices $i, j = 1, \dots, m$ index the matrix components of the τ_r ’s as well as the multiplet components (scalar functions that are transformable by the group).

The Lie algebra valued fields $\tau_r X_\mu^r$ transform adjointly under a local gauge transformation $\gamma(x) = e^{i\tau_r \theta^r(x)}$ for $\theta^r(x)$ real scalars as

$$ig \tau_r X_\mu^r \rightarrow (ig \tau_r X_\mu^r)' = ig \left[\gamma \tau^r X_\mu^r \gamma^{-1} + \frac{i}{g} (\partial^\mu \gamma) (\gamma)^{-1} \right].$$

6.2. Our Old Friend SU(2)

Let’s check this explicitly for the above SU(2) theory. This obviously isn’t necessary, but we might want to rename the vector fields to differentiate them from the other theories. Let’s label the three vector fields required for local SU(2) symmetry W_μ^a for $a = 1, 2, 3$. Inserting gauge covariant derivatives, we have

$$\begin{aligned} \mathcal{L}_{SU(2)\text{-Gauged}} &= (D_\mu^{ij} \psi_j)^\dagger D_\mu^\mu \psi^j - m \psi^\dagger \psi \quad \text{for } i, j = 1, 2 \\ (D_\mu^{ij} \psi_j)^\dagger &= (\delta^{ij} \partial_\mu - ig T_a^{ij} W_\mu^a) \psi_j^\dagger \\ D_\mu^\mu \psi^j &= (\delta_{ij} \partial^\mu + ig T_{ij}^a W_\mu^a) \psi^j \end{aligned}$$

Since we’ve seen this before with the U(1) case, we need only check explicitly one derivative. For a gauge transformation $\Omega \in SU(2)$

$$\begin{aligned}
 D^\mu \psi &\rightarrow (D^\mu \psi)' = \left(\partial^\mu + ig \left[\Omega T^a W_a^\mu \Omega^{-1} + \frac{i}{g} (\partial^\mu \Omega) (\Omega)^{-1} \right] \right) \Omega \psi \\
 &= \left[\Omega (\partial^\mu \psi) + (\partial^\mu \Omega) \psi \right] + ig \Omega T^a W_a^\mu (\Omega)^{-1} \Omega \psi - (\partial^\mu \Omega) (\Omega)^{-1} \Omega \psi \\
 &= \Omega (\partial^\mu \psi) + ig \Omega T^a W_a^\mu \psi = \Omega (D^\mu \psi)
 \end{aligned}$$

Similarly, $(D_{\mu}^{ij} \psi_j)^\dagger \rightarrow (D_{\mu}^{ij} \psi_j)^\dagger \Omega^{-1}$, and we see at the Lagrangian level that

$$\mathcal{L} \rightarrow \mathcal{L}' = (D_{\mu}^{ij} \psi_j)^\dagger \Omega^{-1} \Omega (D_{ij}^{\mu} \psi^j) - m \psi^\dagger \psi = (D_{\mu}^{ij} \psi_j)^\dagger D_{ij}^{\mu} \psi^j - m \psi^\dagger \psi = \mathcal{L}$$

Thus, confirming gauge invariance. We saw before the U(1) gauge field transformed adjointly, and the SU(2) fields did the same here.

$$T_a^{ij} W_\mu^a \rightarrow (T_a^{ij} W_\mu^a)' = \Omega T_a^{ij} W_\mu^a (\Omega)^{-1} + \frac{i}{g} (\partial^\mu \Omega) (\Omega)^{-1}$$

If we want a kinetic term for the gauge field, all we need to do is commute the GCD.

$$[D_\mu, D_\nu] = ig W_{\mu\nu}^a \quad \text{where } W_{\mu\nu}^a = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a + ig \epsilon^a_{bc} [W_\mu^b, W_\nu^c]$$

We now have the tools at our disposal to write down a gauge theory that's gauge invariant under the gauge symmetry of our liking.

6.3. Gauge Theories of the Standard Model

Before we wrap this section up, we should write down a couple other gauge theories relevant to the standard model. *Quantum chromodynamics* is the SU(3) invariant theory of the strong interaction. SU(3) is an eight-dimensional manifold whose Lie algebra elements, in the fundamental representation, are often called the Gell-Mann matrices.

$$\begin{aligned}
 \lambda_1 &= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \\
 \lambda_5 &= \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \quad \lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad \lambda_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}
 \end{aligned}$$

These eight matrices correspond to eight vector fields, called gluons, that mediate the strong interaction. We will label these gauge fields G_μ^b for $b=1, \dots, 8$. The Lagrangian is,

$$\mathcal{L}_{QCD} = i \bar{q}^\dagger \gamma^\mu D_\mu^j q - m \bar{q}^\dagger q + G_{\mu\nu}^b G_b^{\mu\nu}$$

where $D_{ij}^\mu q^j = (\delta^{ij} \partial_\mu + ig \tau_b^{ij} G_\mu^b) q^j$, $\tau_b = \frac{1}{2} \lambda_b$ and $i, j = 1, 2, 3$.

The q^j is a triplet of Dirac spinors (which could be inferred by the presence of γ^μ) which correspond to three flavors of quark/antiquark. The coupling constant g represents the color charge—the QCD analogue to electric charge in QED.

The other gauge theory in the SM, *quantum electroweak theory*, is a real piece of work—so we won't go into it much here. However, it's different from QED and QCD in the way that (before symmetry is broken anyway) it has a double-gauge group—it's simultaneously SU(2) and U(1) gauge invariant. You see, back in the good ol' days, around $10^{(-12)}$ seconds after the big bang, before the Higgs field and its mysterious mechanism spontaneously broke SU(2) symmetry, the universe was fully SU(3) × SU(2) × U(1) gauge invariant and the three gauge bosons (W^\pm , Z^0) were massless. Just to give you an idea of what a two-fold gauge-group theory looks like, here's an example of an SU(2) × U(1) invariant theory for a Dirac doublet:

$$\mathcal{L}_{\text{Not QEW}} = i\psi^\dagger_i \gamma^\mu D_\mu^j \psi_j - m\psi^\dagger_i \psi_i - \frac{1}{4} W_{\mu\nu}^a W_a^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}$$

where $D_\mu^j \psi_j = \left(\delta^{ij} \partial_\mu - \frac{1}{2} ig \sigma_a^{ij} W_\mu^a + \frac{1}{2} ig' Y^{ij} B_\mu \right) \psi_j$.

The matrix $Y^{ij} \equiv I_{2 \times 2}$ is known in the SM as weak hypercharge.

The independent gauge transformations are:

$$\underline{\text{U(1)}}: \psi \rightarrow \psi' = e^{\frac{i\alpha}{2}} \psi$$

$$\underline{\text{SU(2)}}: \psi \rightarrow \psi' = e^{\frac{i\beta^a \sigma_a}{2}} \psi$$

In the SM, the reason this set up is meaningful is largely due to the Higgs mechanism—which after breaking symmetry combines the two terms $ig T_a^{ij} W_\mu^a$ and $\frac{1}{2} ig' Y^{ij} B_\mu$ in a very specific way. The reader is encouraged to consult literature on the *Glashow-Weinberg-Salam model of the first generation of matter* for further discussion on spontaneous symmetry breaking and electroweak interactions.

7. An Extension: Vector Dark Matter

7.1. A Natural Extension

In this section of this article, we pose a straightforward extension of the traditional recipe for gauge theories stated at the beginning of section V. Extensions of this type have been recently studied in the context of “beyond the standard model” theories of dark matter [4] [5]. Because, apart from noticing a matter-like presence in galaxy clusters beyond what is visible, we haven't directly observed or detected the phenomena known as dark matter (DM), we don't really know anything about it—if it's a particle, we don't know its spin or any other quantum numbers it may carry. Even if we assume DM has a particle like description, since we don't know if it's fermionic, bosonic, or something else we haven't seen before, a wide range of guesses for the structure of its field are feasible (scalar, spinor, vector, tensor, etc.). For this reason, we extend our idea of a gauge theory to any theory containing a GCD that facilitates a local gauge symmetry, where the “scalar functions on spacetime” from before are no longer necessarily such—they could be a (multiplet of) field(s) of any spin.

7.2. The Simplest Model in This Regime

Most of these currently relevant theories are $SU(2) \times U(1)$ gauge symmetric and therefore utilize an electroweak-type GCD. There is a very good reason for this, as Nobel laureate Frank Wilczek and others have proposed that the Higgs mechanism could act as a “portal” to theorizing species of particles outside the current SM theory. In this context, DM particles are usually assumed to be weakly interacting massive particles (WIMPs). When a Lagrangian description of whatever field is selected to serve as a dark matter candidate is constructed, call it \mathcal{L}_{DM} , it’s often studied as an extension of the standard model Lagrangian. The extension \mathcal{L}_{DM} could contain various couplings to chosen fields already present in the standard model (most notably the Higgs and weak bosons) and its usually added to the standard model as $\mathcal{L}_{SM} + \mathcal{L}_{DM}$ before spontaneous symmetry breaking or other phenomena are considered. Since the Higgs mechanism often plays a crucial role in these theories and, because the Higgs field and symmetry breaking has not been discussed in this article, we look to construct a simpler model of this type to play with.

The simplest model should of course have the simplest symmetry, $U(1)$. We state the following general idea with $U(1)$ in mind—it should be known for this case the field structure is at a minimum a complex scalar.

Suppose again we have an m -dimensional representation of some n -dimensional Lie group. Rather than defining Φ a scalar function on spacetime (yet an m -vector in the group transformation space), we could define an m -fold multiplet of spacetime vector fields V_μ^i or k -tensor fields $T_{\mu_1, \dots, \mu_k}^i$ (for $i = 1, \dots, m$) whose components transform properly under the group action. This is still an m -vector in the group transformation space, but the components are now maps from spacetime coordinates to spacetime vectors or tensors.

If the group is $U(1)$, all that’s necessary is that every component of every vector/tensor in the multiplet is a map $U \rightarrow \mathbb{C}$ that transforms under $U(1)$ in the correct way.

We can formulate a theory of the kind we now wish to explore the following way. Take a Lagrangian of two non-interacting four-vector fields $A^\mu = (\varphi, \mathbf{A})$ and $G^\mu = (\xi, \mathbf{G})$

$$\mathcal{L} = -\frac{1}{4} G_{\mu\nu} G^{\mu\nu} - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} .$$

As we have it, the vector fields are non-interacting. Both terms are identical in structure to kinetic terms for photons—simply $G_{\mu\nu} = \partial_\mu G_\nu - \partial_\nu G_\mu$ and $F_{\mu\nu} = \partial_\mu F_\nu - \partial_\nu F_\mu$. To make the fields interact in a gauge covariant way, replace the derivatives in the kinetic term of one of the fields with gauge covariant derivatives where the gauge field is the other vector field already present. It becomes clear we must take the vector field “being gauged” to be complex so that elements of the $U(1)$ gauge group can act on its components. Doing this we obtain the following Lagrangian:

$$\mathcal{L} = -\frac{1}{2}\Gamma_{\mu\nu}^*\Gamma^{\mu\nu} - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}$$

where the kinetic term for G_μ is now the norm of a complex field strength tensor

$$\Gamma_{\mu\nu}^* = D_\mu^*G_\nu^* - D_\nu^*G_\mu^*$$

whose components are being “gauged” with A_μ since $D_\mu = \partial_\mu + iqA_\mu$. As before stated, the components of G_μ are necessarily made complex to transform nicely under the gauge transformations suggested by the gauge coupling. We can check that the Lagrangian is gauge invariant, and it is. The transformation rules are no different than usual:

$$G_\mu \rightarrow e^{i\alpha(x)}G_\mu$$

$$G_\mu^* \rightarrow G_\mu^*e^{-i\alpha(x)}$$

$$A_\mu \rightarrow e^{i\alpha(x)}A_\mu e^{-i\alpha(x)} - \frac{i}{q}\left[\partial_\mu e^{i\alpha(x)}\right]e^{-i\alpha(x)}$$

What we currently have in hand is a U(1) invariant theory of interacting vector fields. We can readily study the physicality of the theory by calculating, analytically and/or perturbatively, physical quantities. To conclude this section, we’ll play around with this theory calculating explicitly the Euler-Lagrange equations, as well as the conserved current corresponding to the continuous U(1) symmetry. By presenting these calculations here, we hope to give the reader who lacks confidence in manipulating these structures a guide to follow. This specific reader is encouraged to replicate these calculations for any of the other theories presented in this article, as well as any theory they can dream up and put on paper.

7.3. Euler-Lagrange Equations

It is useful to write the action for the theory.

$$S = \int d^4x \left[-\frac{1}{2}\Gamma_{\mu\nu}^*\Gamma^{\mu\nu} - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \right] \\ = -\int d^4x \left[(D_\mu^*G_\nu^*D^\mu G^\nu - D_\mu^*G_\nu^*D^\nu G^\mu) + \frac{1}{2}(\partial_\mu A_\nu \partial^\mu A^\nu - \partial_\mu A_\nu \partial^\nu A^\mu) \right]$$

Variation of the action with respect to each field gives:

$$\delta S = -\int d^4x \left\{ \frac{1}{2}\left[(\delta\Gamma_{\mu\nu}^*)\Gamma^{\mu\nu} + \Gamma_{\mu\nu}^* (\delta\Gamma^{\mu\nu}) \right] + \frac{1}{4}\left[(\delta F_{\mu\nu})F^{\mu\nu} + F_{\mu\nu} (\delta F^{\mu\nu}) \right] \right\} \\ = -\int d^4x \left\{ (D_\mu^*\delta G_\nu^*)\Gamma^{\mu\nu} + \Gamma_{\mu\nu}^* (D^\mu\delta G^\nu) - iq(\delta A_\mu)G_\nu^*\Gamma^{\mu\nu} + iq\Gamma_{\mu\nu}^* (\delta A^\mu)G^\nu \right. \\ \left. + \frac{1}{2}(\partial_\mu\delta A_\nu - \partial_\nu\delta A_\mu)F^{\mu\nu} \right\} \\ = -\int d^4x \left\{ (D_\mu^*\delta G_\nu^*)\Gamma^{\mu\nu} + \Gamma_{\mu\nu}^* (D^\mu\delta G^\nu) - iq(\delta A_\mu)\left[G_\nu^*\Gamma^{\mu\nu} - (\Gamma^{\mu\nu})^* G_\nu \right] \right. \\ \left. - (\partial_\nu\delta A_\mu)F^{\mu\nu} \right\}$$

For the A^μ field:

$$\begin{aligned} \delta S &= -\int d^4x \left\{ -iq(\delta A_\mu) \left[G_\nu^* \Gamma^{\mu\nu} - (\Gamma^{\mu\nu})^* G_\nu \right] - (\partial_\nu \delta A_\mu) F^{\mu\nu} \right\} \\ &= \int d^4x \left\{ iq(\delta A_\mu) \left[G_\nu^* \Gamma^{\mu\nu} - (\Gamma^{\mu\nu})^* G_\nu \right] + \partial_\nu \left[(\delta A_\mu) F^{\mu\nu} \right] - (\delta A_\mu) \partial_\nu F^{\mu\nu} \right\} \end{aligned}$$

The total derivative (second term) vanishes with integration and setting $\delta S = 0$ we get

$$\partial_\nu F^{\mu\nu} = iq \left[G_\nu^* \Gamma^{\mu\nu} - (\Gamma^{\mu\nu})^* G_\nu \right],$$

which is the equation for A^μ .

For the $(G^\mu)^*$ field:

$$\begin{aligned} \delta S &= \int d^4x \left\{ (D_\nu^* \delta G_\mu^*) \Gamma^{\mu\nu} \right\} \\ &= \int d^4x \left\{ \partial_\nu \left[(\delta G_\mu^*) \Gamma^{\mu\nu} \right] - iq A_\nu \left[(\delta G_\mu^*) \Gamma^{\mu\nu} \right] - (\delta G_\mu^*) \partial_\nu \Gamma^{\mu\nu} \right\} \\ &= \int d^4x \left\{ -(\delta G_\mu^*) D_\nu \Gamma^{\mu\nu} \right\} = 0 \end{aligned}$$

The equation of motion for $(G^\mu)^*$ is then

$$-D_\nu \Gamma^{\mu\nu} = 0,$$

and similarly, for G^μ , we get

$$-D_\nu^* (\Gamma^{\mu\nu})^* = 0.$$

7.4. U(1) Conserved Current

By Noether’s theorem, every continuous symmetry corresponds to a current satisfying a continuity equation. Here we are poised to interpret the Lie algebra as the infinitesimal behavior of the group—the Lie algebra of U(1), the number i , is the defining parameter of the small variation of the group action (gauge transformation) on the fields. This infinitesimal variation is easily understood as just the coefficient of the first derivative term in the Taylor expansion of the group action. The following is sometimes referred to as the *covariant variational principle*. Variation of the action with respect to U(1) transformations of the form $e^{i\Lambda(x)} \in U(1)$ corresponds to variation of the fields G_μ / G_μ^* and their covariant derivatives $D_\mu G_\nu / D_\mu^* G_\nu^*$ (where A_μ remains unaltered for obvious reasons—look at how it usually transforms) with respect to infinitesimal U(1) transformations:

$$\delta G_\mu \rightarrow i\delta\Lambda G_\mu \quad \text{and} \quad \delta G_\mu^* \rightarrow -i\delta\Lambda G_\mu^*.$$

So,

$$\begin{aligned} (\Gamma^{\mu\nu})^* \delta \Gamma_{\mu\nu} &\rightarrow 2i(\Gamma^{\mu\nu})^* \left\{ \delta\Lambda \left[\partial_\mu G_\nu + iqA_\mu G_\nu \right] + \delta(\partial_\mu \Lambda) G_\nu \right\} \\ &= 2i(\Gamma^{\mu\nu})^* \left\{ (\delta\Lambda) D_\mu G_\nu + \delta(\partial_\mu \Lambda) G_\nu \right\} \end{aligned}$$

And similarly,

$$(\delta\Gamma_{\mu\nu}^*)\Gamma^{\mu\nu} \rightarrow -2i\{(\delta\Lambda)D_\mu^*G_\nu^* + \delta(\partial_\mu\Lambda)G_\nu^*\}\Gamma^{\mu\nu}.$$

Plugging into $\delta S = \int d^4x \{(\delta\Gamma_{\mu\nu}^*)\Gamma^{\mu\nu} - (\Gamma^{\mu\nu})^* \delta\Gamma_{\mu\nu}\}$ we have

$$\begin{aligned} \delta S &= \int d^4x i(\delta\Lambda) \{D_\mu^*G_\nu^*\Gamma^{\mu\nu} - (\Gamma^{\mu\nu})^* D_\mu G_\nu\} + i\delta(\partial_\mu\Lambda) \{G_\nu^*\Gamma^{\mu\nu} - (\Gamma^{\mu\nu})^* G_\nu\} \\ &= \int d^4x i\delta(\partial_\mu\Lambda) \{G_\nu^*\Gamma^{\mu\nu} - (\Gamma^{\mu\nu})^* G_\nu\} \end{aligned}$$

Which means the current density satisfying $\delta S = \int d^4x j^\mu \delta(\partial_\mu\Lambda) = 0$ is

$$j^\mu = i \{G_\nu^*\Gamma^{\mu\nu} - (\Gamma^{\mu\nu})^* G_\nu\}.$$

It is worth noting that, for convenience and without loss of generality, we could have chosen U(1) transformations $e^{iq\Lambda(x)}$ so that the current density would include the constant factor of interaction strength (for this gauge group, we might as well call it electric charge). If this is done, the current equates precisely to the divergence of the field strength of the gauge field. If we examine the equation of motion for the gauge field, it's easy to notice that

$$\partial_\nu F^{\mu\nu} = iq \left[G_\nu^*\Gamma^{\mu\nu} - (\Gamma^{\mu\nu})^* G_\nu \right] = qj^\mu$$

Which directly implies $qj^\mu = -\frac{\partial\mathcal{L}}{\partial A_\mu} = -\partial_\nu \frac{\partial\mathcal{L}}{\partial F_{\mu\nu}}$, making clear that the current is gauge invariant.

8. One More for Good Measure: A Hypercomplex Theory

The quaternions, $\mathbb{H} \cong \mathbb{R} \oplus \mathbb{R}^3$, are a real four-dimensional anticommutative algebra with general elements $q = a + bi + cj + dk$ where $a, b, c, d \in \mathbb{R}$ and i, j, k are all equal to $\sqrt{-1}$. However, $i \neq j \neq k$. The so-called pure quaternions $\{i, j, k\}$ are linearly independent, anticommuting numbers defined by the relations

- 1) $i^2 = j^2 = k^2 = ijk = -1$
- 2) $ij = -ji = k, ki = -ik = j$ and $jk = -kj = i$

Linearly independent implies the pure quaternions (a linear subspace of the full quaternions) are a three-dimensional vector space (over the real numbers). And by considering the above relations, it is hardly surprising that there's a Lie algebra isomorphism $\mathfrak{su}(2) \cong \{i, j, k\}$ if we give the pure quaternions the bracket, i.e. $[i, j] = ij - ji = 2k$.

By the above algebra isomorphism, it's straightforward to identify the Lie group isomorphism for $SU(2) \cong S\mathbb{H} = \{\Omega = e^{ai+bj+ck} \text{ for } a, b, c \in \mathbb{R}\}$. The group $S\mathbb{H}$ is called the quaternionic sphere. It's interesting to note that as manifolds, both $S\mathbb{H}$ and $SU(2)$ are isomorphic to all points unit distance from the origin in four dimensions—which is often called the 3-sphere S^3 .

Define $u_\ell = (1, -k, j)$ where $\ell = 1, 2, 3$ indexes the pure quaternions multiplied by $-i$ so that we can define $\Omega(x) = e^{i\theta^\ell u_\ell} \in S\mathbb{H}$ for $\theta^\ell \equiv \theta^\ell(x)$ smooth functions to be local transformations over spacetime by the sphere carved out in

four dimensions via quaternion exponentials. The Lie algebra indexed parameter u_ℓ has been defined for no reason other than to mimic the fundamental representation of $SU(2)$ —where we multiply by i before exponentiation.

Define a quaternionic field over spacetime as

$$\Psi(x) = \alpha(x) + i\sigma(x) + j\tau(x) + k\omega(x)$$

for $\alpha, \sigma, \tau, \omega$ real smooth functions of the spacetime coordinates. We may also write, equivalently, $\Psi(x) = \alpha(x) + i\beta^\ell(x)u_\ell$ for $\beta^\ell(x)u_\ell = \sigma(x) - k\tau(x) + j\omega(x)$. The adjoint of the field is the quaternionic conjugate

$$\Psi^*(x) = \alpha(x) - i\sigma(x) - j\tau(x) - k\omega(x) = \alpha(x) - i\beta^\ell(x)u_\ell.$$

Since the field is just a quaternion, itself and its conjugate are primed to be transformed by any given $\Omega \in S\mathbb{H}$ as $\Psi \rightarrow \Omega\Psi$ and $\Psi^* \rightarrow \Psi^*\Omega^*$. With that, we can readily define a gauge theory whose gauge group is the quaternionic sphere.

Define the gauge covariant derivative as $D_\mu = \partial_\mu + igu_\ell X_\mu^\ell$. Under local action by $\Omega(x) \in S\mathbb{H}$ the Lie algebra-valued forms do the usual:

$$iu_\ell X_\mu^\ell \rightarrow \Omega iu_\ell X_\mu^\ell \Omega^* + \frac{i}{g}(\partial_\mu \Omega)\Omega^*.$$

We can now write down an $S\mathbb{H}$ -invariant theory for the particle we defined. For a massive quaternionic scalar field, we can write the following non-Abelian gauge theory:

$$\mathcal{L} = (D_\mu \Psi)^* D^\mu \Psi + m\Psi^* \Psi - \frac{1}{4} X_{\mu\nu}^\ell X_\ell^{\mu\nu}$$

Transforming the gauge covariant derivative with $\Omega(x) \in S\mathbb{H}$:

$$\begin{aligned} (D^\mu \Psi)' &= \partial^\mu (\Omega\Psi) + ig \left(\Omega u_\ell X_\ell^\mu \Omega^* + \frac{i}{g} (\partial_\mu \Omega)\Omega^* \right) \Omega\Psi \\ &= \Omega (\partial^\mu \Psi) + (\partial^\mu \Omega)\Psi + i\Omega u_\ell X_\ell^\mu \Omega^* \Omega\Psi - (\partial_\mu \Omega)\Omega^* \Omega\Psi \\ &= \Omega (\partial^\mu \Psi) + i\Omega u_\ell X_\ell^\mu \Psi = \Omega (\partial^\mu \Psi + iu_\ell X_\ell^\mu \Psi) = \Omega (D^\mu \Psi) \end{aligned}$$

As we have seen many times before:

$$\begin{aligned} \mathcal{L}' &= (D_\mu \Psi)^* \Omega^* \Omega (D^\mu \Psi) + m\Psi^* \Omega^* \Omega \Psi - \frac{1}{4} X_{\mu\nu}^\ell X_\ell^{\mu\nu} \\ &= (D_\mu \Psi)^* D^\mu \Psi + m\Psi^* \Psi - \frac{1}{4} X_{\mu\nu}^\ell X_\ell^{\mu\nu} = \mathcal{L} \end{aligned}$$

As should be clear from the isomorphisms we mentioned, this quaternionic gauge theory is in all ways equivalent to the $SU(2)$ /spinor theory we discussed in section V. But, unlike that previous theory, there is no need for Latin indices to index the components of our particle—our particle is a one-component quaternion!

This one-dimensional formulation of an $SU(2)$ -theory, at first, seems extraordinarily convenient. However, even though we technically have enough real di-

mensions, we will run into a complication between the quaternions and the Lorentz group if we attempt to make a theory with a quaternionic particle description physical (make it look like a Dirac particle). This complication boils down to the fact that notions of length or norms easily defined on the quaternions do not respect the metric signature on Minkowski spacetime. There is a way to fix this, however. The first step is to *complexify* the quaternions, $\mathbb{H} \otimes_{\mathbb{R}} \mathbb{C}$.

Acknowledgements

We gratefully acknowledge Dr. David Lambert, Dr. Akhila Mohan, Dr. Duncan Weathers, Dr. Trever Harborth, and Dr. Charles Conley for their valuable discussions and suggestions during the preparation of this article. H.A. extends sincere appreciation to his former mentors: Stephen Arico, his high school physics instructor, and Dr. Randolph Peterson, his undergraduate advisor; and he is especially grateful to Y.R. for his guidance throughout the publication process. We also thank the Department of Physics at the University of North Texas for partial financial support.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Venkataraman, G.G. (n.d.) A Brief Introduction to the Lie Bracket. Reed College. <https://people.reed.edu/~jerry/332/projects/venkataraman.pdf>
- [2] Fulton, W. and Harris, J. (2004) Representation Theory: A First Course. Springer.
- [3] Gökeler, M. and Schücker, T. (1999) Differential Geometry, Gauge Theories, and Gravity. Cambridge University Press.
- [4] Maru, N., Okada, N. and Okada, S. (2018) $SU(2)_L$ Doublet Vector Dark Matter from Gauge-Higgs Unification. *Physical Review D*, **98**, Article ID: 075021. <https://doi.org/10.1103/PhysRevD.98.075021>
- [5] Escalona, P., Acevedo, S., *et al.* (2024) Vector Dark Matter from the 5-Dimensional Representation of $SU(2)_L$. *JHEP*, **2024**, Article No. 162. [https://doi.org/10.1007/JHEP11\(2024\)162](https://doi.org/10.1007/JHEP11(2024)162)