A Survey of Gauge Theory and Yang-Mills Equation

Hyun Youk
University of Toronto

March 28, 2003

1 Preamble

In classical physics, the concept of force fields is introduced primarily as a means to facilitate our understanding of "action at a distance". That is, we use fields as means to facilitate our understanding of how one particle "knows" how much force that a particle distant from it is subjecting it to. Newton himself was uneasy with the notion of "action at a distance" and introduced "gravitational fields". In classical electromagnetism, for instance, while electromagnetic fields are first introduced to mediate the concept of "action at a distance", they soon take up lives of their own, and are considered to be on equally realistic footing as seemingly more concrete entities like material particles. For instance, we consider energy and momentum of a system of particles to be stored in the electromagnetic fields, and not just in particles. One theory in particular, called informally 'gauge theory' plays a very important role in describing force fields and particles in nature. But comprehending gauge theory is no easy task since it requires an enormous amount of mathematical tools and physical concepts.

In this paper, we try to scratch the surface of some of the fundamental notions of gauge theory and their associated mathematical concepts such as Lie groups, Lie algebra, and bundles in particular. We'll try to apply the mathematical formalisms we develop to various physical situations whenever possible. Our final goal, however, is to combine the concepts we develop early on in our program to elucidate the *Yang-Mills equation* and explain what it does.

Note about notations: From this point on, we'll denote the n-dimensional Euclidean and complex spaces by R^n and C^n respectively. N will denote the set of natural numbers. **g** will denote a Lie algebra associated with a Lie group G. Manifolds will mostly be denoted by M. In addition, unless stated otherwise, all manifolds M under consideration will be smooth manifolds. The letter G will be reserved for groups. We tried to enumerate only those equations that are not "definition-related" and are of some physical importance. Most of them are found towards the end of this paper.

2 Lie Groups

The study of Lie groups is a vast arena itself. Here, we present only those parts of the theory of Lie groups that pertain to our study of Yang-Mills equations and gauge theory.

2.1 Useful Groups in Physics

Before defining what Lie groups are, we first present some groups that will be of utmost importance for us.

Notation	Group	Description
GL(V)	General Linear group	Group of all linear transformations on vector space V.
GL(n, R)	General Linear Group	Group of all invertible $n \times n$ matrices with real entries.
GL(n, C)	General Linear Group	Group of all invertible $n \times n$ matrices with complex entries. Note that GL(n, R) is a subgroup of GL(n, C)
SL(n, R)	Special Linear Group	Group of all $n \times n$ matrices of real entries with determinant equal to 1. In other words, this is the group all volume-preserving linear transformations in R^n.
O(p, q)	Orthogonal Group	Here, p and q are nonnegative integers with p + q = n. Let g be a metric on R^n with signature (p,q). More specifically, we define g to be the following: $g(v,w) = v^1w^1 + ... + v^pw^p - v^{p+1}w^{p+1} - ... - v^{p+q}w^{p+q}$. Then O(p,q) is the group of $n \times n$ matrices T that preserve g. That is, g(Tv,Tw) = g(v,w) for all v, w $\in R^n$
SO(p,q)	Special Orthogonal Group	Group of matrices in O(p,q) that have determinant equal
O(n)		p=n and q=0, with g being the standard Euclidean metric.
SO(n)		
SO(3)		According to above notation, this is just a group of all rotations in R^3.
SO(3,1)	Lorentz Group	Group of 4×4 matrices preserving the standard Minkowski metric. Hence the Lorentz transformation is also contained in this group. In fact, SO(n,1) for any $n \geq 1$ is a Lorentz group in Minkowski spacetime.
U(n)	Unitary Group	Group of all unitary $n \times n$ complex matrices. That is, this is the group of all those linear transformations T such that: $< Tv, Tw >=< v, w >$, where $<, >$ is usual inner product on C^n and v,w $\in C^n$
SU(n)	Special Unitary Group	All those elements in U(n) whose determinant is equal to 1. SU(n) is a subgroup of U(n).

One can easily check that all those listed above are indeed groups with the appropriate standard definitions of inverse and product operations for each.

2.2 Defining Lie Groups and Group Representations

Now, consider the vector space of $n \times n$ matrices, either over R or C. It turns out that the matrix groups SL(n, R), SL(n, C), O(p,q), SO(p,q), and U(n) are all **submanifolds** of the vector space of the $n \times n$ matrices. In addition, the product and inverse operations on these groups can be shown to be smooth maps. Groups of these types are what are called the **Lie groups**. More formally, a group G is called a Lie group, if G is a manifold, with the product and inverse operations being smooth maps on G.

Groups can also 'interact' with other objects, such as vector spaces. We say that a group G **acts** on a vector space V if there exists a homomorphism ρ : G→GL(V) such that ρ(gh)v = ρ(g)ρ(h)v for all v ∈ V. ρ is called **representation** of G on V. For our purposes here, we'll assume that ρ is smooth as well.

Example : Here's one important example of a group representation (though elementary). First, we show that U(1)={$e^{i\theta}$: $\theta \in R$}. The elements of U(1) are just numbers (1×1 matrices) that act on C. Let T ∈ U(1). T can in turn, be expressed as T = $(\alpha)e^{i\theta}$ ($\alpha \in R$). Then,
$\langle$Tx, Ty$\rangle$= Tx$\overline{Ty}$= T$\overline{T}$x$\overline{y}$= x$\overline{y}$= $\langle$x, y$\rangle$, ∀ x, y ∈ C
⇔ T is not zero, and T$\overline{T}$=1
⇔ $((\alpha)e^{i\theta})(\overline{\alpha}e^{-i\theta}) = \alpha\overline{\alpha} = 1 \Rightarrow \alpha = \pm\, 1$.
But $e^{i\theta+\pi}$ = -$e^{i\theta}$, so can just say $\alpha = 1$. Hence, it follows that
U(1)={$e^{i\theta}$: $\theta \in R$}. It's also easy to see that U(1) is isomorphic to SO(2) with the isomorphism ρ given by $\rho(e^{i\theta})$= $\begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}$.
More importantly, for our purpose, we can define a representation,
$\rho(e^{i\theta}) \equiv e^{i\theta}$.
That is, $\rho \equiv$ identity map. In fact, since all matrix groups are already subsets of GL(V), they're naturally equipped with the 'identity transformation' as one possible representation. Such a representation is called the **fundamental representation**.

2.3 Gauge Groups: Ingredients for the Yang-Mills Equation

All the definitions above have been quite formal. But what do Lie groups have to do with the Yang-Mills equation, or physics for that matter? Quite a lot! The relationship between Lie groups and Yang-Mills equation, is that different Lie groups give rise to different Yang-Mills equations, and these in turn describe the various forces in the standard model of physics. The Lie group that corresponds to a given Yang-Mills equations, or informally, the group that provides the 'ingredient' for a given Yang-Mills equation for a particular force under consideration, is called the **gauge group** (or **symmetry group)**. It'd be ideal at this stage to give a brief preview of what we'll be talking about later on, in order to get comfortable with the idea of 'associating' a Lie group with a particular force. In addition, it will turn out that the Yang-Mills equations are linear if and only if the associated gauge group is abelian. But we're getting ahead of ourselves here. Let's first work out an example.

Electromagnetism has U(1) as its gauge group. From our work above, we know that we can consider U(1) as forming a unit circle in the complex plane. The electroweak force has as its gauge group $SU(2) \times U(1)$, while the strong nuclear force (i.e. Strong force) has $SU(3)$ as its gauge group. Finally, the gauge group of the standard model is $SU(3) \times SU(2) \times U(1)$. One of the attempts of the so called **grand unified theories** (GUTs) is putting physics on a "larger and nicer" gauge group by treating $SU(3) \times SU(2) \times U(1)$ as a subgroup of some "larger and nicer" group such as SU(5).

One can also explore the connection between gauge groups and physical charges (both electric and 'colour' charges). The intimate connection between them is that the physical charge of a particle is described naturally by selecting a particular representation for the gauge group in question. This is one of the powers of using group theoretical methods to build models of particle physics. Before moving on, let's take a closer look at one example of a gauge group, namely the gauge group for electromagnetism U(1).

Example : We first define what seems to be quite a 'natural' representation ρ_n of U(1) (ρ_n: U(1) $\rightarrow$ GL(C)). Define $\rho_n(e^{i\theta})$v $= e^{in\theta}$v, where v $\in C$ and n $\in \aleph$. First, we show that this is indeed a representation. This is easy to do since $\rho_n(e^{i\theta}e^{i\phi}) = e^{in\theta}e^{in\phi}$. So indeed, ρ_n is a homomorphism, and noting that $e^{in\theta}$ acts on any v, by effecting a rotation of nθ on v, we see that it's an element of GL(C) indeed. Hence, ρ_n is a representation of U(1) on C. From quantum mechanics, we know that electric charges are quantized. That is, a particle can only have nq charges where q is the elementary charge, which we take to be the charge of an electron for simplicity (we won't go into quark model here for simplicity, which states that $\frac{1}{3}$q is the true 'elementary' charge). Then, note that a particle with charge nq transforms according to the representation ρ_n of U(1) on C. Namely, say we move a particle of charge nq around a loop γ in spacetime, then quantum mechanics tells us that its wavefunction is multiplied by a certain phase, which is an element of U(1), namely by $e^{-\frac{i}{\hbar}nq(\oint_\gamma A)}$, where A is a vector potential. Now, the numerical values of q and $\hbar$ are dependent on the choice of units, which can be arbitrary. We choose our units such that $\hbar =$ q $= 1$. This is often called the "God given units". We then have $e^{-in\oint_\gamma A} =$

$\rho_n(e^{-i\oint_\gamma A})$, with $e^{-i\oint_\gamma A} \in$ U(1).

One of the main goals of gauge theory is to generalize ideas such as this to other gauge groups corresponding to different forces.

3 Lie Algebras

We interrupt our study of Lie groups at this juncture, to introduce an equally important notion called Lie Algebras.

3.1 Preliminary Definition

We start off by defining what a **Lie Algebra** is. Let G be a Lie group. Then a Lie algebra of G, denoted **g**, is the tangent space of the identity element of G. Note that since G is a manifold, above is well-defined. One heuristic way of thinking about a Lie algebra of a Lie group G is to view it as a space of tangent vectors to paths in G that originate at the identity element of G. Note that this is a vector space with same dimension as G. It will turn out that a more abstract definition can be given for Lie algebra. Rather than introducing it right away, we go through an example, and some algebraic formalities to let it come out 'naturally'.

3.2 Lie Algebra of SO(3)

One example of Lie algebra that is useful in physics, and especially in quantum mechanics when one studies angular momentum, is the notion of an 'infinitesimal rotation'. Let γ be a path in SO(3) such that it corresponds to a rotation by the angle t counterclockwise about the z-axis. That is:

$$\gamma(\mathrm{t}) = \begin{pmatrix} \cos(t) & -\sin(t) & 0 \\ \sin(t) & \cos(t) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Then at t=0, the curve is at identity element of SO(3), and the tangent vector to γ as it passes through it is easily computed to be:

$$\gamma'(\mathrm{t}{=}0) = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \equiv J_z$$

Rotating about x and y axes, we obtain the tangent vectors J_x and J_y respectively, where:

$$J_x = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \qquad J_y = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}$$

These are all in the Lie algebra of SO(3), and we denote it by **so(3)**. It's important to keep in mind that we are interested in paths that lie in SO(3) and not in R^3!. After all, SO(3) is a manifold (since it's a Lie group), and when we say a 'path' in SO(3), we can informally think of it as being a string of matrices in SO(3) "sewn" together along this path, such that as one moves along the path, the entries of the matrices change continuously (i.e. 'smoothly').

Now that we have obtained these matrices representing 'infinitesimal rotations' about the identity element, what we now show is an important concept that one can obtain matrices describing finite rotations in R^3 by using exponentials of linear combinations of the infinitesimal rotations J_x, J_y, and J_z.

As we know from classical linear algebra, the 'exponential' of an n × n real matrix A is defined by the 'Taylor Series' of it:

$$\exp(\mathrm{A}) = 1 + A + \tfrac{A^2}{2!} + \tfrac{A^3}{3!} + \dots$$

Before moving on, we need to prove that above series actually makes sense. That is, it converges to some matrix S with the same dimensions as A. This proof would be rather tedious and long, so we carry out only the key parts of it:

Claim: exp(A) is well defined in the sense above.
Proof: We define a norm ‖A‖ of n × n real matrix A as follows:
$$\parallel A \parallel \equiv sup \tfrac{\|Ax\|}{\|x\|},$$
where the supremum is taken over all non-zero vectors v. Also, note that ‖ . ‖ on the right hand side of above equality is the standard norm. Then by linearity, it's easy to show that we can project all the vectors onto the unit sphere so that
$$\parallel A \parallel = sup \parallel Ay \parallel,$$
where the supremum is taken over all unit vectors y. Then this norm obeys the properties of norm which are:
(i) $\parallel \alpha A \parallel = \mid \alpha \mid \parallel A \parallel$
(ii) $\parallel A + B \parallel \leq \parallel A \parallel + \parallel B \parallel$
(iii) $\parallel A \parallel = 0 \Leftrightarrow A = 0$, and $\parallel A \parallel > 0$ otherwise.
It also satisfies one further property which is that
(iv) $\parallel AB \parallel \leq \parallel A \parallel \parallel B \parallel$
Now, we define the sequence $< s_n >$ where
$$s_n \equiv \sum_{k=0}^{n} \frac{A^k}{k!}$$
then for m ≥ n, we have
$$\begin{aligned} \parallel s_n - s_m \parallel &= \parallel \frac{A^{n+1}}{(n+1)!} + ... + \frac{A^m}{m!} \parallel \\ &\leq \parallel \frac{A^{n+1}}{(n+1)!} \parallel + ... + \parallel \frac{A^m}{m!} \parallel, \text{(by (ii))} \end{aligned}$$

$$\leq \frac{\|A^{n+1}\|}{(n+1)!} + ... + \frac{\|A^m\|}{m!} \text{ , (by (iv))}$$

In the last line above, we have $\| s_n - s_m \|$ reduced to the tail-end of the usual exponential series, with $\| A \|$ treated as just a number. And we know that the Taylor series for the exponential function converges. Hence, this means that $\forall \varepsilon > 0, \exists N_\varepsilon \in N$ such that $\| s_n - s_m \|$ ¡ ε, $\forall n, m \geq N_\varepsilon$. Hence, it follows that $\{s_n\}$ is a Cauchy sequence, and by the completeness of C^n (or R^n, depending on the space we're working with), the sequence converges, which means that we can find some n × n matrix S such that $\| S \| = \lim_{n \to \infty} \| s_n \|$. This proves that the exponential of the square matrix is well-defined indeed.

Now that we've convinced ourselves that exp(A) does make sense, it turns out that to obtain a matrix describing a rotation by the angle t around, say the z-axis in R^3, is exp(tJ_z). To see this, we just need to compute the exponential then verify in the end. Noting that:

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

Hence, J_z^3=-J_z, J_z^4=-J_z^2, and so on. Then exploring this relationship, we can easily compute the desired exponential as follows:

$$\exp(\mathrm{t}J_z) = 1 + tJ_z + \frac{t^2}{2!}J_z^2 - \frac{t^3}{3!}J_z - \frac{t^4}{4!}J_z^2 + \ldots$$
$$= 1 + (t - \frac{t^3}{3!} + \ldots)J_z + (\frac{t^2}{2!} - \frac{t^4}{4!} - \ldots)J_z^2$$
$$= 1 + \sin(tJ_z) + (1 - \cos(t))J_z^2$$
$$= \begin{pmatrix} \cos(t) & -\sin(t) & 0 \\ \sin(t) & \cos(t) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Above matrix is indeed a counterclockwise rotation by angle t about the z-axis in R^3. By the similar process of calculation, which we won't do here, it turns out that the matrix describing a counterclockwise rotation of angle t about the unit vector $n = (n^x, n^y, n^z) \in R^3$ is $\exp[t(n^x J_x + n^y J_y + n^z J_z)]$. In this example, we have seen how Lie algebra of a certain Lie group, can be very useful for study of Lie groups.

3.3 Formal Definition of Lie Algebra

In the previous example, we saw how the Lie algebra of a Lie group (namely, SO(3)) allowed us to construct other elements of the group. This, in fact, is not unique to SO(3) but is rather a surprising relationship between Lie Algebra and Lie groups. A lot of important properties we've noted in the previous example carries over to Lie algebras in general. For one, it turns out that any Lie group G, not just the matrix groups, has an **exponential map**. The exponential map exp: **g**→ G, is determined uniquely by the following three properties, which we won't prove here as they are beyond the scope of this paper, and would only lead us astray. The three properties are:

(i) exp(0) = identity element of G. (Here, 0 is the identity element of the Lie algebra of the Lie group G).
(ii) exp(sx)exp(tx) = exp((s+t)x), $\forall$ x $\in \gamma$ and s, t $\in R$. (where γ is any curve that originates from the identity element of G).
(iii) $\frac{d}{dt}\exp(tx)\mid_{t=0}$ = x.

Above properties match the 'regular' properties of the exponential function in C that many of us are more familiar with. As a consequence of above properties and the inverse function theorem, it turns out that exp can map any sufficiently small open neighbourhood containing the identity element 0 of Lie algebra **g** onto an open set that contains the identity of G. Furthermore, any element of the **identity component** of G is the product of elements of the form exp(x). An identity component is a connected component that contains the identity element of the group. Then as we mentioned above, in some sense, most of the structure of a Lie group is encoded in the structure of its Lie algebra. We elaborate this very important idea further.

To do so, we first abstractly define a **Lie algebra**. First, we denote the **Lie bracket** (or **commutator**) by $[\cdot, \cdot]$. Then, we define Lie algebra to be any vector space **g** that is equipped with a map $[\cdot, \cdot]$: **g**$\times$**g** $\rightarrow$ **g**. such that the following three properties are satisfied (properties of Lie brackets):
(i) [v,w] = -[w,v], $\forall$ v,w $\in$ **g**.
(ii) [u, αv + βw] = α[u,w] + β[u,w], $\forall$ v,w, u $\in$ **g**, and scalars α and β.
(iii) (Jacobi identity): [u,[v,w]]+[v,[w,u]]+[w,[u,v]]=0, $\forall$ u,v,w $\in$ **g**.
(Note: α and β are real numbers for Lie algebra **g** that forms a real vector space. They are complex numbers if **g** forms a complex vector space.)

We can also define **homomorphism** between one Lie algebra to another. This is quite naturally defined as a linear map f from the Lie algebra **g** to the Lie algebra **h** such that f([v,w]) = [f(v),f(w)], $\forall$ v,w $\in$ **g**. Of course, f is an **isomorphism** if it's one-to-one and onto.

Note that these notions that we've been talking about here have striking resemblances with Lie brackets defined on vector fields. Indeed, given a manifold M, the space of all vector fields on M is a Lie algebra with the standard Lie bracket defined for vector fields on M. The difference here is that it's an infinite dimensional Lie algebra where as for the Lie groups we've been considering thus far, deal with finite dimensions.

We have seen that every Lie group has a Lie algebra. But what we haven't mentioned so far is that every homomorphism ρ: G $\rightarrow$ H (where G and H are Lie groups) determines an associated homomorphism $d\rho$: **g** $\rightarrow$ **h** between the Lie algebra **g** of G and the Lie algebra **h** of H. More specifically, $d\rho$ is the **pushforward** of tangent vectors at the identity of G. That is,

$d\rho = (\rho)_* : T_{identity}G \rightarrow T_{identity}H$.
Although we choose to omit the proof here, above indeed forms a Lie algebra.

3.4 Application to Physics

Our goal, from the beginning, was to see how Lie algebras and Lie groups are used in various parts of physics. We have developed sufficient amount of definitions and are now in a position to do just that. But before doing so, we need just one more formal definition, namely, the notion of representation of Lie algebras.

Lie group representations determine Lie algebra representations and vice versa. In the spirit of the definition of representation of a Lie group G on a vector space V, we define a **representation** of a Lie algebra **g** on a vector space V to be a Lie algebra homomorphism f: $\mathbf{g} \rightarrow \mathbf{g}[$(V), where $\mathbf{g}[$(V) is the Lie algebra of all linear operators on V, with the standard definition of a commutator. Recall that we have stated exp is well defined for any Lie groups. A very important fact about Lie algebras is that if one starts with a representation f: $\mathbf{g} \rightarrow \mathbf{g}[$(V), one can 'exponentiate' it to obtain a representation ρ: G → GL(V) with $d\rho =$ f, given that G is simply connected.

Example After all the formalities, we can now give a basic sketch of application of Lie algebras in physics. Most of the mathematics in quantum mechanics are deeply rooted in the infinite dimensional Hilbert space H. Under rotation symmetry in 3 dimensional space, there exists a unitary representation U of SU(2) on H. And as we mentioned above, this representation naturally gives rise to a representation dU of the Lie algebra **su(2)** on H. We can then define $dU(\text{ix}) \equiv \text{i}dU(\text{x})$ for any x $\in$ **su(2)**. Physically, the operator $dU(\frac{\sigma_z}{2})$ on H is called the **angular momentum** about the z-axis, where σ_i is a **Pauli Matrix** about the i-axis. The precise entries of this matrix is not important for our purpose here.

Let's elaborate on this further. In quantum mechanics, a system has its associated wavefunction, which encodes all the information about the system, including the measurement values of any observables one might want to measure. Let ψ be a wavefunction for a particle. $\psi \in H$ for physical particles. The statistical expected value of some observable A of a particle is given by: $< \psi, A\psi >$ with $< f, g >$ being the usual inner product on a Hilbert space (i.e. $\int(f\overline{g})$, integral taken over all physical space). Then according to the formalism we've developed above, it follows that the z-component of the system's angular momentum about the z-axis is just $< \psi, dU(\frac{\sigma_z}{2})\psi >$. Generalizing this concept, the angular momentum about any unit vector v $\in R^3$ is given by the operator $dU(v^i/sigma_i/2)$, (where $/sigma_i$ denotes the Pauli Matrix defined in the appendix). In a similar fashion, we can use other symmetry groups to describe other observables. So for instance, translation in space gives momentum, and translation in time gives energy. This shouldn't be a surprising result considering that even in classical mechanics, momentum **p** and energy E exhibit relationships: $\mathbf{p} = \mathbf{F} \cdot (\triangle \mathbf{d})$ and $E = \mathbf{F}\triangle(t)$, where $\triangle t$ and $\triangle \mathbf{d}$ denote time intervals and displacements through which the force **F** is applied. These intimate connections exhibited between time and Energy, and space and momentum, are two of the fundamental relationships in physics, which crop up again and again in such phenomena as Heisenberg's uncertainty principle, which again couples energy with time interval, and momentum with displacements. We see here that these relationships come up in group theoretical treatments of physical models as well.

4 Bundles: Ingredients for Gauge Fields

In order to reach our goal of constructing Yang-Mills equations, we are inevitably forced to study the algebraic concept of bundles, and gauge transformations. Gauge transformations, in particular, are important in many physical models. For instance, even in classical electromagnetism, Maxwell's equations describing any charge and current configurations can be simplified from a computational perspective, if we 'gauge transform' the Maxwell's equations. In Yang-Mills equations, this concept becomes especially important, as we'll see. Here, we present only the very basic aspects of bundles and gauge transformations that pertain to our purposes.

4.1 Bundles: Preliminaries

Gauge theory deals with general fields on spacetime. What exactly are these 'general fields'? To describe them, we need to understand a bit about vector bundles. First, we note the important fact that a vector field on a manifold M assigns to each point p $\in$ M, a tangent vector which lives in the tangent space T_pM at p on M. This means that since the vector field assigns vectors not in some fixed single vector space, we cannot naively compare two tangent vectors at two different points of M, because they live in different spaces. This, in turn, implies that we cannot define a differentiation of a vector field on M the same way we define differentiation of real-valued functions in Euclidean space. This seemingly innocent fact has far reaching consequences in Gauge theory. In gauge theory, fields are different 'segments' of the 'bundles' or the 'collection' of the tangent spaces of M, which assign to each point in spacetime a vector in the vector space to which the point belongs to. What is needed for formation of Yang-Mills equations is a method to compare vectors in these different tangent vector spaces, which is a field called a 'connection'. Yang-Mills equations turn out to be the equations for this connection itself, as we'll see after we familiarize ourselves with some algebraic formalisms.

A **bundle** is an algebraic structure consisting of two manifold E and M, and an onto map π : E $\rightarrow$ M. For example, M can be a real line, and E can be R^2

plane, with π being the standard projection map π(x,y)=x, $\forall$ (x,y) $\in R^2$. With E, M, and π taking roles in the sense of this example, E is called a **total space**, M is called **base space**, and π is not surprisingly called the **projection map**. Furthermore, for each p $\in$ M, we define $E_p = \{q \in E : \pi(q) = p\}$ to be a **fiber over p**. Then it follows that the total space E can be 'built up' from its fibers, E = $\bigcup_{p\in M} E_p$. For this reason, we often refer to E as being a 'bundle', or a 'bundle' over M. In physical applications, M becomes a Euclidean space or a spacetime.

A **tangent bundle** of a manifold M can be defined as follows. In a tangent bundle, the total space TM is just $TM = \bigcup_{p\in M} T_pM$, where T_pM is the tangent space at p of M. In this bundle, the project map π is naturally defined to be π(v) = p, where v $\in T_pM$ and p $\in M$. Hence, the fiber over p is T_pM. But this doesn't guarantee that TM is a manifold because π needs to be guaranteed to be smooth. To do so, we use the idea that specifying a point in TM is the same as specifying a point p in M together with a vector v $\in T_pM$. And since both M and T_pM look locally like R^n (where M is an n-dimensional manifold), TM should locally look like $R^n \times R^n$. Let's make this idea more precise:

Let M be a n-dimensional smooth manifold over R^n. Let $\phi_\alpha : U_\alpha \to R^n$ denote a chart of M, with U_α being an open set on M. Define a subset of TM: $V_\alpha = \{v \in TM : \pi(v) \in U_\alpha\}$. Now, for each v $\in TM$, since $TM = \bigcup_{p\in M} T_pM$, $\exists$ p $\in$ M such that v $\in T_pM$ (i.e. v is a tangent vector at p on M). Hence π(v)=p, and since M is a manifold, there exists an open set U_α that contains p. Hence v $\in V_\alpha$ for some V_α of TM. Next, define maps $\psi_\alpha : V_\alpha \to R^n \times R^n$ by $\psi_\alpha(v) = (\phi_\alpha(\pi(v)), (\phi_\alpha)_* v)$. We can think of $(\phi_\alpha)_* v$ as a vector in R^n though it is really a tangent vector to R^n. We induce a "natural" topology to TM by defining it to be collection of V_α's. That is, a set of TM is open if and only if it is one of V_α's. Then, note that $\phi_\alpha(\pi(V_\alpha)) \subseteq U_\alpha$ and is open in R^n because ϕ_α is continuous, and that $\phi_\alpha(V_\alpha)$ is also open in R^n since the pushforward ϕ_α is smooth. Hence, $\psi_\alpha(V_\alpha)$ is open in $R^n \times R^n$. Although it's tedious, it's easy to check that ψ_α's are charts for TM. Hence it follows that in this setting, TM is a manifold. It's also easy to check that π is a smooth map here. One can easily convince oneself of this fact by considering a smooth curve lying on M, and the tangent vector to the curve varying in a 'continuous' manner as one traverses the curve.

We define a **trivial bundle** over M with **standard fiber** F to be the Cartesian product $E = M \times F$, with the projection map $\pi(p, f) = p$, $\forall$ (p,f) $\in M \times F$. That is, a trivial bundle amounts to the construction of a bundle by choosing a desired base space along with an explicit fiber of your desire. Note that we can give a 'canonical' diffeomorphism between each fiber and F, by sending (p,f) $\in E_p$ to f $\in F$. This property turns out to be unique to trivial bundles.

Bundles that are not 'globally' trivial, but rather 'locally' trivial turn out to be more useful in many physical models. Informally, as you may have guessed, an a bundle is 'locally' trivial if it looks trivial in sufficiently small neighbourhood of any given point. Examples include cylinders (trivial over S^1 with standard fiber R) and Möbius strip over a 'small' portion of S^1. We can in fact, make this definition more precise. But first, we need to introduce some more notions.

Given two bundles, $\pi : E \to M$ and $\pi' : E' \to M'$, a **morphism** from the first to the second is a map $\psi : E \to E'$ together with a map $\phi : M \to M'$ such that ψ maps each fiber E_p into the fiber $E'_{\phi(p)}$. In other words, a morphism "transforms" one bundle into another. A morphism is called an **isomorphism** if ϕ and ψ are both diffeomorphisms.

Suppose we're given two bundles $\pi : E \to M$ and $\pi' : E' \to M'$. We now show that the maps $\psi : E \to E'$ and $\phi : M \to M'$ are a bundle morphism if and only if $\pi' \circ \psi = \phi \circ \pi$. First, suppose that $\psi : E \to E'$ and $\phi : M \to M'$ are a bundle morphism. Then for $y \in E_p$, $\pi' \circ \psi(y) = \pi'(z) = \phi(p) = \phi \circ \pi(y)$, where $z \in E'_{\phi(p)}$ by definition, and p $= \pi(y)$. Hence $\pi' \circ \psi = \phi \circ \pi$. Conversely, say $\pi' \circ \psi = \phi \circ \pi$. Then for $y \in E_p$; $\pi' \circ \psi(y) = \pi'(z) = p'$ where $z \in E'_{p'}$, $\Rightarrow \psi(y) \in E'_{p'}$. On the other hand, $\phi \circ \pi(y) = \phi(t) = p'$, where $t \in M \Rightarrow y \in E_t$, hence ψ maps E_p to $E'_{\phi(p)}$. Hence, we've proven what we set out to do.

Furthermore, it's easy to show that for a given ψ defined for a morphism, if ϕ and ϕ' are two corresponding maps for the base space, then $\phi = \phi'$. To see this, note that since we have a morphism, $\phi' \circ \psi = \phi \circ \pi = \phi' \circ \pi$ by our previous work above. Then now, the rest is very simple, because π is onto map, so $\phi(p) = \phi'(p), \forall p \in E, \Rightarrow \phi \equiv \phi'$. Hence, ψ *uniquely* determines ϕ. For this reason, we can be quite lax and call ψ to be the 'bundle morphism'.

One very useful and familiar example of a bundle morphism is induced by the pushforward of tangent vectors. Let $\phi : M \to M'$ denote a map between manifolds M and M'. Then we have the pushforward $\phi_* : T_pM \to T_{\phi(p)}M'$. We also know that $TM = \bigcup_{p\in M} T_pM$, and $TM' = \bigcup_{p'\in M'} T_{p'}M'$. Then we see that in fact, $\phi_* : TM \to TM'$, with ϕ being smooth $\Rightarrow \phi_*$ being smooth. Hence $\phi_* : TM \to TM'$ is a bundle morphism.

For a bundle $\pi : E \to M$ and a submanifold $S \subseteq M$, a **restriction** to S is defined by taking $E\mid_s = \{q \in E : \pi(q) \in S\}$ as the total space, and taking S as the base space with π restricted to $E\mid_s$ as the projection. Now, we're finally in a position to give a precise formulation of local triviality. A bundle $\pi : E \to M$ is called **locally trivial** if for each p $\in M$, there exists an open neighbourhood U of p, and a bundle isomorphism $\phi : E\mid_U \to U \times F$, such that it sends each fiber E_p to the fiber p $\times F$. Such a ϕ is called a **local trivialization**, and a section of $E\mid_U$ is called a section of E **over** U. When we use charts on manifolds, we are using the concept of 'local analysis'. In a similar spirit, local trivialization of a bundle simplifies our analysis of bundles by allowing us to perform a 'local analysis' of the bundle.

Before moving on, we note the important fact that for any smooth n-dimensional real manifold M, the tangent bundle $\pi : TM \rightarrow M$ is locally trivial. Rather than giving a formal proof, we can, as physicists, intuitively see this as following. Pick a point p on a manifold M, then via a chart local to that region of M, we can find some open subset U of M containing p. Now, for T_pM, we can define a map $\phi(v_p) = (p, v_p)$, and it's quite clear to see that this is a bundle isomorphism, with the standard fiber being R^n, with v_p tangent vector in p in R^n of course.

4.2 Vector Bundles

We now confine our attention to specific kind of bundles, namely the vector bundles. Vector bundles are what we really need in formulation of Yang-Mill's equations and mathematical physics in general.

An n-dimensional **real vector bundle** is a locally trivial bundle $\pi : E \rightarrow M$ such that each fiber E_p is a n-dimensional vector space, with the additional requirement that for each p ∈ M, there exists a neighbourhood U ∈ M of p and a local trivialization $\phi : E\ |_U \rightarrow U \times R^n$ that maps each fiber E_p to the fiber p$\times R^n$ linearly. In this situation, we say that the trivialization is **fiberwise linear**. The definition of complex vector bundle is exactly the same as this one except that we replace R^n with C^n. What we'll now describe for real vector bundles are equally applicable to complex vector bundles with the modification mentioned above.

We first show that TM of M is a vector bundle. We already showed above (at least heuristically), that TM is a trivial bundle. Let E_p be a fiber over p ∈ M, where M is an n-dimensional real smooth (as always) manifold. We know that T_pM is isomorphic to R^n (i.e. they "look" alike). Now then, let $\phi(v_p) = (\pi(v_p), v_p)$. Then, with real scalars α and β, we have:
$\phi(\alpha v_p + \beta w_p) = (p, \alpha v_p + \beta w_p) = (p, \alpha v_p) + (p, \beta w_p) = \phi(\alpha v_p) + \phi(\beta w_p)$ where we define "componentwise addition" above. (p, v_p) + (q, w_q) where p≠q simply wouldn't make sense. It's easy to see that ϕ is a bundle isomorphism, and hence we conclude that the tangent bundle of a manifold is a vector bundle indeed.

We interrupt our development of formalisms to inject an important comment at this junction. Physics should be 'invariant' under any choice of a local coordinate system on spacetime. After all, how nature behaves shouldn't be dependent on what we, as observers of phenomena, pick as our tools to describe them. In the same spirit then, physics is invariant under any local trivialization of whatever vector bundles under consideration when we describe physical fields in spacetime. This is the principle of **gauge invariance** and we'll elaborate on this in more detail later on. But first, let's return to development of some more formalisms.

Given two vector bundles $\pi : E \to M$ and $\pi' : E' \to M'$, a **vector bundle morphism** from first bundle to the second bundle is a bundle morphism $\psi : E \to E'$ such that its restriction to each fiber E_p of E is linear. That is, the morphism restricted to each fiber E_p "preserves" the linearity of the vector space E_p.

Finally, we are one step closer to completing a toolkit to describe fields in spacetime. As aforementioned, physical fields, whether they be electromagnetic fields, or gravitational fields, can be described as 'sections' of vector bundles. A **section** of a bundle $\pi : E \to M$ is simply a function $s : M \to E$ so that $\forall p \in M$, s(p)$\in E_p$. Physically, this means as we 'traverse' in any curve γ on the base space M (which is a smooth manifold), $s \circ \gamma$ traverses through $E_{s \circ \gamma}$. So a section assigns to each point p $\in$ M, a specific vector in the fiber over that point. If you have a trivial bundle E = M $\times$ F with a standard fiber F, then a section of E is just a function $f : M \to F$ such that $s(p) = (p, f(p)) \in E_p$. Or we can define a function f instead, in which case it would induce a corresponding section for the trivial bundle. It's physically obvious that a section of the tangent bundle is a vector field, and we've known all along that a vector field on spacetime manifolds, describe physical fields. Evidently, a physical field can be treated as a 'section'. The subtlety present here is that when we think of vector fields in the Euclidean space, we're really looking at vector fields defined in *one* vector space, while sections describe vectors, or "little arrows" if you will, in *different* spaces, namely the fibers.

We know that certain forces in nature obey the superposition principle. That is, two fields add up vectorially to form the field that we physically observe in nature (i.e. the resultant field). Vector fields can then be added. In a similar spirit then, it only makes sense that sections should have the ability to withstand additions as well. Indeed, they do. Given a vector bundle E over a manifold M, and sections s and s' of E, we define **addition of sections** as:
(s+s')(p) = s(p) + s'(p)
We also define **multiplication of sections by functions** on M as (where f $\in C^\infty(M)$: (fs)(p) = f(p)s(p). $\Gamma(E)$ denotes the set of all sections of E.

Recall that for a vector bundle, E_p is a vector space with all the fibers having the same dimensions. Now, let's imagine the following situation. We have a smooth curve γ in the total space E (remember that E is a smooth manifold), then at some instant t_0, $\gamma(t_0) \in E_{\overrightarrow{x}}$. And we have a basis for the vector space $E_{\overrightarrow{x}}$ at this instant. Then in some infinitesimal time dt later, we have $\gamma(t_0 + dt) \in E_{\overrightarrow{x}+d\overrightarrow{x}}$, and at this instant, each of the basis vectors for the vector space $E_{\overrightarrow{x}}$ has been "perturbed" little to form a basis for the new fiber $E_{\overrightarrow{x}+d\overrightarrow{x}}$. We can consider this as a "moving frame" in which the basis vectors vary smoothly as we traverse through different sections. Indeed, we can make this notion more precise. Given a vector bundle E, **basis of sections** of E is the set of sections $\{e_1, \ldots, e_n\}$ such that any section s $\in \Gamma(E)$ can be *uniquely* expressed as a sum $s = s^i e_i$ (with the Einstein summation notation being used here), where $s^i \in C^\infty(M)$. Although we will not prove here, it turns out that a vector bundle has a basis of sections if and only if it's isomorphic to a trivial bundle. Note, however, that vector bundles, by definition, are locally isomorphic to trivial bundles. Hence, when we perform a local analysis, we can pick a basis of sections over a sufficiently small neighbourhood of any point in the base space.

4.3 Construction of Vector Bundles useful in General Relativity

We've seen above some of the correspondences between vector spaces and vector bundles. This is no coincidence since a vector space is really a vector bundle with base space equal to a single point. That is, we have a very 'direct' isomorphism here. We now develop some notions for vector bundles that are familiar to us in the vector space setting.

Say we have a vector bundle E over M. Each fiber E_p is a vector space and hence it has a well-defined dual space E_p^*. Then we can construct what's called a **dual vector bundle** E^* over M as follows. We let its total space to be $E^* = \cup_{p \in M} E_p^*$, and let the projection $\pi : E^* \to M$ map each E_p^* to p. Hence E_p^* is the fiber over p∈ M. It can be shown that E^* is indeed a manifold and that there's a local trivialization of E^* that's fiberwise linear. Hence E^* is a vector bundle indeed. An important example of a dual vector bundle, for computational purposes, is the **cotangent bundle** T^*M of a manifold M. For this dual vector bundle, its fiber of a point is the cotangent space at the point, and a section of T^*M is a 1-form on M.

Let E and E' be two vector bundles over M. A **direct sum vector bundle** $E \bigoplus E'$ over M has as its fiber over p, $E_p \bigoplus E'_p$. This is a "natural" way to define direct sum of two objects indeed. In a similar spirit, the **tensor product vector bundle** $E \bigotimes E'$ over M has as its fiber over p, $E_p \bigotimes E'_p$. Both the direct sum vector bundle and the tensor product vector bundle are indeed vector bundles.

Now, we reveal perhaps one of the more important constructions that's useful in study of gauge fields, and it's the object called a "G-bundle". Here's how we construct it. First, we start off with some essential ingredients. We have a smooth manifold M along with the open cover $\{U_\alpha\}$ of M. In addition, we select a vector space V, group G, and a representation ρ of G on V, all of which are of our own choosing. So summarizing, our ingredients are: $\{M, \{U_\alpha\}, V, \rho, G\}$. Now, we construct a total space as follows. Consider $T \equiv \bigcup_\alpha U_\alpha \times V$. Now, we 'identify' elements in T as follows. Two points (p,v) $\in U_\alpha \times V$ and (p,v') $\in U_\beta \times V$ are considered 'equal' if $v = \rho(g_{\alpha\beta}(p))v'$, where $g_{\alpha\beta} : U_\alpha \bigcap U_\beta \to G$ is a **transition function**. We denote above condition as $v = g_{\alpha\beta}v'$ for simplicity. Informally, what we're saying is that for each point p $\in$ M, we also have an associated group element g $\in$ G. And associated with this g *acts* on elements of V. If this g acts on v' to get v, then we will only put one of them in the fiber over p, since there's a way to obtain the other vector via the interaction between M and G. Now, there are two constraints on this procedure in order to insure consistencies.

1.) $g_{\alpha\alpha} \equiv 1$ on U_α: For otherwise, we'd have to identify (p,v) and (p, $g_{\alpha\alpha}$v), both of which lie in $U_\alpha \times V$. But we don't want to identify two different elements in the same trivial bundle. This constraint prohibits this situation.

2.) **Cocycle condition -** $g_{\alpha\beta}g_{\beta\gamma}g_{\gamma\alpha} \equiv 1$ on $U_\alpha \bigcap U_\beta \bigcap U_\gamma$. This allows identification of (p,v), (p, $g_{\alpha\gamma}$v), and (p, $g_{\beta\gamma}g_{\gamma\alpha}$v) as one element (p, $g_{\alpha\beta}g_{\beta\gamma}g_{\gamma\alpha}$v).

It turns out that other similar consistency conditions follow from the requirements 1) and 2) above. To finalize our construction of this "special" vector bundle, we denote the equivalence class $[p, v]_\alpha$ to be the "point" of E (rather the representative) corresponding to (p,v) $\in U_\alpha \times V$. We then define the projection $\pi : E \to V$ by $\pi[p, v]_\alpha = p$. Then the fiber E_p is the set of all points in E of form $[p, v]_\alpha$. So we've finally constructed the fibers! Due to the conditions 1) and 2) imposed above, it turns out that $\pi : E \to M$ is a vector bundle. This kind of vector bundle is what we called a **G-bundle** above. The group G is called the **gauge group** of bundle, and V is called the **standard fiber**. As you recall, we've already defined what a gauge group is before. Physical gauge fields are described by sections as mentioned before. More specifically, they're the sections of G-bundles, and by varying the group G, we can describe different forces. This was exactly what we said a gauge group is all about in the previous sections.

5 Gauge Transformations

5.1 Gauge Theory and the Standard Model

	Neutron	Proton
Fermion/Boson	Fermion	Fermion
Spin	$\frac{1}{2}$	$\frac{1}{2}$
Mass	$1839m_e$	$1836m_e$
Electric Charge	0	+e
Stability	Mean life $\approx$ 15 minutes	Absolutely Stable

(Note: -e $\equiv$ charge of an electron, $m_e \equiv$ mass of an electron)

Looking over the table of comparisons between neutrons and protons, one notices that the two types of particles share some stark similarities. Both the neutron and proton belong to a class of particles called **hadrons** which are particles that interact with each other via the strong force. For example, the strong force between proton and neutron keeps the nuclei atoms from breaking apart by overcoming the electric repulsion between any two protons clustered in the nuclei. To account for some of the properties of protons and neutrons, Heisenberg introduced the notion of **isospin**. Using this notion, if we ignore all the interactions between a proton and a neutron resulting from any force other than the strong force, then the two particles can be considered as being merely two different states of a single particle, which we call a **nucleon**.

There are two different spins that a nucleon can have. A proton is arbitrarily assigned a 'isospin-up' state which is represented by a unit vector in C^2, $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$, while the neutron is assigned the 'isospin-down' state, $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$. For this reason, we say that nucleon is an isospin **doublet**. There are different 'spin' representations of SU(2), which are useful in describing different spin states of particles, which we won't go into much detail here.

In the standard model of physics, one makes a distinction between a **force** and an **interaction**. Every force is carried by a **force carrier particle** such as photons, gluons, and pions (for strong force). The result of force carried by these mediator particles from one particle to another results in the 'interaction' between the two particles. There are three pions, each with a positive, negative, or a neutral electric charge. Once again, assuming that the strong force is the only force present, these three pions can be described as three different states of a single particle, the pion. These transform according to so-called 'spin-1' representation of SU(2), and the isospin **triplet**.

Describing the dynamical aspects of a particle requires more than their isotropic spin considerations. We need to analyze the 'matter wave' representation of a particle according to quantum mechanics for a full understanding of the dynamics. For example, for a nucleon, we have a "field" (i.e. matter wave) $\psi : R^4 \to C^2$ on the Minkowski spacetime (C^2 because a nucleon is a 'doublet'), while for a pion we have the 'field' $\phi : R^4 \to C^3$. But the spins of particles depend on the polarization direction of the external force field such as the electromagnetic field. That is, a spin is either up or down *parallel* to the polarization axis of the external magnetic field that's used to 'measure' the spin. This observation, along with the fact that SU(2) is S^3 as a manifold, tells us that if we act on any valid fields ϕ and ψ defined above by an element g $\in$ SU(2), so that we have $\psi'(x) = g\psi(x)$ and $\phi'(x) = g\phi(x)$, then both ψ' and ϕ' are also valid matter waves on Minkowski spacetime. Here, g $\in$ SU(2) acts according to spin-1/2 and spin-1 representations, whose actual forms are not of interest to us here.

In their paper published in 1954, C.N. Yang and R. L. Mills have investigated the conservation of isotopic spin and isotopic gauge invariance. What they noticed was that as mentioned before, when strong force is the sole force under consideration, distinguishing between a neutron and a proton is a purely arbitrary process. But at one space-time point, if one chooses what to call a proton

and what to call a neutron by assigning the aforementioned spins, then one is not able to swap the labels on the two particles at other spacetime points. Yang and Mills thought that this was inconsistent with the localized field concept that underlies the usual physical theories; that this seemingly arbitrary choice at one point in spacetime results in globally affecting the physics of the system in the spacetime bothered them. Hence, they searched for the field equations that possess more symmetry, namely, symmetries under transformations of the form:
$\psi'(x) = g(x)\psi(x)$, and $\phi'(x) = g(x)\phi(x)$
where g:$R^4 \to$ SU(2) is any function that's SU(2)-valued on spacetime. This difficult problem was solved by Yang and Mills by mimicking Maxwell's equations, and resulted in an SU(2) 'gauge theory' of various sorts of hadrons.

Standard model does not take into account gravity, and in fact, there are fundamental inconsistencies that arise when we try to combine general relativity with quantum field theory. Currently, efforts are being put to develop new theories to take gravity into account, and standard model serves as a basis for these candidate theories.

5.2 Gauge Transformations for Vector Bundles

For calculations involved in Yang-Mills equations that we'll perform later on, we need to make the concept of a gauge transformation more precise for vector bundles. This is what we set out to do in this section.

Let V be a vector space. Then the linear functions from V to itself (T:V $\to$ V) are called **endomorphisms** and the set of all endomorphisms of V is denoted **End(V)**. It's easy to see that End(V) is a vector space with usual definitions of addition of two linear functions on V and multiplication by any scalar. It's also an *algebra* with the product of two linear functions S and T defined by (ST)(v) = S(T(v)), where v $\in$ V. Note that End(R^n) is the algebra of $n \times n$ real matrices with matrix multiplication as the product. In fact, more useful definition can be devised by using the basis of V and the dual basis of V^*. We can define End(V) to be $V \otimes V^*$.

Let E be a vector bundle over a manifold M. Then End(E) is called an **endomorphism bundle** of E, which is the bundle $E \otimes E^*$. Let's consider what this is saying in more detail. Given a fiber of End(E) over some a point p $\in$ M, the fiber is same as End(E_p) because End(E) $\equiv E \bigotimes E^*$ by definition, so indeed over p, the fiber will be End(E_p) $\equiv E_p \bigotimes E_p^*$. Therefore, for any section T of End(E), we get a map from E to itself sending v $\in E_p$ to T(p)v $\in E_p$, which is a vector bundle morphism. So in fact, the name 'endomorphism' is quite fitting here since 'endomorphism' in Greek, depicts the notion of sending something *inside* itself, as is the case with End(E).

We need to define one more terminology before defining what "gauge invariance" means rigorously. Say p $\in U_\alpha$ (with U_α defined as in G-bundle). Then T **lives in** the group G (in G-bundle), if it's of form $[p,v]_\alpha \longmapsto [p,gv]_\alpha$ for some g $\in$ G. Now we are finally in a position to define succinctly the notions of "gauge invariance" and "gauge transformations". Let T $\in$ End(E) and p $\in$ M. Then T(p) $\in$ End(E_p) as we saw above. If T(p) lives in G, $\forall$ p $\in$ M, then T is said to be a **gauge transformation**. Ω denotes the set of all gauge transformations, and it's easy, yet tedious, to show that it is a group with products and inverses defined by:
(gh)(p) = g(p)h(p)
g^{-1}(p) = $(g(p))^{-1}$
Gauge theory's based on the foundation that physical fields are sections of G-bundles, and that the systems of differential equations that govern the evolution of the physical systems are **gauge invariant**. That is, if section s is a solution for the system of differential equations, then gs is also a solution for *any* g $\in \Omega$.

6 Connections and Curvature for Vector Bundles: A step towards Yang-Mills Equations

In our course (APM421: General relativity), we've already encountered the notions of connections and curvature. But how do these notions arise in the setting of bundles, which are things we haven't explicitly considered in our course so far? In this section, we address and try to answer this question with minimum amount of formalism as possible, keeping in mind what we already know from our course.

6.1 Introduction to Connections for Bundles

"Differentiating" a section of a vector bundle is not as simple as differentiating any real valued smooth functions because a section assigns vectors in different fibers as one traverses along the base space. This, in turn, results in the "addition" of two vectors assigned by a section to be not well defined in many situations. A 'connection' is simply a way to differentiate sections as a result of this complication. More specifically, a **connection** D on manifold M (base space) assigns to each v $\in$ Vect(M) a function $D_v : \Gamma \to \Gamma$ satisfying the following conditions:
(i) $D_v(\alpha s) = \alpha D_v s$ (Linearity)
(ii) $D_v(s+t) = D_v s + D_v t$ (Linearity)
(iii)$D_v(fs) = v(f)s + fD_v s$ (Leibnitz Rule)
(iv) $D_{v+w}s = D_v s + D_w s$
(v) $D_{fs} = fD_v s$

($\forall v, w \in Vect(M)$, $\forall s, t \in \Gamma(E)$, $\forall f \in C^{\infty}(M)$, and any scalar α).
One might feel uneasy about what we just defined as 'connection' as somehow embodying the notion of 'differentiation'. This, of course, is no coincidence. The Leibnitz and linearity rules of the 'connection' characterizes it to behave like 'differentiation'. For a section s and a vector field v, $D_v s$ is called the **covariant derivative of s in the direction v**. But of course, we knew this well ahead from our previous work in our course, but with different language.

6.2 Vector Potentials

For simplicity, we will denote D_{∂_μ} by D_μ. Using the notation developed for basis of sections, let $\{e_i\}$ denote the basis of sections for the vector bundle, E over U. Then $D_\mu e_i$ is an element of $\Gamma(E)$, so it can be expressed uniquely as a linear combination of the basis of sections, with smooth functions on U as coefficients. That is, letting $A^i_{\mu j}$ denote the coefficients, then we have that $D_\mu e_j = A^i_{\mu j} e_i$. We call $A^i_{\mu j}$ to be the **components of vector potential**. We write A to denote the **vector potential**. There are dual purposes for introducing this vector potential, one of which is that it allows us to compute $D_v s$ for any section E over U. This is done as follows:

$D_v s = D_{v^\mu \partial_\mu} s$ (since any v is expressible in terms of *its* basis.)
$= v^\mu D_\mu s$ (using property (v) of covariant derivatives)
$= v^\mu D_\mu (s^i e_i)$ (expressing section in terms of its basis)
$= v^\mu ((\partial_\mu s^i) e_i + A^j_{\mu i} s^i e_j)$ (using Leibnitz law)
$= v^\mu (\partial_\mu s^i + A^i_{\mu j} s^j) e_i$ (switch indices i and j in the second term)

Looking at the second line in the above calculation, by writing $D_\mu s \equiv (D_\mu s)^i e_i$, then matching up the components, we get:

$$(D_\mu s)^i = \partial_\mu s^i + A^i_{\mu j} s^j$$

The second purpose of introduction of the vector potential is physical. In classical electrodynamics, we introduce the **magnetic vector potential** which encodes the information about the underlying magnetic field. Ironically, as in classical electrodynamics, the vector potential *by itself* has *no* physical meaning. It's only when we perform certain operations on them that the *actual* physical values arise. This is clarified in further detail below and leads us to defining "End(E)-valued 1-form".

As it stands at the moment, the vector potential seems to depend on local coordinates and hence a local trivialization of the bundle E under consideration. But we can obtain a "global" picture of vector potentials by "sewing" all the different fibers together. That is, view it as a section of a certain bundle. In the last line of above calculation, we note that we obtain a form of $A^j_{\mu i} v^\mu s^i e_j$, and that this is a section of E over U. It's also linear in both v and s, and linear over $C^\infty(U)$. In light of this, we can view a vector potential that acts on a vector field and a section of E over U, and produce a new section of E over U in a smooth (i.e. $C^\infty(U)$) manner. Hence, we can think of A as an **End(E)-valued 1-form** on U. In other words, a vector potential is a section of the bundle $End(E\mid_U)\bigotimes T^*U$. Here's an explicit calculation to clarify this notion:

In our claim above, we're led to define a vector potential A as:
$A = A^j_{\mu i} e_j \bigotimes e^i \bigotimes dx^\mu$
Now, let A 'act' on a vector field v on U:
$A(v) = A^j_{\mu i}(e_j \bigotimes e^i) dx^\mu(v) = A^j_{\mu i} v^\mu (e_j \bigotimes e^i)$
But $A^j_{\mu i} v^\mu (e_j \bigotimes e^i)$ is a section of End(E) over U. Now, letting A *then* act on a section s of E over U, we see that:
$A(v)s = A^j_{\mu i} v^\mu s^i e_j$
and above is a section of E over U again.

From above calculation, combining with the computation for $D_v s$ done previously, we see that
$(D_v s)^i = vs^i + (A(v)s)^i$
For notational convenience, we suppress the internal indices and write the vector potential in terms of its components A_μ as $A_\mu = A^j_{\mu i} e_j \bigotimes e^i$, with each component being a section of End(E) over U. In light of all this then, we see that gauge theory, in a sense, generalizes classical electromagnetism by allowing more freedom in the vector potential involved in the physical theory. Namely, it allows vector potentials to be more than just 1-forms.

6.3 Curvature

We've treated the notion of 'curvature' in a considerable amount of detail in our course, but here we remind the reader very briefly what it is in our new 'endomorphism' setting.

We first remind ourselves what a 'curvature' is. Given two vector fields v and w on a smooth manifold M, with E being a vector bundle over M with connection D, **curvature** F(v,w) is an operator that acts on sections s of E, with:
$F(v,w)s = D_v D_w s - D_w D_v s - D_{[v,w]} s$

It's of course possible to define p-forms, other than p=1 on endomorphisms and this is what we set out to do here. But first, we remind ourselves what an "exterior algebra" of a vector space is. Given a vector space V, the **exterior**

algebra over V, denoted $\bigwedge V$, is the algebra generated by V with the anti-commutation relation $v \wedge w = -w \wedge v$, $\forall v, w \in$ V. $\bigwedge V$ is in fact an algebra as can be easily seen, because in its definition, we see that we "start" with a certain set (in fact a vector space V), then generate all possible elements through the formal products of the form $v_1 \wedge ... \wedge v_p$ then take all possible linear combinations of them, putting each of them in our $\bigwedge V$, (where $\wedge$ denotes the usual 'wedge product') with the constraint of the anti-commutation relation define above. With this definition, we construct, for a vector bundle E over M, an **exterior algebra bundle** $\bigwedge E$ over M. $\bigwedge E$ over M is defined to have a total space $\bigcup_{p \in M} \bigwedge E_p$, and the projection map $\pi : \bigwedge E_p \to p$. Though we won't do this here, $\bigwedge E$ can be made into a manifold such that $\pi : \bigwedge E \to M$ is a vector bundle. In the specific instance of a tangent bundle TM, we form the dual (cotangent bundle) T^*M, then we can form its exterior algebra bundle $\bigwedge T^*M$. This special exterior algebra bundle is given its own name, **form bundle**. A differential form on a manifold M is then a section of the form bundle $\bigwedge T^*M$.

To derive Yang-Mills equations, we really need to do some "calculus" with connections and curvatures, as we'll see soon. Towards that end, we define **End(E)-valued differential forms**. We've already defined 1-forms of this type and here we generalize the notion to higher degree forms. These are the sections of the bundle $End(E) \bigotimes \bigwedge T^*M$, or its restriction to some open set in M. Furthermore, given a vector bundle E over a smooth manifold M with a connection D, we define **E-valued p-form** to be a section of $E \bigotimes \bigwedge^p T^*M$. Then an E-valued 0-form is just a section of E as we see. Last but not least, we can define **exterior covariant derivative** d_D of E-valued differential forms. For an E-valued 0-form (i.e. section) of E, we define it to be

$$d_D s(v) = D_v s, \forall v \in Vect(M)$$

Note that this is a generalization of the well-known formula df(v) = v(f). In local coordinates x^μ on an open subset U $\subseteq$ M, we have: $d_D s = D_\mu s \otimes dx^\mu$. We can now define the same notion on arbitrary E-valued differential forms. Let s be a section of E and ω be an ordinary differential form on M. Then we define it as:

$$d_D(s \otimes \omega) \equiv d_D s \wedge \omega + s \otimes d\omega$$

More importantly, if ω and μ are ordinary differential forms, then
$(T \otimes \omega) \wedge (s \otimes \mu) = T(s) \otimes (\omega \wedge \mu)$.

7 The Yang-Mills Equation

Recall from our course work, that one of the two Maxwell's equations (in differential forms setting) was just the result of the **Bianchi identity**. In this section, we show what Yang-Mills equations are, and how they can be derived from generalizing the Maxwell's equations. In addition, we derive Yang-Mills equations from two different perspectives. One will be by considering fields in space directly, the second approach will be a Lagrangian formulation.

7.1 The Hodge Star Operator and Maxwell's Equations

We first introduce the **Hodge Star Operator** which will allow us to write down our equations in a very compact form. Looking over the Maxwell's equations for classical electromagnetic fields, we notice that there's a "duality symmetry" in the equations. That is, we have the following Maxwell's equations:

$$d_s B = 0; \qquad \partial_t B + d_s E = 0 \tag{1}$$

and,

$$d_s E = \rho; \qquad d_s B - \partial_t E = 0 \tag{2}$$

where E and B are electric and magnetic fields, t is time, and s is the spatial coordinate s=(x,y,z). Then looking at the above equations, if we switch E and B in (1) with a sign change, then we "almost" obtain the second pair of Maxwell's equations (2). So we see that this "symmetry" is not quite perfect. The physical reason for this "symmetry" being not "complete" is that there are no magnetic charges in the theory of classical electrodynamics where as electric charges do exist. Experimentalists have been trying to find the "magnetic monopoles" but have not been successful yet. Physicists such as Paul Dirac and Julian Schwinger have expounded on the existence of magnetic monopoles. The lack of magnetic monopoles is not just a matter of nuisance. In the 1930's, Dirac published a short paper in which he showed that the existence of magnetic charges would in fact explain why the electric charge is quantized. From a mathematician's point of view, one sees that the incompleteness of the "symmetry" described above is due to the fact that E is a 1-form on space and B is a 2-form on space in (1), where as in (2), E acts as a 2-form and B acts as a 1-form. So it seems that if we can somehow "adjust" E and B into appropriate degree of forms, then this symmetry might be completed, save for the fact that there's still no magnetic monopole. Hodge star's job is to do just that when dealing with equations embodying this type of "near-symmetry". It converts 1-forms in 3-dimensional space into 2-forms and vice versa. To do so, it requires a metric and a choice of orientation on the manifold.

Let's now formally define the **Hodge star operator for differential forms**. On a 3-dimensional Riemannian manifold M, at any point p $\in$ M and a 1-form v associated with p, we define:

$$\star v = \omega \wedge \mu$$

Here, ω and μ are both 1-forms with $\omega \wedge \mu$ being *orthogonal* to v. Conversely, the Hodge star operator maps $\omega \wedge \mu$ to v. Geometrically, in a 3-dimensional space, we can interpret what the Hodge star operator does as follows. We can certainly view v as a little arrow sticking out of point p $\in$ M. Then $[\omega, \mu, v]$ form a basis of the 3-dimensional space with $\omega \wedge \mu$ representing the area element of the parallelogram spanned by ω and μ, with v being orthogonal to the plane of the parallelogram.

We can generalize the construction of the Hodge star operator to n-dimensions as follows:

$\star$: p-forms $\rightarrow$ (n-p)-forms

and it acts in a **linear** manner. It takes each "p-dimensional area element" to an orthogonal "(n-p)-dimensional area element". To make this precise, we need an inner product of differential forms. So let's do that. Let M be an n-dimensional oriented semi-Riemannian manifold. We define an **inner product of two p-forms** ω and μ on M to be $< \omega, \mu >$ as a function on M. With this notion of inner product, we then define the **Hodge star operator** $\star : \Omega^p(M) \rightarrow \Omega^{n-p}(M)$ as a unique linear map such that $\forall \mu, \nu \in \Omega^p(M)$

$\star(\omega \wedge \mu) \equiv (\omega \wedge \star\mu) =< \omega, \mu > vol$

where $\Omega^p(M)$ is the set of all p-forms on M and vol is the volume form associated with the metric on M. For example, vol at a point $\in$ M is just $vol_p = e_1 \wedge ... \wedge e_n$, where $[e_1, ..., e_n]$ is an oriented orthogonal basis of cotangent vectors at p. Note that $\omega \wedge \star\mu$ is an n-form because $\star\mu$ is an (n-p)-form and ω is a p-form. We say that $\star\mu$ is a **dual** of μ.

Now that we've defined the Hodge star operator, what we want to do is get familiar with using it explicitly in computations involving differential forms (and later on, involving vector bundles). By working out some explicit examples, we'll try to convince ourselves that the Hodge star operator is indeed well-defined, and also get some perspectives on some of its important properties.

First, we state an equivalent definition of Hodge star operator that will be computationally more useful. Let $[e_1, ..., e_n]$ denote a positively-oriented orthonormal basis of 1-forms on some chart of semi-Riemannian manifold M. Hence $< e^\mu, e^\nu > = 0$ if and only if $\mu \neq \nu$, and $< e^\mu, e^\mu > = \epsilon(\mu)$, where $\epsilon(\mu) = \pm 1$ with the sign depending on the chosen orientation. We claim that for any *distinct* set of indices $1 \leq i_1 ..., i_p \leq n$:

$\star(e^{i_1} \wedge ... \wedge e^{i_p}) = \pm e^{i_{p+1}} \wedge ... \wedge e^{i_n}$

where $\{i_{p+1}, ..., i_n\} = \{1, ...n\} - \{i_1, ..., i_p\}$

Which sign (either + or -) is assigned above is determined by:

sign$(i_1, ..., i_n)\epsilon(i_1)...\epsilon(i_p)$

where sign$(i_1, ..., i_n)$ denotes permutation that takes (1,...,n) to $(i_1, ..., i_n)$. We claim that this working definition of Hodge star is equivalent to the one given previously, though we won't prove the equivalence here.

The definition given above seems even more complicated than the original definition of Hodge star, and in fact, it's probably correct to say so. But from above definition, we can deduce the following very nice properties for 1-forms dx, dy, and dz in R^3, taking the standard Euclidean metric and orientation:

$$\star dx = dy \wedge dz, \qquad \star\, dy = dz \wedge dx, \qquad \star\, dz = dx \wedge dy \tag{3}$$

And from above, we can further deduce the following, since Hodge star was defined to be an *inverse* of itself:

$$\star(dx \wedge dy) = dz, \qquad \star\,(dz \wedge dx) = dy, \qquad \star\,(dy \wedge dz) = dx \tag{4}$$

For any 0-form 1 and volume form $dx \wedge dy \wedge dz$, we have the following properties:

$$\star 1 = dx \wedge dy \wedge dz, \qquad \star\,(dx \wedge dy \wedge dz) = 1 \tag{5}$$

From now on, for notational convenience, we'll just write $\star\omega \wedge \mu$ instead of $\star(\omega \wedge \mu)$. When we compute expressions involving the Hodge star operator in R^3 space, we only need to refer to (3), (4), and (5) to grind through our calculations. This is why the second definition of Hodge star is useful indirectly.

Let's work out some explicit calculations with the Hodge star operator to get a "feel" for it. While still working in R^3 with standard metric and orientation, let $\omega = \omega_i dx^i$ and $\nu = \nu_i dx^i$ be 1-forms defined on R^3. Then keeping in mind (3), (4), and (5), we compute $\star(\omega \wedge \nu)$ in the following manner:

$$\star(\omega \wedge \nu) = \star[(\omega_x\nu_y - \omega_y\nu_x)dx \wedge dy + (\omega_y\nu_z - \omega_z\nu_y)dy \wedge dz + (\omega_z\nu_x - \omega_x\nu_z)dz \wedge dx] \tag{6}$$

but $\star$ acts linearly so we obtain

$$\star(\omega \wedge \nu) = \star[(\omega_x\nu_y - \omega_y\nu_x)dx \wedge dy] + \star[(\omega_y\nu_z - \omega_z\nu_y)dy \wedge dz] + \star[(\omega_z\nu_x - \omega_x\nu_z)dz \wedge dx] \tag{7}$$

then invoking the *original* definition of Hodge star

$$\star(\omega \wedge \nu) = (\omega_x\nu_y - \omega_y\nu_x)\star dx \wedge dy + (\omega_y\nu_z - \omega_z\nu_y)\star dy \wedge dz + (\omega_z\nu_x - \omega_x\nu_z)\star dz \wedge dx \tag{8}$$

which in turn becomes

$$\star(\omega \wedge \nu) = (\omega_x\nu_y - \omega_y\nu_x)dz + (\omega_y\nu_z - \omega_z\nu_y)dx + (\omega_z\nu_x - \omega_x\nu_z)dy \tag{9}$$

which is just the familiar **cross-product** from vector calculus.

Let's do one more important example. We'll end up deriving another well known operation in vector calculus. Using R^3 again with the same standard orientation and metric, let ω be a 1-form in R^3. Then we compute the 1-form $d\omega$ as follows (we'll work step-by-step slowly):

$$\star d\omega = \star(d\omega_i \wedge dx^i) = \star(\partial_\mu \omega_i dx^\mu \wedge dx^i) \tag{10}$$

which in turn becomes

$$= \star(-\partial_y\omega_x dz + \partial_z\omega_x dy + \partial_x\omega_y dz - \partial_z\omega_y dx - \partial_x\omega_z dy + \partial_y\omega_z dx) \tag{11}$$

$$= \star[(\partial_x \omega_y - \partial_y \omega_x)dz + (\partial_z \omega_x - \partial_x \omega_z)dy + (\partial_y \omega_z - \partial_z \omega_y)dx] \tag{12}$$

then using the linearity of $\star$,

$$= (\partial_x \omega_y - \partial_y \omega_x) \star dz + (\partial_z \omega_x - \partial_x \omega_z) \star dy + (\partial_y \omega_z - \partial_z \omega_y) \star dz \tag{13}$$

Then recalling (3) we have,

$$\star d\omega = (\partial_x \omega_y - \partial_y \omega_x) dy \wedge dz + (\partial_z \omega_x - \partial_x \omega_z) dz \wedge dx + (\partial_y \omega_z - \partial_z \omega_y) dx \wedge dy \tag{14}$$

which is just the **curl** of ω in vector calculus. In a similar manner, a computation of above type will prove that $\star d \star \omega$ is the **divergence** of ω.

Above lengthy computations are not in vain, for we can now use them to *reduce* Maxwell's equations (1) and (2) down to the following:

$$dF = 0, \qquad \star\, d \star F = J \tag{15}$$

where J is the electric current, and F is the electromagnetic field. So F encodes the information about both the electric field E and the magnetic field B.

7.2 Derivation of the Yang-Mills Equation via field-theoretic method

In the previous section, we managed to reduce the Maxwell's equations to a more compact version by using the Hodge star operator defined on differential forms. We noted that F is the electromagnetic field that encodes both the electric E and magnetic B fields. Then we can 'separate' F into its magnetic and electric parts on Minkowski spacetime

$$F = B + E \wedge dt \tag{16}$$

Here, B and E are time-dependent 2-forms and 1-forms on space respectively. In certain physical situations, we can write the entire electromagnetic field F as an *exterior* derivative of some vector potential A, as noted previously,

$$F = dA \tag{17}$$

Then we note that since d(dV)=0 for any vector potential V, that in fact the first of Maxwell's equation (16) follows directly from this general fact about differential forms without recourse to any physical reasoning embodied in electromagnetic fields. We remember, from our course, that this gives rise to the well known Bianchi identity:

$$d_D F = 0 \tag{18}$$

where D is any well-defined connection.

In order to generalize the other half of Maxwell's equation, we need to introduce the Hodge star operator for *endomorphism-valued* differential forms. Towards this end, we let $\pi : E \to M$ be a vector bundle over an oriented semi-Riemannian manifold M. We define the **Hodge star operator** $\star$ acting on End(E)-valued differential forms to be the $C^\infty(M)$-linear operator such that for any section T of End(E) and differential form ω:

$$\star(T \otimes \omega) = T \otimes \star\omega \tag{19}$$

where the $\star$ on the right hand side of the equation denotes the Hodge star operator that was already defined on ordinary differential forms. Note the striking similarity between the Hodge star operators on ordinary differential forms and End(E)-valued differential forms.

Using above definition, for any End(E)-valued 1-form J defined on M, which is to represent a "current", the so-called **Yang-Mills equation** with **Yang-Mills field** F is

$$\star d_D \star F = J \tag{20}$$

Now, let's explore some of the physical meanings of this equation. Suppose E is the trivial U(1)-bundle (recall the definition of G-bundle for a Lie group G), equipped with the standard fiber given by the fundamental representation of U(1). Recall that we've worked out explicitly this fundamental representation in detail in the section dealing with Lie groups. In this case, we see right away that (20) reduces to the second of the familiar Maxwell's equation (15). So we've seen here a remarkable example of how the Yang-Mills equation can reduce to the governing equations for known physical fields. The Yang-Mills equation (20) can be reduced to equations describing other gauge field models like the strong force, and weak force. This is exactly what we have been advertising from the beginning to show.

We might argue that this all seems a bit too abstract. What are some more physical pictures that (20) can depict? After all, it's important to keep in mind the different physical meanings contained in (20), and there's a lot of them. But to unravel them, it would be helpful to 'separate' spacetime into space and time components. We consider a 4-dimensional *static* spacetime M, where M = $R \times S$. Here, R is to represent the time component. In such a case, we can also separate the metric g on M into its time and space component $\widehat{g}$:

$$g = -dt^2 + \widehat{g} \tag{21}$$

Furthermore, in analogy with the Maxwell's equations, we can may express the Yang-Mills field F in terms of its separate components

$$F = B + E \wedge dt \tag{22}$$

where B is the Yang-Mills version of **magnetic field** and E is the Yang-Mills version of **electric field**. Here, B is still an End(E)-valued 2-form, with accordance with our previous statement that physical fields are sections of G-bundles.

E, in turn, is an End(E)-valued 1-form on space. We'll stick with the convention of denoting electric fields by E, and total space also by E. But keep in mind that the 'E' in 'End(E)' is *not* the same as E representing the electric field! We also break down the current J into space and time components as

$$J = j - \rho dt \tag{23}$$

where j is an End(E)-valued 1-form on space and ρ is a section of End(E). Both of them are time dependent functions. Remembering that D is a connection and our definition of exterior covariant derivatives associated with connections, we can even decompose the exterior covariant derivative into its space and time parts as follows:

$$d_D\omega = dt \wedge D_t\omega + d_S\omega \tag{24}$$

Where S denotes the spatial part (x,y,z) of course. Furthermore, if we recall the Bianchi identity for connection D, we can separate that too into its respective space and time components using (24) to get

$$d_D F = d_S B + dt \wedge (D_t B + d_S E) = 0 \tag{25}$$

But the two components are 'separate', so above equality can only hold if both the spatial and time components vanish

$$d_S B = 0, \qquad D_t B + d_S E = 0 \tag{26}$$

Now, using $\star_S$ to denote the Hodge star operator on End(E)-valued forms on S, we have, from (25) and (22),

$$\star F = \star(B + E \wedge dt) = \star_S E - \star_S B \wedge dt \tag{27}$$

Then combining (27) with (20), we obtain

$$\star d \star F = -D_t E - \star_S d_S \star_S E \wedge dt + \star_S d_S \star_S B = j - \rho dt \tag{28}$$

Once again, as in (26), equating the space and time components, we see that

$$\star_S d_S \star_S E = \rho, \qquad - D_t E + \star_S d_S \star_S B = j \tag{29}$$

We now show one of the important properties of Yang-Mills equations, **gauge invariance**. We've defined gauge invariance before, but recall that this means that if we have a connection D that satisfies the Yang-Mills equations, then any *gauge transformation* of D also satisfies it as well. This is not an easy matter and will take some work on our part. But here it goes.

7.3 Gauge Invariance of Yang-Mills Equation

As above, let D be a connection on the Yang-Mills electric field E, which is an End(E)-valued 1-form on space. We let g be a gauge transformation which was defined in the early part of our program. Furthermore, let D' denote a gauge transformation of D by g, where

$$D'_v s = gD_v(g^{-1}s) \tag{30}$$

with s being a section and v a vector field. We've defined what a curvature of connections is in the previous section. Recalling the definition, let F denote the curvature of D and let F' denote the curvature of D'. Then we have by definition of curvature,

$$F'(u,v)s = D'_u D'_v s - D'_v D'_u s - D'_{[u,v]}(g^{-1}s) \tag{31}$$

then substituting (30) into (31), we obtain

$$F'(u,v)s = gD_u D_v(g^{-1}s) - gD_v D_u(g^{-1}s) - gD'_{[u,v]}(g^{-1}s) \tag{32}$$

but this is just the definition of curvature on $g^{-1}s$ acted on by g,

$$F'(u,v)s = gF(u,v)(g^{-1}s) \tag{33}$$

We denote this as $F' = gFg^{-1}$ to be economical with our notations.

Now, we use our work above to show that indeed Yang-Mills equation is gauge invariant. Let F be the field satisfying the Yang-Mills equation (20). With the same notations as above, and using the definitions of Hodge star operator given for End(E)-valued forms and curvature of connections given earlier, we get that

$$\star d_D \star F = \frac{1}{2}[D_\mu, F_{\nu\lambda}] \otimes \star(dx^\mu \wedge \star(dx^\nu \wedge dx^\lambda)) \tag{34}$$

This looks rather complicated but it follows strictly from the definitions if we just substitute in. More specifically, D_μ is the covariant derivative on sections of E, and*not* on sections of End(E). Then replacing D and F with the gauge transformed connection D' and curvature F', we work out that

$$\star d_{D'} \star F' = \frac{1}{2}[D'_\mu, F'_{\nu\lambda}] \otimes \star(dx^\mu \wedge \star(dx^\nu \wedge dx^\lambda)) = \frac{1}{2}[gD_\mu g^{-1}, gF_{\nu\lambda}g^{-1}] \otimes \star(dx^\mu \wedge \star(dx^\nu \wedge dx^\lambda)) \tag{35}$$

where in the last part, we've simply invoked the definition of D' and F', both of which were previously defined. We can simplify above expression even further by noting that

$$[gD_\mu g^{-1}, gF_{\nu\lambda}g^{-1}] = gD_\mu g^{-1}gF_{\nu\lambda}g^{-1} - gF_{\nu\lambda}g^{-1}gD_\mu g^{-1} = gD_\mu F_{\nu\lambda}g^{-1} - gF_{\nu\lambda}D_\mu g^{-1} = g[D_\mu, F_{\nu\lambda}]g^{-1} \tag{36}$$

Then (35) can be further simplified to

$$\star d_{D'} \star F' = \frac{1}{2}g[D_\mu, F_{\nu\lambda}]g^{-1} \otimes \star(dx^\mu \wedge \star(dx^\nu \wedge dx^\lambda)) = g(\star d_D \star F)g^{-1} \tag{37}$$

but $g(\star d_D \star F)g^{-1}$ is just the gauge transformed current J'. Hence we have shown that

$$\star d_{D'} \star F' = J' \tag{38}$$

That is, we have shown that when we gauge transform D, F, and J, we still satisfy the Yang-Mills equation (38). So what we have shown here is that Yang-Mills equation is indeed **gauge-invariant**.

As is the case in classical electrodynamics, gauge invariance is a useful concept in that it allows us to pick some specific gauge transformations such that our equations look 'nicer' and simpler. In electrodynamics, for instance, the Lorentz gauge and Coulomb gauge simplify the Maxwell's equations tremendously, saving us tons of calculations. In a similar spirit, we can use gauge-invariance of Yang-Mills equation to write (28) in a nicer looking form on Minkowski space. First, we note that every vector bundle in R^n is trivial as was explained before. We can use this fact to our advantage and write any given connection D in terms of the **standard flat connection** and a vector potential. Remember that we've dealt with standard flat connection in our course. Componentwise, this would be

$$D_\mu = \partial_\mu + A_\mu \tag{39}$$

where A is a vector potential. Through a concept called **temporal gauge**, which we won't go into here, it turns out that we can always select A such that the time component A_0 is zero. Then (39) becomes

$$D_t = \partial_t, \qquad D_i = \partial_i + A_i \qquad (i = 1, 2, 3) \tag{40}$$

Then substituting in (40) into (26), we obtain

$$E = E_i dx^i, \qquad B = \frac{1}{2}\epsilon_{ijk} B^i dx^j \wedge dx^k \tag{41}$$

where we have $E_i = -\partial_t A_i$ and $B^i = \epsilon^{ijk}(\partial_j A_k - \partial_k A_j + [A_j, A_k])$. So note that what we have done in (41) was really nothing more than equating components. Now, we can proceed even further if we consider A to be an End(E)-valued 1-form on space dependent on time. Then the Bianchi identity reduces to

$$\partial^i B_i + [A^i, B_i] = 0, \qquad \partial_t B^i + \epsilon^{ijk}(\partial_j E_k + [A_j, E_k]) = 0 \tag{42}$$

and in light of this, the Yang-Mills equation is now

$$\partial^i E_i + [A^i, E_i] = \rho, \qquad -\partial_t E^i + \epsilon^{ijk}(\partial_j B_k + [A_j, B_k]) = j^i \tag{43}$$

We can immediately see the striking resemblance between (43) and the Maxwell's equations. After all, this is expected since we've shown that Maxwell's equation is merely a Yang-Mills equation when we choose the underlying Lie group to be U(1). The difference between (43) and the Maxwell's equation is that (43) is nonlinear when the commutators involved in them do not vanish. But this occurs if and only if the underlying gauge group on which the Yang-Mills equation (43) is based on is **nonabelian**. Recall that at the very beginning of our program, when we first introduced the notion of Lie groups, we claimed that the nonabelian nature of the Lie group is intimately linked with the Yang-Mills equation becoming nonlinear. We see right here that indeed this is the case.

7.4 Derivation of Yang-Mills Equation via a Lagrangian formulation

Here we derive Yang-Mills equation through its Lagrangian, rather than using vector fields as we have previously. This method will reveal the symmetries and several important properties of the equation more vividly. The Lagrangian formulation also helps us to "quantize" fields, which is the basis of quantum field theory.

We first ask ourselves where we should start. We start by noticing that the Yang-Mills equation is founded upon its underlying vector bundle. Namely, let E be a vector bundle over a semi-Riemannian oriented n-dimensional manifold M. Now, we'll show that the Yang-Mills Lagrangian is an n-form that's to be integrated over M to get its associated *action*.

Towards formulating the Yang-Mills Lagrangian, it turns out to be necessary to define the *trace* of an End(E)-valued form. We can define this notion in analogy with the trace of a square matrix. Note that V is a vector space, so we can select a basis for it and write elements of End(V) as matrices and define our desired trace to be the trace on these matrices. This, however, turns out to be unnecessary because our trace turns out to be invariant under change of basis. We have shown previously that End(V) is isomorphic to $V \bigotimes V^*$, and this isomorphism is independent of choice of basis. Then we can define the following linear map

$$tr : End(V) \rightarrow R, \quad where tr(v \otimes f) \rightarrow f(v) \tag{44}$$

Let's pick some basis $\{e_i\}$ of V and let $\{e^j\}$ be a dual basis of V^*. Then pick T $\in End(V)$. Note that we can express $T = T^i_j e_i \otimes e^j$ as shown when we introduced the notion of endomorphisms in the previous section. Then we have

$$tr(T) = T^i_j e^j(e_i) = T^i_j \delta^j_i = T^i_i \tag{45}$$

which is the sum of the diagonal entries. So for a section T of End(E), we can define tr(T) on the base space M such that for p $\in$ M, tr(T)(p) = tr(T(p)). Note that T(p) is an endomorphism of the fiber E_p. Similarly, we define trace of an End(E)-valued form as follows:

$$tr(T \otimes \omega) = tr(T)\omega \tag{46}$$

where T is a section of End(E) and ω is a differential form.

Using above definition of trace, given a connection D on E, and a curvature F of D, the **Yang-Mills Lagrangian** is defined to be

$$L_{YM} = \frac{1}{2} tr(F \wedge \star F) \tag{47}$$

How we derive (47) from *physical* principles is quite a delicate matter that is beyond the scope of this paper. To acquire the **Yang-Mills action**, we simply need to integrate (47) over M to get

$$S_{YM} = \frac{1}{2} \int_M tr(F \wedge \star F) \tag{48}$$

Then using the *Euler-Lagrange equations*, we obtain the Yang-Mills equation that we got through the field-theoretic method in the previous section. We choose do not go into the actual derivation in this paper.

8 Epilogue

In this paper, we have tried to learn some basic facts about Lie groups, Lie algebras, vector bundles, and gauge theories while sampling different snippets of physics whenever possible. We caution the reader, however, that each of the aforementioned areas constitutes a vast field of study in itself and that this paper serves no justice to any one of them. We hope that we have been successful in wetting the reader's appetite in each of these different, yet linked areas and sparked his or her curiosity. For more information on these areas, we suggest that the reader consult some of the materials listed as references in this paper. We have found the book "*Gauge fields, knots and gravity*" by J. Baez and J. P. Muniain to be especially helpful in the development of this paper. As one can see just from the amount of material we've presented here, gauge theory is quite a mature subject. Yet, it's also important to point out that many parts of gauge theory have yet to be developed, and its development is crucial to understanding some of the most fundamental questions about nature, especially about gravity. The next revolution in physics will undoubtedly come in part from extension of some of the ideas presented here to the development of a coherent theory for gravity. None other than C. N. Yang himself has said it best and we end our treatise with a quote from him; *"Electromagnetism is, as we have seen, a gauge field. That gravitation is a gauge field is universally accepted, although exactly how it is a gauge field is a matter still to be clarified."*

References

J. Baez, J. P. Muniain. *Gauge Fields, Knots and Gravity.* World Scientific. New Jersey, 1994.

C. N. Yang, R. L. Mills. *Conservation of Isotopic Spin and Isotopic Gauge Invariance.* Physical Review Vol. 96, No. 1, Oct. 1, 1954.

S. H. Friedberg, A. J. Insel, L. E. Spense. *Linear Algebra*, 3rd Ed., Prentice Hall, New Jersey, 1997.

D. J. Griffiths. *Introduction to Electrodynamics*, 3rd Ed., Prentice Hall, New Jersey, 1999.

M. Peskin, D. Schroeder. *An Introduction to Quantum Field Theory*, Westview Press, 1995.

G. E. Simmons. *Differential Equations with Applications and Historical Notes*, 2nd Ed., McGraw-Hill, 1990.