

# An Introduction to Gauge Theory

Department of Physics, Drexel University, Philadelphia, PA 19104

## Quantum Mechanics II

Frank Jones

### Abstract

Gauge theory is a field theory in which the equations of motion do not change under coordinate transformations. In general, this transformation will make a problem easier to solve as long as the transformation produces a result that is physically meaningful. This paper discusses the uses of gauge theory and its applications in physics. Gauge transformations were first introduced in electrodynamics, therefore we will start by deriving a gauge transformation for electromagnetic field in classical mechanics and then move on to a derivation in quantum mechanics. In the conclusion of this paper we will analyze the Yang Mills theory and see how it has played a role in the development of modern gauge theories.

## 1 Introduction

From the beginning of our general physics class we are taught, unknowingly, the ideas of gauge theory and gauge invariance. In this paper we will discuss the uses of gauge theory and the meaning of gauge invariance. We will see that some problems have dimensions of freedom that will allow us to manipulate the problem as long as we apply transformations to the potentials so that the equations of motion remain unchanged. Examples of gauge transformations will be discussed including electromagnetism, the Loterntz condition, and phase transformations in quantum mechanics. Finally Yang-Mills gauge theory will be addressed as an analogy to classical electromagnetic gauge transformations. We begin by defining a gauge theory as a field theory in which the equations of motion remain unchanged after a transformation. A local gauge is a coordinate system that can change from point to point in space-time, while a gauge transformation allows one to move from one location in the coordinate system to another. The ability to move from point to point in this coordinate system without changing the equations of motion is known as gauge invariance. We will try to explain this concept by beginning our discussion with two simple examples.

### 1.1 Gauge Freedom, Transformation, and Invariance In two Simple Examples

In a general physics course the discussion of gravitational potential energy is often out first introduction to a gauge transformation. The potential energy of a mass,  $m$ , suspended a height,  $h$ , is given by the equation:

$$V = mgh$$

In general, we take  $h$  to be the distance above the ground the mass is suspended. However, we could just as reasonably have taken it to be the height above a table, or from the center of the earth, and we normally choose a base location that simplifies the problem we are trying to solve. In fact, we have considerable freedom in selecting the origin of our coordinate system. This is known as gauge freedom. If we choose the floor to be the origin of our system (which we have the freedom to do), but would like to measure the potential energy of a mass with respect to a table, 1 meter tall, we could just as easily move the origin from the ground to the table through a gauge transformation. A gauge transformation says that we are allowed to transform coordinate system, and in this case the potential, so long as the equations of motion remain unchanged. In our case, adding a constant potential to shift the origin off of the ground and to the table is an example of a gauge transformation. As long as this transformation is carried out throughout the problem, the physics of the problem will remain unchanged. This is known as gauge invariance.

Another simple analogy to a gauge transformation can be seen in introductory electronics. In electronics, the absolute value of a potential has little meaning. Instead we are often concerned with the potential difference between two points. For example, a battery has a potential difference of 1.5V. A gauge transformation allows us to set the potential at one side of the battery to any arbitrary value, say 100V. However, we must also add this 100V potential to the other side of the battery, giving 101.5V. In the end, the potential difference is still 1.5V and the physical system will remain invariant.

## 2 Gauge Transformations In Electrodynamics

In general, a gauge transformation will make a problem easier to solve by exploiting symmetries in a physical system. By changing coordinates, our example of gravitational potential energy became simpler to solve mathematically. Up until now, we have made transformations by adding a constant to the potential. This has allowed us to preserve the forces on a system:

$$F = -\nabla\phi \tag{1}$$

We can transform  $\phi$  by adding a constant  $C$ :

$$\phi' = \phi + C$$

$$F = -\nabla\phi' = -\nabla(\phi + C) = -\nabla\phi$$

And the force is conserved. However, this may not be the case in more advanced physics. From Maxwell's Equations we know

$$\nabla \cdot B = 0 \tag{2}$$

Where  $B$  is the magnetic field. Knowing this, we can write:

$$B = \nabla \times A \tag{3}$$

where  $A$  is the vector potential of the magnetic field (analogous to  $\phi$ , the scalar potential of an electric field). There is a considerable degree of freedom with which we can determine  $A$ . The addition of any gradient of a scalar function to the vector potential will restore the original equation for the magnetic field

$$A' = A + \nabla\chi \quad (4)$$

$$B = \nabla A' = \nabla \times (A + \nabla\chi) = \nabla \times A$$

This gauge transformation leaves the magnetic field invariant.

It is also important to note that this transformation must be carried through the entire physical system, and will therefore cause a change in the electric field. We know

$$E = -\nabla\phi - \frac{\partial A}{\partial t} \quad (5)$$

Substituting our new  $A'$  into this equation yields:

$$E = -\nabla\phi - \frac{\partial A}{\partial t} - \frac{\partial \nabla\chi}{\partial t} = E - \frac{\partial \nabla\chi}{\partial t}$$

This does not result in field invariance. Clearly something is wrong. Luckily, just as with the  $A$  vector potential in the magnetic field, we can apply a gauge transformation to the scalar potential,  $\phi$ , of the magnetic field that will cancel out this new term, and result in field invariance.

We have:

$$E = -\nabla\phi - \frac{\partial A}{\partial t} - \frac{\partial \nabla\chi}{\partial t}$$

Which can also be written as:

$$E = -\nabla\phi - \frac{\partial A}{\partial t} - \nabla\frac{\partial\chi}{\partial t}$$

Transforming  $\phi$  into  $\phi' = \phi + \frac{\partial\chi}{\partial t}$  changes the electric field equation to:

$$E = -\nabla\phi' - \frac{\partial A}{\partial t} - \nabla\frac{\partial\chi}{\partial t}$$

$$E = -\nabla\left(\phi + \frac{\partial\chi}{\partial t}\right) - \frac{\partial A}{\partial t} - \nabla\frac{\partial\chi}{\partial t}$$

$$E = -\nabla\phi + \nabla\frac{\partial\chi}{\partial t} - \frac{\partial A}{\partial t} - \nabla\frac{\partial\chi}{\partial t} = -\nabla\phi - \frac{\partial A}{\partial t} = E$$

And our electric field is invariant.

### 3 Maxwell's Equations of Motion

The following table is Maxwells equations:

Maxwell's Equations	
Gauss's Law	$\nabla \cdot E = -\mu_0\rho$
Gauss's Law for Magnetism	$\nabla \cdot B = 0$
Maxwell-Faraday Equation	$\nabla \times E = -\frac{\partial B}{\partial t}$
Ampere's Circuit Law	$\nabla \times B = \mu_0 J + \mu_0\epsilon_0 \frac{\partial E}{\partial t}$

Taking the Divergence of our electric field equation above:

$$\nabla \cdot E = \nabla^2\phi + \frac{\partial \nabla \cdot A}{\partial t} \quad (6)$$

And substituting it into Gauss's Law gives us:

$$\nabla^2\phi + \frac{\partial \nabla \cdot A}{\partial t} = -\mu_0\rho \quad (7)$$

Substituting our equations for electric (5) and magnetic (3) field into amperes law:

$$\nabla \times (\nabla \times A) = \mu_0 J + \frac{1}{c^2} \left( \frac{\partial}{\partial t} (\nabla\phi + \frac{\partial \nabla \cdot A}{\partial t}) \right) \quad (8)$$

However, the  $\nabla \times (\nabla \times A)$  can be rewritten as:

$$\nabla \times (\nabla \times A) = \nabla(\nabla \cdot A) - \nabla \cdot (\nabla A)$$

And (8) can be rewritten as:

$$\nabla^2 A + \frac{1}{c^2} \frac{\partial^2 A}{\partial t^2} - \nabla(\nabla \cdot A + \frac{1}{c^2} \frac{\partial \phi}{\partial t}) = -\mu_0 J \quad (9)$$

Where J is the total current density,  $\mu_0$  is the permeability of free space,  $\epsilon_0$  is the permittivity of free space, and  $\rho$  is the total charge density.

#### 3.1 Lorenz Gauge Condition

These equations can be simplified if we can find a gauge transformation where:

$$\nabla \cdot A + \frac{1}{c^2} \frac{\partial \phi}{\partial t} = 0 \quad (10)$$

This is known as the Lorenz Gauge Condition.

Using the same transformation as before:

$$A' = A + \nabla\chi$$

$$\phi' = \phi - \frac{\partial \chi}{\partial t}$$

Substituting in the gauge condition gives:

$$\nabla \cdot A' + \frac{1}{c^2} \frac{\partial \phi'}{\partial t} = \nabla \cdot A + \nabla^2\chi + \frac{1}{c^2} \frac{\partial \phi}{\partial t} - \frac{1}{c^2} \frac{\partial^2 \chi}{\partial t^2} = 0 \quad (11)$$

$$\nabla^2\chi - \frac{1}{c^2} \frac{\partial^2 \chi}{\partial t^2} = \nabla \cdot A + \frac{1}{c^2} \frac{\partial \phi}{\partial t} \quad (12)$$

### 3.2 Aside

This result could also have been obtained by taking the Divergence of the A vector potential and the time derivative of the electric potential and then adding the two equations together.

$$\nabla \cdot A' = \nabla \cdot A + \nabla^2 \chi \quad (13)$$

$$\frac{\partial \phi'}{\partial t} = \frac{\partial \phi}{\partial t} - \frac{\partial^2 \chi}{\partial t^2} \quad (14)$$

$$\nabla \cdot A' + \frac{\partial \phi'}{\partial t} = \nabla \cdot A + \frac{\partial \phi}{\partial t} + \nabla^2 \chi - \frac{\partial^2 \chi}{\partial t^2} = 0 \quad (15)$$

However, we recall from the Lorenz transformation that:

$$\nabla \cdot A + \frac{1}{c^2} \frac{\partial \phi}{\partial t} = 0$$

so we require

$$\nabla^2 \chi - \frac{1}{c^2} \frac{\partial^2 \chi}{\partial t^2} = 0$$

This is a constraint on our transformation. The Lorenz condition will be satisfied so long as we choose a gauge that satisfies this requirement. Doing so leaves us with:

$$\nabla \cdot A = -\frac{1}{c^2} \frac{\partial \phi}{\partial t} \quad (16)$$

When substituting this into Gauss's Law and Amperes Law we obtain the three following equations:

$$\nabla^2 \phi - \frac{1}{c^2} \frac{\partial^2 \phi}{\partial t^2} = -u_0 \rho \quad (17)$$

$$\nabla^2 A - \frac{1}{c^2} \frac{\partial^2 A}{\partial t^2} = -u_0 J \quad (18)$$

$$\nabla^2 \chi - \frac{1}{c^2} \frac{\partial^2 \chi}{\partial t^2} = 0 \quad (19)$$

Analyzing these results we see that each equation is actually just the standard wave equation which describes a wave traveling at the speed of light. We see that changes in the electric and magnetic fields travels at the speed of light.

## 4 Gauge Theory in Quantum Mechanics

As we all know, the phase of a wave function does not visibly affect the physics of a system. As an example, we choose to ignore(absorb into constants) the phase factor when computing the transmission matrix of a wave function[4]. This means that we can arbitrarily pick a phase to start with and as long as we carry this phase through the problem, no physical effects will be observed(the particle will end up in the correct location). Here the gauge transformation

is regarded as the change in phase of a wave function, while the potential will "connect" the phase at one space-time to the phase at another location.

First let us define the transformed wave function:

$$\psi = \psi e^{-i\alpha} \quad (20)$$

For an electromagnetic field, the momentum operator becomes:

$$-i\hbar\nabla - \frac{e}{c}A \quad (21)$$

The time dependent schrodinger equation says:

$$\left[\frac{1}{2m}(-i\hbar\nabla - \frac{e}{c}A)^2 + e\phi\right]\psi e^{i\alpha} = i\hbar \frac{\partial \psi e^{i\alpha}}{\partial t} \quad (22)$$

Substituting our new  $\psi'$  into the schrodinger equation, we get:

$$\left[\frac{1}{2m}(-i\hbar\nabla - \frac{e}{c}A)^2 + e\phi\right]\psi e^{i\alpha} = i\hbar \frac{\partial \psi e^{i\alpha}}{\partial t} \quad (23)$$

Using the identity  $e^f(\frac{\partial}{\partial x})g = (\frac{\partial}{\partial x} - \frac{\partial f}{\partial x})g$  [4] we can rewrite the schrodinger equation to read:

$$\left[\frac{1}{2m}(-i\hbar\nabla - \frac{e}{c}A + i\hbar(i\nabla\alpha))^2 + e\phi\right]\psi e^{i\alpha} = (i\hbar \frac{\partial}{\partial t} - i\hbar(i\frac{\partial\alpha}{\partial t}))\psi e^{i\alpha} \quad (24)$$

We see that if  $\alpha$  is constant, then  $\nabla\alpha$  and  $\frac{\partial\alpha}{\partial t}$  will be zero and the original schrodinger equation will be restored. However, a local gauge says that  $\alpha$  can change from point to point in space-time, so  $\alpha$  may also be a function of  $x$  and  $t$ . In this case we must apply a gauge transformation to the potentials  $A$  and  $\phi$  to cancel the derivatives generated from  $\alpha$ .

$$A' = A + \frac{c\hbar}{e}\nabla\alpha \quad (25)$$

and

$$\phi' = \phi - \frac{\hbar}{e}\frac{\partial\alpha}{\partial t} \quad (26)$$

Finally we see that just as in electrodynamics, we define a gauge transformation to the potentials such that:

$$A' = A - \nabla\chi \text{ and } \phi' = \phi + \frac{1}{c}\frac{\partial\chi}{\partial t}$$

Applying the transformed wave function and potential to this equation gives the schrodinger equation back again. The terms generated by the gradient cancel off with the transformed vector potential, while the extra time derivative is cancelled off by the transformed electric potential. This exercise shows that the phase of a wave function can be regarded as a local variable that can be changed under a gauge transformation. In the end, the same physical equations are obtained, so no observable change occurs by changing the phase of the wave function. (Actually, a change is seen, however, it is the relative phase change that is observed. See Aharanov-Bohm effect)

## 5 Yang Mills Gauge Theory

The final part of this paper will discuss modern gauge theory and how it originated from a paper that was initially dismissed due to mass requirements that were not observed. We begin by making the assumption that the proton and neutron are actually the same particle, just with a different spin orientation. It is not a far stretch of the imagination to recognize this assumption. After all, the proton and neutron are very similar, their masses are nearly identical, nuclear forces between them are the same, in fact the only large difference between the two particles is charge. It was suggested (by Heisenberg) that, just as electrons have spin up and spin down, neutrons and protons are also spin up and spin down states[1]. In 1952 Yang and Mills suggested a field analogous to the electromagnetic field, controlled the strong nuclear interactions. They hypothesized that if we were to ignore the charge difference between a neutron and proton, and ignore all electromagnetic effects, and just concentrate on the nuclear forces, we would not be able to tell the difference between which particle is a proton and which is a neutron[1]. All we know is that one particle has an isotopic spin up and the other, down. We are free to assign the proton to the up state and the neutron to the down state. Now, this isotopic spin state is a local variable, meaning that it can be different for each location in space time, this is called the isotopic gauge[2]. The up state at one location does not have to be the same at any other location. We are free to choose which direction is up and which is down at any other space time point.

Once we choose, however, which state will correspond to a proton and which will correspond to neutron, we must carry this assignment throughout all space-time. How can we do this when isotopic spin is local to each point in space-time? There must be a "connection" between two space-time points that Yang and Mills attributed to the isotopic spin potential field, just as how the electromagnetic potential connects the phase of wavefunctions at different points.

The connection between two points, the isotopic spin potential, must be able to rotate the isotopic spin direction between two points in space[1]. Remember that the up direction at one location may not be oriented the same way at another location, so the potential must transform the spin as the particle travels from one point to another. The potential does not actually rotate the spin but is actually a "generator of the rotation." [1] A rotation of a wave function can be written as:

$$R\psi = \psi e^{-i\theta T}$$

which is analogous to the phase change described in section 4. We saw above that the transformed potential was proportional to the phase change  $\chi$ . In this case, in order for the equations of motion to remain invariant, the potential must be proportional to the isotopic spin momentum  $T$ .

### 5.1 Yang Mills Derivation

Just as we saw how the phase of a wave function transforms, we apply the same procedure to the isotopic spin:

$$\psi' = S(x, t)^{-1}\psi \tag{27}$$

where  $S(x,t)$  is the isotopic spin rotation at a space-time location.

Now, again, analogously to the electric potential, we define the isotopic spin potential so that we can cancel out the effects of the rotation  $S$  applied to the isospin wave function. To cancel off these extra terms generated by taking the gradient of the potential, we replace the gradient operator with the covariant derivative:

$$D = \nabla - i\epsilon A(x, t) \quad (28)$$

where  $\epsilon$  is the coupling constant which describes the field strength[2].

The potential then becomes:

$$A' = S^{-1}AS + \left(\frac{i}{\epsilon}\right)S^{-1}\nabla S \quad (29)$$

$A$  itself, as stated previously, is proportional to the isotopic spin angular momentum  $T$ :

$$A_\mu = \sum_a A_\mu^a T^a \quad (30)$$

The lagrangian can be written:

$$L = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} - \psi^T \gamma_\mu D_\mu \psi \partial^\mu \psi - m\psi^T \psi \quad (31)$$

We should also note that this is the Lagrangian for an electron in an electromagnetic field. Here,  $\frac{1}{4}F^{\mu\nu}F_{\mu\nu}$  is the energy density of the field,  $\psi^T \gamma_\mu D_\mu \psi \partial^\mu \psi$  is the kinetic energy of the electron, and  $m\psi^T \psi$  is the mass of the electron, and  $D$  is the covariant derivative[1]. In the Yang Mills case, this equation does not refer to the electron. Instead, this is the lagrangian for the Yang Mills particle, which they call the  $b_\mu$  quanta, in the B field. This b quantum is the propagator of the B field. For a Neutron to transform into a proton, one b quantum must be absorbed. A b quantum is emitted to go the other direction. Up until this point we have neglected charge differences between neutrons and protons. Now we see that the Yang Mills field must also carry the charge associated with the change between particles. Because of this, the b quantum must have three charge states associated with it, +e, -e and 0[2]. The mass of the b quantum presents a problem in the Yang Mills gauge theory. Adding a mass of the gauge field into the Lagrangian violates gauge invariance. There is no extra term that can cancel off this mass term. Therefore we require that the mass of the b quantum be zero. This does not sound far fetched, after all, the photon also has zero mass. However, upon closer examination, we see that a massless particle would allow the nuclear force to have large effects at much longer distances than are observed, similar to those of the electromagnetic force. In addition, a massless particle is expected to already have been observed experimentally, whereas no such particle has been found. This one restriction, that the mass be zero, put great doubt on the validity of the Yang Mills theory. In fact it wasn't until the discovery of symmetry breaking that the Yang Mills theory was rediscovered and evolved into what is now modern gauge theory.



## 6 Conclusion

We have addressed the application of gauge transformations on the electromagnetic field. In doing this we have seen how the transformation must keep the lagrangian invariant and therefore produce the same equations of motion. Addressing this same problem through quantum mechanics, we derived the same transformations of the electric and magnetic potentials using on the conditions that the phase of a wave function is an extra degree of freedom for us to manipulate, and that the schrodinger equation remain invariant. Following this, we discussed the theory behind the Yang Mills field, and discussed why an additional massless particle created problems for the acceptance of this new Gauge theory.

## 7 Resources

- [1] K. Moriyasu, *An Elementay Primer For Gauge Theory*, World Scientific Publishing Co Pte Ltd, Singapore, 1983
- [2] C. N. Yang and R. L. Mills, *Conservation of Isotopic Spin and Isotopic Gauge Invariance*, Phys. Rev, Vol 96, 1954
- [3] R. Healey, *Gauging What's Real*, Oxford University Press, 2007
- [4] R. Gilmore *Class Notes*