A brief introduction to gauges.

Rocco Tenaglia

March 2, 2018

Introduction.

Reference to gauge freedom in electromagnetism can be found as early as Maxwell. In deriving his equations Maxwell noted that there was some freedom of choice in the magnetic vector potential. Choosing to keep things simple, he used what became known as the Coulomb gauge, $\nabla \cdot \mathbf{A} = 0$. Later on, Lorenz used his eponymous gauge, $\nabla \cdot \mathbf{A} = -\mu_0 \varepsilon_0 \frac{\partial V}{\partial t}$, to produce a general solution to Maxwell's equations (the derivation of which we will reproduce below) (Jackson, 2001).[1] Amidst discussions among the community of which of these potentials was "correct", Fock (1926; see translation in Fock, 2010) wrote down equations for the transformation from one gauge to another, as well as a similar equation which transformed the wave equation, thus resolving the question. All this was completed before the name of gauge was ever uttered.

Beginning in 1928 with his book *Gruppentheorie und Quantenmechanik*, Hermann Weyl "enshrined as fundamental the principle of gauge invariance", according to Jackson, coining the term in the process. On top of repeating Fock's result, Weyl first suggested the interpretation that these equations governed the relation between electromagnetism and matter. Many others took this idea up, starting with Klein in 1938, who attempted to describe the electromagnetic and weak forces by way of gauge symmetry. He failed in his goal, but in 1954, Yang and Mills released a paper which introduced an important theoretical result on non-Abelian gauge symmetries, the Yang-Mills Equation. This paved the way for the creation of the current electroweak theory by Glashow, Salam, and Weinberg in the 1960s and the development of a distinct gauge theory describing the strong force by Gell-Mann in the 1970s; these two theories form the description of all non-gravitational forces in the Standard Model (Jackson).

In this paper we focus on the electromagnetic aspects of gauge theory. We begin by showing as an example how Maxwell's equations can be solved in general form using the Lorenz gauge. We will explore the problem from the perspective of two different gauges to illustrate how difficult it may become. We then discuss gauge theory as a whole, and derive an important result to which gauge invariance is fundamental.

Maxwell's and Schrödinger's equations.

The first question one might ask about gauge theory is "what is a gauge?" There are two different ideas of gauges present in theoretical physics as a whole, making the term itself somewhat difficult to pin down. However, an intuitive understanding of gauges can be obtained by using electromagnetism as an example. Here, we are already familiar with idea of gauges, in the form of the magnetic vector potential $\mathbf{A}$:

$$\mathbf{A}' = \mathbf{A} + \nabla\chi, \tag{1}$$

where $\mathbf{A}$ and $\mathbf{A}'$ are both perfectly valid vector potentials. The scalar function χ is known as a *gauge transformation*, which is used to transform $\mathbf{A}$ into $\mathbf{A}$'. Note that we are adding the gradient of χ; since the curl of a gradient is always zero, the electromagneetic fields remain unchanged. The fact that we have the freedom do this is known as *gauge invariance*, and this symmetry plays a vital role in modern theoretical physics.

[1] This was eventually attributed to H. A. Lorentz, although Lorenz (and Riemann as well) had proved it long before his time (Jackson, 2001).

As it turns out, the same χ can be used to transform both the electric scalar potential V and the wave equation ψ. We list the corresponding transformations, due to Fock (1926) and rewritten in modern terms by Jackson (2001), here:[2]

$$V' = V - \mu_o \varepsilon_o \frac{\partial \chi}{\partial t}, \tag{2}$$

$$\psi' = \psi e^{ie\chi/\hbar c}. \tag{3}$$

Note that each of these transformations keep some value invariant. (1) maintains the magnetic field, since the curl of any gradient is zero. (2) maintains the electric field, as we will show below. (3) transforms the quantum mechanical wave function; it maintains the resulting probability:

$$|\psi'|^2 = |e^{ie\chi/\hbar c}|^2 |\psi|^2 = |\psi|^2. \tag{4}$$

This non-uniqueness of gauges can be useful for practical reasons, by way of *gauge fixing* - that is, choosing a function χ to perform calculations in. A powerful application of this is the derivation of the general solution to Maxwell's equations. To do this, we will use the Lorenz gauge, defined in SI units by $\nabla \cdot \mathbf{A} + \mu_o \varepsilon_o \frac{\partial V}{\partial t} = 0$. This derivation is due to Griffiths (1999).

We begin by reducing Maxwell's equations to two equations involving only V and $\mathbf{A}$. We first note that there is currently no scalar potential, since the curl of $\mathbf{E}$ is nonzero. We correct this using Faraday's Law and the definition of $\mathbf{A}$:

$$\nabla \times \mathbf{E} = -\frac{\partial}{\partial t}(\nabla \times \mathbf{A}),$$

$$\implies \nabla \times (\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t}) = 0. \tag{5}$$

The quantity in (5) does have a vanishing curl. We thus define V from that equation:

$$\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t} = -\nabla V. \tag{6}$$

(Note here that the gauge transformation (2) is satisfied by this definition; the gradient of the time derivative obtained by plugging in (2) cancels out the time derivative of the gradient obtained by plugging in (1).) Now we have

$$\nabla^2 V + \frac{\partial}{\partial t}(\nabla \cdot \mathbf{A}) = -\frac{1}{\varepsilon_o}\rho, \tag{7}$$

and finally,

$$\nabla \times (\nabla \times \mathbf{A}) = \mu_o \mathbf{J} - \mu_o \varepsilon_o \frac{\partial}{\partial t}\left(\nabla V + \mu_o \varepsilon_o \frac{\partial \mathbf{A}}{\partial t}\right)$$

$$\nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A} = \mu_o \mathbf{J} - \mu_o \varepsilon_o \nabla \frac{\partial V}{\partial t} - \mu_o \varepsilon_o \frac{\partial^2 \mathbf{A}}{\partial t^2}$$

$$\left(\nabla^2 \mathbf{A} - \mu_o \varepsilon_o \frac{\partial^2 \mathbf{A}}{\partial t^2}\right) - \nabla\left(\nabla \cdot \mathbf{A} + \mu_o \varepsilon_o \frac{\partial V}{\partial t}\right) = -\mu_o \mathbf{J}. \tag{8}$$

(6) is the potential form of both the curl of $\mathbf{E}$ and the divergence of $\mathbf{B}$. We used (6) and Gauss' Law to create (7) and we used (6) and Ampere's Law to create (8). Thus, equations (7) and (8) are Maxwell's Equations in potential form.

Here now we apply the Lorenz gauge to simplify Maxwell's equations to

$$\nabla^2 V - \mu_o \varepsilon_o \frac{\partial^2 \mathbf{V}}{\partial t^2} = -\frac{1}{\varepsilon_o}\rho, \tag{9}$$

$$\nabla^2 \mathbf{A} - \mu_o \varepsilon_o \frac{\partial^2 \mathbf{A}}{\partial t^2} = -\mu_o \mathbf{J}. \tag{10}$$

Note that (9) relies only on V, while (10) relies only on $\mathbf{A}$. These noticeably simpler (and more symmetric) equations are known as *wave equations*, and they are well-known differential equations which can be solved

[2] As is the case throughout this paper, Jackson's notation differs slightly from the more familiar Griffiths' notation. Whenever necessary, I will translate his notation to Griffiths', which tends to be less obtuse.

relatively easily. From here, the electric and magnetic fields can be found by taking a gradient and a curl, respectively. Note that this question would have been significantly more difficult in the Coulomb gauge, since we would have to solve (7) and (8) in this form:

$$\nabla^2 V = -\frac{1}{\varepsilon_o}\rho \tag{11}$$

$$\left(\nabla^2 \mathbf{A} - \mu_o\varepsilon_o\frac{\partial^2 \mathbf{A}}{\partial t^2}\right) - \mu_o\varepsilon_o\nabla\frac{\partial V}{\partial t} = -\mu_o\mathbf{J}. \tag{12}$$

(11) is actually simpler than (9), being Poisson's equation. (12), however, is quite complicated. It still involves V, and it has three terms to boot (the first term did not cancel out because it does not actually involve taking the divergence of $\mathbf{A}$). Rather than attempt to solve it, we can now a different method - taking a gauge transformation from the Lorenz gauge to the Coulomb gauge and then calculating the electromagnetic fields of each. We refer to Jackson (2002) for the solutions here:[3]

$$V_L(\mathbf{r},t) = \frac{1}{4\pi\varepsilon_o}\int\frac{\rho(\mathbf{r}',t_r)}{\mathscr{r}}d\tau', \tag{13}$$

$$\mathbf{A}_L(\mathbf{r},t) = \frac{\mu_o}{4\pi}\int\frac{\mathbf{J}(\mathbf{r}',t_r)}{\mathscr{r}}d\tau', \tag{14}$$

and for the gauge transformation,

$$\chi = \frac{1}{\mu_o\varepsilon_o}\int\frac{1}{\mathscr{r}}\int_t^{t_r}\rho(\mathbf{r}',t')dt'd\tau' + \chi_o, \tag{15}$$

where t_r is the so-called *retarded time*, $t_r = t - \mathscr{r}/c$. This is necessary because, as Griffiths puts it, "electromagnetic 'news' travels at the speed of light" (1999), and no faster. This causes a lag in the electromagnetic effects of a moving charge distribution at any point in space. In the transformation (15), the limits of the time integral are a consequence of the scalar potential condition for the gauge transformation:

$$\mu_o\varepsilon_o\frac{\partial\chi}{\partial t} = V_L(\mathbf{r},t) - V_C(\mathbf{r},t). \tag{16}$$

The Lorenz scalar potential depends on t_r, but the Coulomb potential (which is the general solution to Poisson's equation and can be found easily without a gauge transformation) does not. At any time t, then, the difference is given by the limits of integration. We now write down Jackson's solution for the Coulomb gauge vector potential:

$$\mathbf{A}_C = -\frac{1}{\mu_o\varepsilon_o}\nabla\times\int\int_t^{t_r}(t-t')\frac{\hat{\boldsymbol{\mathscr{r}}}\times\mathbf{J}(\mathbf{r}',t')}{\mathscr{r}^2}dt'd\tau'. \tag{17}$$

As implied by the significantly more difficult differential equation (12), the Coulomb vector potential is much more complicated than the Lorenz potential. The introduction of a different gauge vastly simplified both our problem and its solution.[4]

General gauge theory.

Electromagnetism is the quintessential example of a gauge theory, by virtue of being both the first and the inspiration for all other gauge theories. Since its inception, it has done much to incite other physicists to further explore the symmetry inherent in gauges. Yang-Mills theory, around which the Standard Model was based, was designed in great part by copying Maxwell's equations. Indeed, the Yang-Mills equation as stated in Baez (1994) is nearly identical to the single-equation statement of Maxwell's equations in the same text:

$$dF = 0,$$

[3]Jackson does not actually provide a derivation for the general solutions, choosing instead to state the solution and then work with it to achieve the rest of the results given here. A derivation of this can be found instead in Griffiths (1999).

[4]We can now show that these solutions do indeed satisfy their respective equations and then calculate the electric and magnetic fields to show that they are equal as expected. Griffiths (1999) and Jackson (2002) respectively provide these derivations.

$$\star d \star F = J; \tag{18}$$
$$d_D F = 0,$$
$$\star d_D \star F = J. \tag{19}$$

Equations (18) are Maxwell's equations; the statement here is equivalent to another pair of equations formed directly from the usual Maxwell's equations by a substitution (Baez, 1994):

$$\nabla \cdot \mathcal{E} = \rho,$$
$$\nabla \times \mathcal{E} = i\left(\frac{\partial \mathcal{E}}{\partial t} + \mathbf{J}\right), \tag{20}$$

where $\mathcal{E} = \mathbf{E} + i\mathbf{B}$. (Note that the form (20) implies the existence of magnetic charge and current, although none has been shown to exist.) Although the form (18) and the so-called Bianchi identity and Yang-Mills equation (19) are both beyond the scope of this paper, the direct parallel of notation is a clear indication of the fact that Yang-Mills theory is merely a generalization of Maxwell's equations.[5]

Yang-Mills theory as a whole is too big of a subject for the scope of this paper; we shall attempt to provide an introduction by way of group theory. In algebra, a group (often denoted by G) is a set with a binary operation $\cdot$ (i.e., one which takes in two inputs), such that the following three properties hold:

1. For any three elements f, g, and h in G, we have $f \cdot (g \cdot h) = (f \cdot g) \cdot h$;
2. There is an element e in G such that for every g in G, $e \cdot g = g$;
3. For any element g in G, there is an element g^{-1} such that $g \cdot g^{-1} = e$.

The integers coupled with the binary operation of addition are an example of a group. In particular, however, the groups we need to discuss are known as Lie groups. These groups have the added property that they are a *manifold*, i.e., a space which locally looks like Euclidean space ($\mathbf{R}^n$). To describe this mathematically, we say that for every element g in G, there is an open set[6] containing it such that we can map that set continuously to Euclidean space (and back).

We will not go into the specifics of Lie group theory here; rather, we will introduce the concept of a *representation*. A representation ρ of a group G is a function whose range is the set of invertible linear transformations on a vector space V, a group in its own right. A representation must also distribute across the group operation - that is, for any g and h in G, $\rho(g \cdot h) = \rho(g)\rho(h)$. Gauge theory is built upon these concepts - the gauge invariance in any physical force can be described as the representation of a Lie group on a manifold such as spacetime. For electromagnetism, the Lie group in question is $U(1)$, the set of complex numbers on the unit circle in the complex plane with multiplication as its operation. This is a Lie group. To see this, note that the function $f(\theta) = e^{i\theta}$ is continuous and invertible, with its inverse mapping the complex unit circle to the real interval $[0, 2\pi)$.[7] The Lie group for the weak force is known as $SU(2)$, and it follows that the theory of the electroweak force is described by $SU(2) \times U(1)$. The strong force Lie group is $SU(3)$; accordingly, all the forces in the Standard Model are described collectively by the cross products of these groups, $SU(3) \times SU(2) \times U(1)$.[8]

We now have almost all the tools we need to properly discuss an important derivation given by Griffiths' *Introduction to Elementary Particles* (1987). The last tool we need is known as the Dirac Lagrangian, called a Lagrangian because it is derived in much the same way as the Lagrangian from classical mechanics, using the differential equation

$$\frac{\partial \mathscr{L}}{\partial \overline{\psi}} = i\hbar c \sum_{\mu=0}^{1} (\gamma_\mu \partial_\mu)\psi - mc^2\psi, \tag{21}$$

[5]Note that there is a significant amount of differential geometry hidden in the equations (19). To accurately define the Bianchi identity and the Yang-Mills equation, we need an understanding of vector bundles and connections, whereas (18) follows after only an introduction to differential forms and manifolds. Even the star operators and the "current density" J are not the same between the two pairs of equations. The symmetry here is suggestive, but can be misleading.

[6]Under the metric defined on the group.

[7]The mapping here is a little more subtle than we have described, since the complex unit circle is a loop. This actually causes our function to be discontinuous at $\theta = 0$. To account for this, we instead define *several* functions, each with a slightly different domain. This gives us the open sets described in our definition of manifolds.

[8]Although gravity is not included in the Standard Model or described by Yang-Mills theory in general, it is still governed by a form of gauge theory. The tenets of general relativity require that the laws of physics are the same in any reference frame - that is, physics is invariant under *coordinate transformations of spacetime* (Wald, 1984). The gauge transformations, mathematically, are diffeomorphisms - bicontinuous, bijective, bidifferentiable functions from spacetime to spacetime - which cannot be collected in a Lie group the way that the other gauge transformations can.

the four values of μ referring to the four coordinates of space and time.[9] The immediate solution to this equation is

$$\mathscr{L} = i\hbar c\overline{\psi}\sum_{\mu=0}^{3}(\gamma_\mu\partial_\mu)\psi - mc^2\overline{\psi}\psi. \tag{22}$$

Lagrangians are central to constructing gauge theory - the Yang-Mills equations are derived from a Lagrangian. This one in particular is used in relativistic field theory to describe what is known as a "spin-1/2 field", essentially a wave function for a spin-1/2 particle (Griffiths, 1987). We now use the Dirac Lagrangian to derive an important connection between particles and electromagnetism.

We begin with (3). Note that the transformation factor $e^{ie\chi/\hbar c}$ in this equation is in fact an element of $U(1)$ for any transformation χ; thus, equation (3) is in fact a representation of $U(1)$ on the wave function. Also recall from (4) that invariance clearly holds for the probability distribution of ψ, and that if χ is a constant, then so is the Dirac Lagrangian. If it is *not* constant, however, then another term is added:

$$\mathscr{L} = i\hbar c\overline{\psi}\sum_{\mu=0}^{3}(\gamma_\mu\frac{ie}{\hbar c}(\partial_\mu\theta)\psi + e^{ie\chi/\hbar c}\partial_\mu\psi) - mc^2\overline{\psi}\psi. \tag{23}$$

The new term is due to the derivative of χ, and it means that the Lagrangian is no longer invariant. To correct the issue, we must add a correcting term to the Lagrangian. This term comes from the vector potential $\mathbf{A}$:

$$\mathscr{L} = i\hbar c\overline{\psi}\sum_{\mu=0}^{3}(\gamma_\mu\partial_\mu)\psi - mc^2\overline{\psi}\psi - \overline{\psi}\sum_{\mu=0}^{3}(\gamma_\mu\psi A_\mu), \tag{24}$$

where A_μ refers to the components of $\mathbf{A}$. Now we see that the gauge transformation in the added term cancels itself out: By (1), we are subtracting a new set of terms $\overline{\psi}\gamma_\mu\psi\partial_\mu\chi$, which are equal to and cancel out the miscreant first term in the summand from (22). Thus, the new term makes the Dirac Lagrangian invariant under gauge transformations.

The most important part of this result is the implied connection between particles and the electromagnetic fields. In the words of Hermann Weyl, "...I now believe that this gauge invariance does not tie together electricity and gravitation, but rather *electricity and matter*...." (Quoted from Jackson, 2001) Here we see that this is in some respect the case - in order to have our Lagrangian be invariant under gauge transformations, it is necessary to intrinsically tie together the field of a particle and the electromagnetic field, suggesting the claim made by Weyl above. In fact, Griffiths (1987) suggests that (24) implies that "the condition of local phase invariance... generates all of electrodynamics and specifies the current produced by Dirac particles." The vital connection of gauge theory to this result only serves to drive home the intrinsic importance of gauge theory and its connection to electromagnetism, not just as a computational tool, but also as a fundamental aspect of physics.

[9] The γ here is known as a Dirac matrix, and it comes from the Dirac equation. Griffiths (1987) provides an explanation of this.

References.

1. Baez, J. C., & Muniain, J. P. (1994). Gauge fields, knots, and gravity. Singapore: World Scientific.
2. Fock, V. (2010). On the invariant form of the wave and motion equations for a charged point mass. Physics-Uspekhi, 53. Retrieved from http://iopscience.iop.org.proxy2.cl.msu.edu/article/10.3367/UFNe.0180.201008h.0874/meta.
3. Fock, V. (1926). Über die invariante Form der Wellen- und der Bewegungsgleichungen für einen geladenen Massenpunkt. Zeitschrift für Physik, 39, 226-232. Retrieved from https://link-springer-com.proxy2.cl.msu.edu/article/10.1007%2FBF01321989.
4. Griffiths, D. J. (1999). Introduction to electrodynamics. Upper Saddle River, N.J: Prentice Hall.
5. Griffiths, D. J. (1987). Introduction to elementary particles. New York: Harper & Row.
6. Jackson, J. D. (2002). From Lorenz to Coulomb and other explicit gauge transformations. American Journal of Physics, 70, 917. Retrieved from http://aapt.scitation.org.proxy1.cl.msu.edu/doi/10.1119/1.1491265
7. Jackson, J. D.; Okun, L.B. (2001) Historical roots of gauge invariance. Reviews of Modern Physics, 73, 663. Retrieved from https://journals-aps-org.proxy1.cl.msu.edu/rmp/abstract/10.1103/ RevModPhys.73.663
8. Wald, R. M. (1984). General relativity. Chicago: University of Chicago Press.