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 finally coined the term whilst he, according to Jackson, "enshrined as fundamental the principle of gauge invariance". 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. In 1954, a paper by Yang and Mills introduced an important theoretical result on non-Abelian gauge symmetries, the Yang-Mills Equation. This paved the way for the creation of the 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 provide an example of gauge invariance by discussing the transformation from the Lorenz gauge to the Coulomb gauge and deriving the electromagnetic fields from both separately. Then as a final remark, we discuss the presence of gauge theory in theoretical physics as a whole.

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 rest of Maxwell's equations remain unchanged. The fact that we have the freedom do this is known as *gauge invariance*, and it plays a vital role in understanding symmetry, a central phenomenon in theoretical physics.

[1] This was eventually attributed to H. A. Lorentz, although Lorenz (and Riemann as well) had proved this 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]

$$\begin{aligned} V' &= V - \mu_o \varepsilon_o \frac{\partial \chi}{\partial t}, \\ \psi' &= \psi e^{i e \chi / \hbar c}. \end{aligned} \tag{2}$$

This non-uniqueness can be useful for many practical reasons, by way of *gauge fixing*. In short, we choose a function χ to work with. By (1) and (2), the choice of function is irrelevant; the electric and magnetic fields derived under any particular gauge will be the same for *all* gauges. As a direct application of this point, we now derive 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 $\mathbf{A}$:

$$\begin{aligned} \nabla \times \mathbf{E} &= -\frac{\partial}{\partial t}(\nabla \times \mathbf{A}), \\ \implies \nabla \times (\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t}) &= 0. \end{aligned} \tag{3}$$

The quantity in (3) does have a vanishing curl. We thus define V based on that:

$$\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t} = -\nabla V. \tag{4}$$

Now, using (4) and Gauss' Law,

$$\nabla^2 V + \frac{\partial}{\partial t}(\nabla \cdot \mathbf{A}) = -\frac{1}{\varepsilon_o}\rho. \tag{5}$$

Finally, using (4), the definition of $\mathbf{A}$, Ampere's Law, and a vector identity:

$$\begin{aligned} \nabla \times (\nabla \times \mathbf{A}) &= \mu_o \mathbf{J} - \mu_o \varepsilon_o \frac{\partial}{\partial t}\left(\nabla V + \mu_o \varepsilon_o \frac{\partial \mathbf{A}}{\partial t}\right) \\ \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A} &= \mu_o \mathbf{J} - \mu_o \varepsilon_o \nabla \frac{\partial V}{\partial t} - \mu_o \varepsilon_o \frac{\partial^2 \mathbf{A}}{\partial t^2} \\ \left(\nabla^2 \mathbf{A} - \mu_o \varepsilon_o \frac{\partial^2 \mathbf{A}}{\partial t^2}\right) - \nabla\left(\nabla \cdot \mathbf{A} + \mu_o \varepsilon_o \frac{\partial V}{\partial t}\right) &= -\mu_o \mathbf{J}. \end{aligned} \tag{6}$$

(4) is the vector form of both the curl of $\mathbf{E}$ and the divergence of $\mathbf{B}$. We used (4) and Gauss' Law to create (5) and we used (4) and Ampere's Law to create (6). Thus, equations (5) and (6) are Maxwell's Equations in potential form.

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

$$\nabla^2 V - \mu_o \varepsilon_o \frac{\partial^2 \mathbf{V}}{\partial t^2} = -\frac{1}{\varepsilon_o}\rho, \tag{7}$$

$$\nabla^2 \mathbf{A} - \mu_o \varepsilon_o \frac{\partial^2 \mathbf{A}}{\partial t^2} = -\mu_o \mathbf{J}. \tag{8}$$

Note that (7) relies only on V, while (8) relies only on $\mathbf{A}$. These noticeably simpler (and more symmetric) equations are known as *wave equations*, and each can be solved to obtain general forms for the potentials in the Lorenz gauge. From here, the electric and magnetic fields can be found by taking a gradient and a curl, respectively.

The most intriguing claim we have made is, of course, that the electric and magnetic fields found using this method are independent of gauge choice. To verify this claim, our preference would be to compare solutions to Maxwell's equations in two separate gauges. Unfortunately, solutions to Maxwell's equations are difficult to come by. Observe the form of (5) and (6) in the case of the Coulomb gauge, $\nabla \cdot \mathbf{A} = 0$:

[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.

$$\nabla^2 V = -\frac{1}{\varepsilon_o}\rho \tag{9}$$

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

(9) is exceedingly simple; it is exactly Poisson's equation. (10), however, is incredibly 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 $\mathbf{A}$). Finding a solution to this equation is significantly more complicated than finding solutions in the Lorenz gauge. Rather than attempt to do so, we will use a different method - taking a gauge transformation from the Lorenz gauge to the Coulomb gauge and then calculating the magnetic field of each. We refer to Jackson (2002) for the calculations here, as they are too complicated for the scope of this paper. First, the solution to Maxwell's equations in the Lorenz gauge is as follows:[3]

$$V_L(\mathbf{r},t) = \frac{1}{4\pi\varepsilon_o}\int\frac{\rho(\mathbf{r}',t_r)}{\mathscr{r}}d\tau', \tag{11}$$

$$\mathbf{A}_L(\mathbf{r},t) = \frac{\mu_o}{4\pi}\int\frac{\mathbf{J}(\mathbf{r}',t_r)}{\mathscr{r}}d\tau', \tag{12}$$

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.

For completeness, we now write down the explicit gauge transformation χ which Jackson uses to move from the Lorenz gauge to the Coulomb gauge:

$$\chi = \frac{1}{\mu_o\varepsilon_o}\int\frac{1}{\mathscr{r}}\int_t^{t_r}\rho(\mathbf{r}',t')dt'd\tau' + \chi_o. \tag{13}$$

Here the limits of the time integral are a consequence of the scalar potential condition for the gauge transformation:

$$\mu_o\varepsilon_o\frac{\partial\chi}{\partial t} = V_L(\mathbf{r},t) - V_C(\mathbf{r},t). \tag{14}$$

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 vector potential:

$$\mathbf{A}_C = -\frac{1}{\mu_o\varepsilon_o}\nabla\times\int\int_t^{t_r}(t-t')\frac{\hat{\boldsymbol{\mathscr{r}}}\times\mathbf{J}(\mathbf{r}',t')}{\mathscr{r}^2}dt'd\tau'. \tag{15}$$

As implied by the significantly more difficult differential equation (10), the Coulomb vector potential is a significantly uglier solution.

The above derivation is just one application of gauges. Several others can be found in Jackson (2002).

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. The Yang-Mills theory in particular 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 book:

$$\begin{gathered} dF = 0, \\ \star d\star F = J; \end{gathered} \tag{16}$$

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

$$
\begin{aligned}
d_D F &= 0, \\
\star d_D \star F &= J.
\end{aligned} \tag{17}
$$

Equations (16) 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):

$$
\begin{aligned}
\nabla \cdot \mathcal{E} &= \rho, \\
\nabla \times \mathcal{E} &= i\left(\frac{\partial \mathcal{E}}{\partial t} + \mathbf{J}\right),
\end{aligned} \tag{18}
$$

where $\mathcal{E} = \mathbf{E} + i\mathbf{B}$. (Note that the form (18) implies the existence of magnetic charge and current, although none has been shown to exist.) Although the form (16) and the Bianchi identity and Yang-Mills equation (17) are both beyond the scope of this paper, the direct parallel of notation is a clear identification of the fact that the Yang-Mills theory is merely a generalization of Maxwell's equations.[4]

The Yang-Mills theory is the basis for the entire Standard Model. Clearly electromagnetism is a gauge theory under the Yang-Mills theory, but it is also true that both the weak force and the strong force are as such. The gauge invariance in each force is described by a structure known as a Lie group, whose action on a manifold (such as 3-dimensional space) characterizes the way that the potential for each theory can be transformed while maintaining the equations for the fields. For electromagnetism, the Lie group involved is known as $U(1)$, and it can be simply stated as every value of the form $e^{i\theta}$, for θ a real number.[5] 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 $SU(3)$; thus, the entire Standard Model thus far is described by the group $SU(3) \times SU(2) \times U(1)$.

Finally, we add that general relativity as well is described by a gauge theory. In this case we do not use a Yang-Mills theory, although it still takes Maxwell's equations as inspiration. The defining feature of a gauge theory is the symmetry it describes. With the Yang-Mills theories, the symmetries all exist in the Lie groups, with each element of the group acting in such a way that the fields - the physically realizable interaction of each theory - remain invariant. In general relativity, the symmetry is somewhat different. Rather than an invariance of a field under transformations of a potential, the general relativistic invariance is under *coordinate transformations of spacetime* (Wald, 1984). This is a vastly different situation than the Standard Model, where the set of gauge transformations was "small" in a sense (compact). Here we have that the fundamental laws of physics in spacetime must be invariant under "diffeomorphisms" - bicontinuous, bijective, bidifferentiable functions from spacetime to spacetime. Despite this, the fundamental principle of a gauge theory - that the system is symmetric under some transformation - holds.

[4]Note that there is a significant amount of differential geometry hidden in the equations (17). To accurately define the Bianchi identity and the Yang-Mills equation, we need the language of vector bundles and connections, whereas (16) 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 dangerously misleading.

[5]The "action" described is more complicated than simple multiplication, and we will not discuss it here. A good reference can be found in Baez. Suffice it to say that it is equivalent to the gauge transformations (1) and (2).

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. 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
6. 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
7. Wald, R. M. (1984). General relativity. Chicago: University of Chicago Press.