Gauge Transformations in Electrodynamics

Fangjun Zhu

January 18, 2021

Let's recall Maxwell's equations in Lorentz-Heavyside units:

$$\nabla \cdot E = \rho, \quad \nabla \times E = -c^{-1}\partial_t B,$$
$$\nabla \cdot B = 0, \quad \nabla \times B = c^{-1}J + c^{-1}\partial_t E$$

where the charge density ρ and the current density J satisfy:

$$\partial_t \rho + \nabla \cdot J = 0.$$

One can view these as an IVP for a pair of vector fields E, B on space via:

$$c^{-1}\partial_t \begin{pmatrix} E \\ B \end{pmatrix} = \begin{pmatrix} 0 & \nabla\times \\ -\nabla\times & 0 \end{pmatrix} \begin{pmatrix} E \\ B \end{pmatrix} + \begin{pmatrix} -c^{-1}J \\ 0 \end{pmatrix}$$

with constraints:

$$\nabla \cdot E = \rho$$
$$\nabla \cdot B = 0.$$

We can view this as an ODE for a function valued in L^2 functions on space and so solutions do indeed exist given sufficiently nice initial data (E_0, B_0) on space. In this sense, Maxwell's equations do indeed describe the (unique) time evolution of an electromagnetic field.

If we wish to obtain these equations from variational formulae, we take the second curl of both E and B and denote $\Box := c^{-2}\partial_t^2 - \Delta$ to obtain:

$$\Box E = -c^{-2}\partial_t J - \nabla \rho$$
$$\Box B = c^{-1}\nabla \times J$$

which are a pair of inhomogeneous wave equations. These aren't quite equivalent to Maxwell's equations above, but this problem is essentially handled by our constraints.

However, there are several reasons why one might want to instead describe the state of an electromagnetic field as an equivalence class of potentials:

1. The potentials (ϕ, A) used the this alternate formulation have a total of 4 components, as opposed to the redundant 6 components of (E, B).

2. The potentials (ϕ, A) may have some physical significance (see: Aharonov-Bohm effect), although some object to this (give reference).

3. The action functional used to obtain the variational formulation of Maxwell's equations depends on the derivatives of the fields E, B and is not a terribly physical quantity. On the other hand, the version with (ϕ, A) admits a much more physical variational formulation.

So, let's briefly recall the formulation in terms of potentials. Since $\nabla \cdot B = 0$ we can write $B = \nabla \times A$ for some vector field A (at least locally) and so

$$\nabla \times (E + c^{-1}\partial_t A) = 0$$

meaning that there is some function ϕ (again, locally) such that

$$-\nabla\phi = E + c^{-1}\partial_t A.$$

We can then rewrite Maxwell's equations in terms of the potentials (ϕ, A) as:

$$\begin{gathered} -\Delta\phi - c^{-1}\partial_t \nabla \cdot A = \rho \\ \Box A + \nabla(\nabla \cdot A + c^{-1}\partial_t\phi) = c^{-1}J. \end{gathered}$$

Using $E = -\nabla\phi - c^{-1}\partial_t A$ and $B = \nabla \times A$ we can immediately see that these equations arise as variational formulae for the action:

$$S[\phi, A] = \frac{1}{2}\int d^4x \, (|E|^2 + |B|^2).$$

However, there is a major problem here:

these equations do **not** describe a well-posed IVP for (ϕ, A)!

Indeed, for any function $\chi = \chi(x, t)$ we can consider the following action on the space of fields:

$$(\phi, A) \mapsto (\phi - c^{-1}\partial_t\chi, \ A + \nabla\chi).$$

Notice that the observables E, B are invariant under this action and so if (ϕ, A) is some solution to Maxwell's equations corresponding to any choice of initial data at $t = 0$ then any choice of χ with support away from $t = 0$ will yield a new solution $(\phi - c^{-1}\partial_t\chi, \ A + \nabla\chi)$ with the same initial data!

Aside: It should be remarked that while in electrodynamics the fields E, B are themselves invariant under gauge transformations, this is no longer the case in electroweak theory or chromodynamics. In these cases the field strength will transform under the adjoint representation of a non-abelian Lie group and only its magnitude will remain gauge invariant.

Proposition 0.1. Utiyama's Theorem
Any gauge invariant local functional of (ϕ, A) is actually a functional of E, B.

In this way we see that it is reasonable to identify the state (E, B) of the electromagnetic field with the gauge equivalence class of its potential.

Nevertheless, there is one important subtlety here. We need not identify everything in a gauge orbit in order to obtain a well-posed IVP. For example, given any gauge $(\widetilde{\phi}, \widetilde{A})$ consider solving the following ODE for χ:

$$\phi - c^{-1}\partial_t \chi = 0.$$

This always has a solution by Picard-Lindelöf and so, given E, B, there always exists a choice of gauge (ϕ, A) in which $\phi = 0$ identically. This is called **temporal gauge** and we can notice that here we have

$$\begin{aligned} E &= -c^{-1}\partial_t A \\ B &= \nabla \times A. \end{aligned}$$

Furthermore, Maxwell's equations become:

$$\begin{aligned} -c^{-1}\partial_t \nabla \cdot A &= \rho \\ \Box A + \nabla(\nabla \cdot A) &= c^{-1} J. \end{aligned}$$

Notice that by assuming $\phi = 0$ we get a subset of the original gauge orbit which is preserved under gauge transformations χ satisfying $\partial_t \chi = 0$. In particular, our previous trick involving a time-dependent gauge transformation $\chi(x, t)$ no longer works! One can actually demonstrate that in temporal gauge we obtain a well-posed IVP for A, invariant under time-independent gauge transformations.

There are several other choices of gauge (such as the Lorenz and Coulomb gauges) which, when combined with temporal gauge, actually specify a single element in each gauge orbit. Since these will be discussed by others later, we will continue focusing on temporal gauge. The equations of motion in temporal gauge arise as Euler-Lagrange equations for the action:

$$S[A] := \frac{1}{2} \int d^4x \, (|\nabla \times A|^2 + c^{-2}|\partial_t A|^2)$$

with the imposed constraint:

$$-c^{-1}\partial_t \nabla \cdot A = \rho.$$