Dirac Quantisation

Chris Elliott

October 19th, 2012

In this talk I'll aim to demystify the following statement:

"The classical theory of electromagnetism on a 4-manifold M is a gauge theory whose fields are connections on principal $U(1)$-bundles on M."

How does one go from the E and B fields one learns about in school to this rather baroque geometric construction, and what does it mean for the physics? The answer to this question involves a rather beautiful story, whose most critical component is an argument given by Dirac in his 1932 paper "Quantised Singularities in the Electromagnetic Field", following geometric ideas of Weyl.

1 Electromagnetism

Let's start off by reviewing the theory of electric and magnetic fields, and their unification in the 19th century. We'll work through several approaches for mathematically describing these fields, and their behaviour. The most traditional is the following.

Definition 1.1 (First Approximation)**.** The *electric field* $\boldsymbol{E}$ and the *magnetic field* $\boldsymbol{B}$ are vector fields on $\mathbb{R}^3$, varying with time.

We can investigate these fields by asking how they act on a charged particle: the 'Lorentz force' experienced by the particle. A particle of charge q moving along a path $\gamma(t)$ experiences a force

$$\boldsymbol{F}(t) = q\left(\boldsymbol{E}_{\gamma(t)}(t) + \dot{\gamma}(t) \times \boldsymbol{B}_{\gamma(t)}(t)\right).$$

One of the great triumphs of 19th century physics was the discovery of exactly how these fields relate to one another: the *unification* of electricity and magnetism. These are famously summarised by Maxwell's equations, which – in a vacuum – take the form

$$\nabla \cdot E = 0 \qquad\qquad \nabla \cdot B = 0$$
$$\nabla \times E = -\frac{\partial B}{\partial t} \qquad\qquad \nabla \times B = \frac{\partial E}{\partial t}$$

Here, and throughout this talk, I'll be using units where the speed of light $c = 1$. Later on I'll probably set Planck's constant $\hbar = 1$ also. On $\mathbb{R}^3$ this is all well and good, but we'd like to understand the behaviour of electric and magnetic fields on a more *general* 3-manifold M. What will replace these divs and curls? There is a natural way of rephrasing this story that will make the answer apparent: dualise everything, and phrase everything in terms of *differential forms*.

Definition 1.2 (Second Approximation)**.** The *electric field* E and the *magnetic field* B are 1-forms in $\Omega^1(\mathbb{R}^3)$, varying with time.

What form does the Lorentz force law now take? Given a time dependent vector field X, there is a time dependent 1-form built out of the 1-forms $\boldsymbol{E}$ and $\boldsymbol{B}$, namely

$$F(t) = E(t) + \dot{X}(t) \vee *B(t)$$

where $\dot{X} \vee -$ denotes contraction with the vector field $\dot{(X)}$, and $*$ denotes the Hodge star (coming from the standard metric on $\mathbb{R}^3$). This is what we might call the *field strength* associated to the electric and magnetic fields E and B: the Lorentz force law says that that a particle with charge 1 following a trajectory given by X experiences force given by the dual of this 1-form (again, using the metric). Notice here that the force does not depend on B itself, but rather on its Hodge dual. So let's treat this 2-form as fundamental instead:

Definition 1.3 (Third Approximation)**.** The *electric field* E is a 1-form in $\Omega^1(\mathbb{R}^3)$, and the *magnetic field* B is a 2-form in $\Omega^2(\mathbb{R}^3)$, varying with time.

We can explicitly include the time dependence in our setup by defining a *field strength 2-form* $F = E \wedge dt + B \in \Omega^2(\mathbb{R}^4)$. The Lorentz force law now says that the force experienced by a particle is given by

$$\dot{X} \vee F$$

where now $\dot{X}$ describes the "4-velocity" of a particle moving in $\mathbb{R}^4$.

In this language, Maxwell's equations take a particularly simple form. In a vacuum as before they now say, simply that

$$dF = 0 \quad d^*F = 0$$

where $d^* = *d*$ is the formal adjoint to d under the metric. In other words, *electromagnetic* fields are nothing but 2-forms on $\mathbb{R}^4$ satisfying these two equations: "harmonic" 2-forms. Even more satisfying is that these equations make perfect sense where $\mathbb{R}^3$ *any* 3-manifold M with a Riemannian metric, or even when the whole *spacetime* is replaced by a general 4-manifold N with a Riemannian or pseudo-Riemannian metric. In light of the theory of relativity, being able to work on a general Lorentzian manifold is extremely important.

At this point, I can't resist mentioning the evident symmetry here: the Hodge star acts on the space $\Omega^2(N)$, preserving the harmonic 2-forms. On the level of electric and magnetic fields, this is visible as the symmetry

$$E \mapsto B, B \mapsto -E$$

of the space of solutions to Maxwell's equations. This is usually referred to as "electric-magnetic duality". It no longer holds once one introduces an electric charge distribution into the setting, in which case the second Maxwell equation becomes $d^*F = J$, but one can restore the symmetry by positing the existence of a dual "magnetic charge distribution".

This discussion would not be complete without mentioning the natural Lagrangian framework these equations live in. Maxwell's equations arise as the equations of motion for a certain *Lagrangian*. That is, solutions to Maxwell's equations are nothing but critical points of the "Yang-Mills functional"

$$S(F) = \int_M F \wedge *F.$$

2 Geometry

Now, I'll give a quick outline of some of the geometry that'll be important in our eventual description of electromagnetic fields, namely the concept of a connection on a principal $U(1)$-bundle. Let M be a smooth n-manifold, and let $\pi \colon P \to M$ be a principal circle bundle. Such a bundle is determined up to isomorphism by its first Chern class, an element $c_1(P) \in H^2(M; 2\pi\mathbb{Z})$. In other words, this is the *degree* of the associated complex line bundle $P \times_{U(1)} \mathbb{C}$.

The basic idea of a *connection* on P is the following. We want to be able to describe *parallel transport* for tangent vectors on P. In order to do this, we *either* need to specify the space of horizontal vectors in each tangent space T_pP (so that we have a canonical splitting of each tangent space into horizontal and vertical vectors), *or* equivalently we need to specify a 1-form which vanishes in (annihilates) the horizontal directions. When one unpacks what this means in the case of a circle bundle, one arrives at the following definition.

Definition 2.1. A *connection* on P is a 1-form $A \in \Omega^1(P)$, which is invariant for the action of $U(1)$ on P, and restricts to the fundamental 1-form $i\frac{dz}{z}$ on each fibre (the infinitesimal generator of the action of $U(1)$ on itself). The *horizontal* vector fields on P are given by the kernel of the action of A on vector fields.

The way the physicists like to think about connections is the following: choose a trivialisation ($s_\alpha \colon U_\alpha \to P$) of P, and pull back the connection 1-form to a collection of 1-forms $A_\alpha \in \Omega^1(U_\alpha)$. Physicists call these "local gauge potentials". Modifying the trivialisation by $s_\alpha \mapsto f_\alpha s_\alpha$, where $f_\alpha \colon U_\alpha \to U(1)$ has the effect of modifying the gauge potentials by

$$A_\alpha \mapsto A_\alpha + f_\alpha^{-1} ds_\alpha.$$

In particular, the derivative dA_α is unchanged. Therefore the derivatives glue together to form a *global* closed 2-form

$$F_A \in \Omega^2(M)$$

the *curvature* of the connection A. This curvature has an interesting cohomological property

Fact (Chern-Weil)**.** The curvature F_A of any connection A on a principal $U(1)$-bundle P satisfies

$$[F_A] = c_1(P) \in H^2(M).$$

In particular $[F_A]$ is an *integral* cohomology class.

To summarise, to the connection A on the principal bundle P, we have associated a *closed, integral* 2-form, the *curvature* of A. In fact, all such 2-forms arise in this way, as the curvature of some connection on the principal bundle P with the appropriate first Chern class, although in general there will be more than one connection yielding this particular curvature.

3 Dirac Quantisation

So what does all this have to do with electromagnetism? We argued that classical electromagnetism on a 4-manifold was the theory of closed 2-forms satisfying a particular differential equation (harmonic 2-forms). What does this have to do with connections? Well, we'll see that once one introduces a little quantum mechanics into the picture, the electromagnetic field strength has to satisfy a so-called *quantisation condition*, making it into a *closed integral* 2-form, i.e. exactly the curvature of a connection on a principal $U(1)$-bundle. I'll sketch this argument, then explain what it all means geometrically.

For simplicity, consider a 4-manifold of form $M^3 \times \mathbb{R}$, where M is an oriented Riemannian 3-manifold. We'll consider charged *quantum* particles moving in a fixed electromagnetic field. What does this mean? Well, let's review some basic quantum mechanics. A quantum particle moving in a potential A is described by a "wavefunction" ψ (i.e. an L^2-function) satisfying the Schrödinger equation

$$i\frac{\partial \psi}{\partial t} = -(\nabla + iA)^2 \psi.$$

We'll work in local charts U_α for M. On these charts we can write $F|_{U_\alpha} = dA_\alpha$ as the derivative of a local electromagnetic potential, which we think of as the potential energy of a charged particle moving under this field. What happens when we try to solve Schrödinger's equation in this electromagnetic potential? We can do this locally, to get local solutions ψ_α on U_α. But when can we glue these solutions to form a global solution?

In order to see this we'll analyse what happens when we choose a *different* local potential, $A_\alpha + d\chi$. How does this affect the solutions to the Schrödinger equation? It is not hard to check that is ψ_α was a solution under potential A_α, then $e^{-i\chi}\psi_\alpha$ is a solution under potential $A_\alpha + d\chi$. What does it mean for these solutions to glue to a global gauge invariant wavefunction? Well, let's imagine travelling around a closed loop γ in M. Whenever we pass between two neighbouring charts U_α and U_β we pick up a transition function $d\chi_{\alpha\beta} = A_\alpha - A_\beta$. Thus, travelling around the whole loop we find that we require a compatibility condition for gauge invariance:

$$\psi_\alpha = e^{-i(\chi_{\alpha_1}(0) - \chi_{\alpha_1}(1) + \cdots + \chi_{\alpha_n}(0) - \chi_{\alpha_n}(1))}\psi_\alpha.$$

That is, we need $\sum_{i=1}^n \chi_{\alpha_i}(0) - \chi_{\alpha_i}(1) \in 2\pi\mathbb{Z}$. Here, I'm thinking of my transition functions χ_{α_i} as functions $[0,1] \to M$ partitioning the loop γ.

Ok, so in order for solutions to be independent of the choice of gauge (i.e. to glue to a single-valued global solution instead of a mult-valued one) we need a condition for any collection of local gauge transformations around any loop. We can think of the global wavefunctions as not really functions on $M \times \mathbb{R}$, but sections of a possibly non-trivial $U(1)$-bundle P which we might call the *phase bundle*, and imagine the fibre as keeping track of the "phase" of the quantum particle. This is starting to sound promising. I'll give a Čech argument that phrases the restriction on A in a more cohomological way.

Claim. This condition is equivalent to the condition that the curvature F_A has *integral periods*, i.e. that for any closed surface $\Sigma \subseteq M \times \mathbb{R}$ (without loss of generality $\subseteq M$), we have

$$\int_\Sigma F_A \in 2\pi\mathbb{Z},$$

or equivalently $[F_A] \in H^2(M \times \mathbb{R}; 2\pi\mathbb{Z})$.

Proof. We apply Stokes' theorem. Choose any cell decomposition of M. Each two-cell D_i is contractible, so we can locally write $F|_{D_i} = dA_i$ for $A_i \in \Omega^1(D_i)$. Furthermore, on each 1-cell ℓ_{ij}, we can write $A_i|_{\ell_{ij}} - A_j|_{\ell_{ij}} = d\chi_{ij}$ for $\chi_{ij} \in \Omega^0(\ell_{ij})$ a local function. We used the orientation to order the pair (i,j) here. From this point of view we can repeatedly apply Stokes' theorem to localise the electromagnetic flux to an expression on the 0-skeleton. That is:

$$\begin{aligned}
\int_\Sigma F &= \sum_{i=1}^n \int_{D_i} dA_i \\
&= \sum_{i=1}^n \int_{\partial D_i} A_i \\
&= \sum_{i=1}^n \sum_{j<i} \int_{\ell_{ij}} A_i|_{\ell_{ij}} - A_j|_{\ell_{ij}} \\
&= \sum_{i=1}^n \sum_{j<i} \int_{\ell_{ij}} d\chi_{ij} = \sum_{i=1}^n \sum_{j<i} \chi_{ij}(1) - \chi_{ij}(0).
\end{aligned}$$

Asking for this to land in $2\pi\mathbb{Z}$ for any cell decomposition is equivalent for asking for the analogous sum to land in $2\pi\mathbb{Z}$ around any loop, which is exactly our quantisation condition. □

So the result is, we've shown that – for physical reasons – any electromagnetic field must be described by a 2-form with integral periods, i.e. precisely *the curvature of a connection on a principal $U(1)$-bundle*. So I've met the goal I set myself at the outset of this talk. But what does it really mean?

Well, let's think about what role the connection itself played. The data of a connection on a principal $U(1)$ bundle was the data of a local *potential* for the electromagnetic field strength in each chart of a trivialisation. These potentials do not have to glue to form a global 1-form, but the ambiguity in the attempt to glue is limited by our quantisation condition: roughly we can glue "up to local circle valued functions $e^{i\chi}$". The potentials are only defined up to a gauge symmetry, which corresponds to changing the trivialisation of the bundle, or in other words to *automorphisms* of the circle bundle.

A particularly neat consequence of this geometric description is that it now makes sense for *groups other than* $U(1)$! We can look at principal G-bundles for any compact Lie group G, and say that the "fields" in a non-abelian gauge theory are connections on such a bundle. Now, we can once more write the action

$$S(A) = \int_N \mathrm{Tr}(F_A \wedge *F_A)$$

using the metric on spacetime N, and look at the extrema of this action. The study of these PDEs is called *Yang-Mills* theory, and is important both in mathematics (e.g. the study of the topology of 4-manifolds), and in physics (e.g. for $G = SU(3)$, the quantum version of this theory – with matter – models the strong nuclear force).