Apparent overdetermination in Maxwell's equations and the weirdness of curl

Daniel Duffy, dld34@cam.ac.uk

Maxwell's equations *seem* to have more equations than degrees of freedom ('unknowns') in many reasonable problems, forming an overdetermined system. Writing the equations as

$$\nabla \cdot \boldsymbol{E} = \rho/\varepsilon_0, \tag{1}$$
$$\nabla \cdot \boldsymbol{B} = 0, \tag{2}$$
$$\nabla \times \boldsymbol{E} = -\dot{\boldsymbol{B}}, \tag{3}$$
$$\nabla \times \boldsymbol{B} = \mu_0 \boldsymbol{J} + \mu_0 \varepsilon_0 \dot{\boldsymbol{E}}, \tag{4}$$

we see that we have 8 scalar equations in total, since (3) and (4) each have 3 components. Let's think about a very common and reasonable class of problem: given some specified $\boldsymbol{J}$ and ρ that satisfy the charge continuity equation[1] $\dot{\rho} + \nabla \cdot \boldsymbol{J} = 0$, what are the resulting $\boldsymbol{E}$ and $\boldsymbol{B}$? The degrees of freedom are the components of $\boldsymbol{E}$ and $\boldsymbol{B}$, so we have only 6 dofs and things don't look good for this problem — it seems like there won't in general be enough freedom in $\boldsymbol{E}$ and $\boldsymbol{B}$ to satisfy the equations!?

There's a fairly well known 'solution' to this alarming observation [1–4], which goes something like this: Take the divergence of (3) and (4) and exchange the order of time and space derivatives to get

$$0 = -\frac{\mathrm{d}}{\mathrm{d}t} \left(\nabla \cdot \boldsymbol{B} \right), \tag{5}$$
$$0 = \nabla \cdot \boldsymbol{J} + \frac{\mathrm{d}}{\mathrm{d}t} \left(\nabla \cdot \varepsilon_0 \boldsymbol{E} \right). \tag{6}$$

If we sub our charge continuity condition into the latter it becomes

$$0 = \frac{\mathrm{d}}{\mathrm{d}t} \left(\nabla \cdot \boldsymbol{E} - \rho/\varepsilon_0 \right). \tag{7}$$

From (5) and (7) we see that as long as the *initial* $\boldsymbol{E}$ and $\boldsymbol{B}$ satisfy (1) and (2), both (1) and (2) are actually redundant in describing the evolution, because

[1] The reason I'm specifying this is that if ρ and $\boldsymbol{J}$ don't satisfy charge continuity, then there are *definitely no* solutions to Maxwell's equations. To see this, take the divergence of (4) and then sub in (1) and the charge continuity equation pops out — so charge continuity *has* to hold for Maxwell's equations to hold!

(3) and (4) will automatically ensure that (1) and (2) remain satisfied for all time[2].

I don't find this 'solution' fully satisfying for at least one reason: it leans on time dynamics in a big way, but electrostatics should be a perfectly well defined (and self-contained) subset of electromagnetic theory, completely consistent with Maxwell's equations and with no overdetermination problems. In electrostatics we can for sure specify whatever charge distribution ρ we like, and then find the resulting $\boldsymbol{E}$, while $\boldsymbol{B} = \boldsymbol{J} = \mathbf{0}$. But the Maxwell's equations corresponding to such an electrostatic problem also appear to be overdetermined, in just the same way as before! The components of $\boldsymbol{E}$ are our 3 dofs, and we have to solve the equations

$$\nabla \cdot \boldsymbol{E} = \rho/\varepsilon_0, \tag{8}$$
$$\nabla \times \boldsymbol{E} = 0, \tag{9}$$

given some specified static ρ. This is 4 scalar equations! In my opinion, a complete understanding of the overdetermination problem must be able to explain this problem, in which time does not feature at all.

I believe that it basically boils down to the subtleties of the curl operator, and that (9) is effectively only 2 scalar equations. Consider how we actually solve this problem usually: we replace (9) with $\boldsymbol{E} = -\nabla\phi$, and sub into (8) to get the Poisson equation for ϕ (see my note *Electrostatics and simply connectedness* for a standard derivation of this). Notice what's happened here — as soon as we replace (9) with $\boldsymbol{E} = -\nabla\phi$, we have only a 1-dof field ϕ, whereas before we had 3 dofs. In other words, the 3 scalar equations (9) actually only eliminate 2 dofs to leave us with 1, even though you might expect them to eliminate 3. . . sorcery!

We can look at magnetostatics similarly:

$$\nabla \cdot \boldsymbol{B} = 0, \tag{10}$$
$$\nabla \times \boldsymbol{B} = \mu_0 \boldsymbol{J}, \tag{11}$$

for some given static $\boldsymbol{J}$ that has $\nabla \cdot \boldsymbol{J} = 0$. Again, 3 dofs and 4 scalar equations on the face of it. But if we again take the curl equation to only be 2 scalar

[2] This is reminiscent of the fact that when solving $\nabla^2 \boldsymbol{A} = -\mu_0 \boldsymbol{J}$ for some source $\boldsymbol{J}$ that has $\nabla \cdot \boldsymbol{J} = 0$, if you impose $\nabla \cdot \boldsymbol{A} = 0$ as a BC then your solution will actually have $\nabla \cdot \boldsymbol{A} = 0$ everywhere (see my *Consistency of Coulomb gauge* post). Sometimes equations can be replaced just with BCs!

equations effectively, then we see that the problem is actually properly determined.

I think these statics examples tell us that the curl of any field in some sense only 'operates on' two degrees of freedom that the field has, hence why setting the curl equal to zero still leaves the field with a single dof. This weirdness of curl is consistent with the Helmholtz decomposition, which splits a very general 3-dof field into a 1-dof $\nabla\phi$ term, and a divergenceless part that has a curl — the part that has a curl *must* therefore only correspond to 2 dofs! Sadly, I haven't really come up with any intuition for why curl has this property — you might say "it's because $\operatorname{div}\operatorname{curl} = 0$ shows that the 3 components of a curl have a constraint relating them, so there are only 2 dofs" ... but you could say something similar about $\operatorname{curl}\operatorname{grad} = 0$, which doesn't eliminate the single dof that a term like $\nabla\phi$ has! Plus, I don't think $\operatorname{div}\operatorname{curl} = 0$ is any more intuitive really.

Anyway, inspired by the static cases, my preferred 'solution' to the apparent overdetermination of the full Maxwell equations is that the two curl equations (3) and (4) should each be interpreted as 2 scalar constraints each rather than 3. This is somewhat the other way round from the conventional 'solution': I'd say that the divergence equations (1) and (2) provide 1 constraint each, removing 1 dof from $\boldsymbol{E}$ and 1 from $\boldsymbol{B}$, and also guaranteeing that the curl equations (3) and (4) have divergenceless fields on both sides and so are mathematically consistent. Via the weirdness of curl, (3) then constrains both of the remaining 2 dofs in $\boldsymbol{E}$, and (4) does the same for $\boldsymbol{B}$. Thinking in Helmholtz decomposition terms, (1) and (2) are equations constraining the (1-dof) grad parts of $\boldsymbol{E}$ and $\boldsymbol{B}$ respectively, while (3) and (4) are equations constraining the other (2-dof, divergenceless) parts. Everything adds up, and this way of looking at things works for both the static and time-varying cases!

Finally, I'll just note that another angle on all this: if we write $\boldsymbol{E}$ and $\boldsymbol{B}$ in terms of potentials $\boldsymbol{E} = -\nabla\phi - \dot{\boldsymbol{A}}$ and $\boldsymbol{B} = \nabla \times \boldsymbol{A}$, and go back to the full Maxwell's equations, then (2) and (3) are satisfied automatically, so we're left with 4 scalar equations, and 4 scalar unknowns ϕ and $\boldsymbol{A}$. But the system you're left with,

$$-\nabla^2\phi - \nabla \cdot \dot{\boldsymbol{A}} = \rho/\varepsilon_0, \tag{12}$$

$$\nabla \times (\nabla \times \boldsymbol{A}) = \mu_0 \boldsymbol{J} - \mu_0\varepsilon_0(\nabla\dot{\phi} + \ddot{\boldsymbol{A}}), \tag{13}$$

is actually *under*determined! The curl equation is again really only providing 2 constraints, even though it is 3 scalar equations. We can also see it like this: if you have a solution to the above system, then you get another solution by adding any ∇f to $\boldsymbol{A}$ as long as you also subtract $\dot{f}$ from ϕ. To get a properly determined system we have to add an arbitrary 'gauge fixing' condition! This freedom is called gauge invariance, and I think it's another manifestation of the same weirdness of curl. That curl's weirdness is intertwined with both gauge invariance *and* the overdetermination problem is part of why it's a good way to think about these things imo.

If I'm right about curl's weirdness being the key thing in all this, then perhaps apparent overdetermination can turn up similarly in other theories where the equations have curl in them, e.g. fluid dynamics?

References

[1] Stratton, *Electromagnetic Theory*

[2] Rosen, *Redundancy and superfluity for electromagnetic fields and potentials*

[3] `https://physics.stackexchange.com/questions/20071`

[4] `https://physics.stackexchange.com/a/557016`