Topology, holes and sources

Alexander Afriat

July 15, 2012

Abstract

The Aharonov-Bohm effect[1] is typically called "topological." But it seems no more topological than magnetostatics, electrostatics or Newton-Poisson gravity (or just about any radiation, propagation from a source). I distinguish between two senses of "topological."

1 The Aharonov-Bohm effect

A wavefunction is split into two, and these, having enclosed a (simply-connected) region ω containing a solenoid, are made to interfere on a screen. The enclosing wavefunction is sensitive to any enclosed electromagnetism inasmuch as the electromagnetic potential[2] A, a one-form, contributes a phase

$$\exp i \oint_{\partial\omega} A$$

to (the wavefunction along) the boundary $\partial\omega$ and hence to the interference pattern on the screen. The electromagnetism on ω is related to the circulation around the boundary by Stokes' theorem

$$\oint_{\partial\omega} A = \iint_{\omega} dA. \tag{1}$$

The electromagnetic field[3] $F = dA$ produced by the solenoid is circumscribed to a middle region $\lambda \subset \omega$ surrounded by an isolating region[4] $\lambda' = \omega - \lambda$ where F vanishes but not A. The full *Aharonov-Bohm effect* can be considered the 'differential' or 'incremental' sensitivity of the interference pattern to variations in the current through the solenoid.

[1] Ehrenberg & Siday (1949), Aharonov & Bohm (1959)

[2] By "potential" I just mean *primitive*: the potential of the electromagnetic two-form $F = dA$ is its primitive $A \leftrightarrow (\mathbf{A}, \varphi)$, the potential of the magnetic two-form $\mathbf{B} = d\mathbf{A}$ is its primitive $\mathbf{A}$, the potential of the electric one-form $\eta = *\mathbf{E} = d\varphi$ is its primitive φ (the Hodge dual $*$ being taken in three dimensions), the potential of the three-form density $\rho = d\mathbf{E}$ is its primitive $\mathbf{E}$.

[3] It is perhaps easiest to think of F as a purely *magnetic* field $\mathbf{B}$ produced by the current density $\mathbf{J} = d{*}\mathbf{B}$ in the solenoid.

[4] It will be convenient to view λ and ω as concentric disks.

2 The topological$_1$ interpretation

I will distinguish between two (related but) different senses of "topological":

1. *Topological*$_1$: related to the absence or presence of a hole (which may or may not contain a source, like a solenoid or a charge).

2. *Topological*$_2$: invariant under appropriate continuous deformations.

The *topological*$_1$ *interpretation*[5] of the Aharonov-Bohm effect can be formulated as follows: If A were closed throughout a simply-connected region ω it would also be exact, and hence expressible as the gradient $A = d\mu$ of a zero-form μ (a real-valued function); the flux

$$\oint_{\partial\omega} d\mu = \iint_{\omega} d^2\mu$$

through the boundary $\partial\omega$ would then vanish, since $d^2 = 0$. But here A is closed on λ'; from $dA = 0|_{\lambda'}$ it does not follow that A is exact, nor that the flux through the enclosing loop vanishes: it may or may not.

The existence of the source responsible for the effect is therefore ruled out by one topology (A closed throughout a simply-connected region) but not another.

The same applies to a simply-connected three-dimensional region Ω enclosed by a two-dimensional boundary $\partial\Omega$. If the two-form $\mathbf{E}$ were closed throughout Ω it would also be exact, and hence expressible as the curl $\mathbf{E} = d\zeta$ of a one-form ζ; the flux

$$\iint_{\partial\Omega} d\zeta = \iiint_{\Omega} d^2\zeta$$

[5] Aharonov & Bohm (1959, p. 490): "in a field-free multiply-connected region of space, the physical properties of the system still depend on the potentials." Wu & Yang (1975, p. 3845): "The famous Bohm-Aharonov experiment [...] showed that in a multiply connected region where $f_{\mu\nu} = 0$ everywhere there are physical experiments for which the outcome depends on the loop integral [...] around an unshrinkable loop." And p. 3856: "$f_{\mu\nu}$ underdescribes electromagnetism because of the Bohm-Aharonov experiment which involves a doubly connected space region." Nash & Sen (1983, p. 301): "We [...] consider the consequence of assuming the field **F** to be identically zero in some region Ω. At first one may think that there will be no physically measurable electromagnetic effects in such a region Ω. This is not so, effects may arise if the topology of Ω is non-trivial, e.g. if Ω is not simply connected. [...] In terms of parallel transport one says that zero curvature does not imply trivial parallel transport if the region in which the curvature is zero is not simply connected. This underlies the fact that there is a sense in which the connection is a more fundamental object than the curvature, even though a connection is gauge dependent and not directly measurable." Ryder (1996, p. 101-4): "the Bohm-Aharonov effect owes its existence to the non-trivial topology of the vacuum [...]. The Bohm-Aharonov effect is the simplest illustration of the importance of topology in this branch of physics. [...] The relevant space in this problem is the space of the vacuum, i.e. the space outside the solenoid, and that space is not simply connected. [...] It is thus an essential condition for the Bohm-Aharonov effect to occur that the configuration space of the vacuum is not simply connected. [...] in other words, it is because the gauge group of electromagnetism, U_1, is not simply connected that the Bohm-Aharonov effect is possible. [...] The configuration space of the Bohm-Aharonov experiment is the plane $\mathbb{R}^2$ [...] with a hole in, and this is, topologically, the direct product of the line $\mathbb{R}^1$ and the circle [...]. There is, nevertheless, a positive effect on the interference fringes. The mathematical reason for this is that the configuration space of the null field (vacuum) is the plane with a hole in [...]." Martin (2003, p. 48): "in the case of non-trivial spatial topologies, the gauge-invariant interpretation runs into potential complications. [...] So-called holonomies [...] encode physically significant information about the global features of the gauge field." Agricola & Friedrich (2010, p. 275): "so ist das verbleibende Gebiet $\Omega - S$ der Ebene aus Sicht des Elektrons nicht mehr einfach zusammenhangend." See also Nounou (2003).

through the boundary would then vanish. But if the region on which $\mathbf{E}$ is closed has a hole in it, the flux through the enclosing surface may or may not vanish.

This is precisely what we have in electrostatics, where the electric field $\mathbf{E} = *d\varphi$ is (Hodge-dual to) the gradient $d\varphi$ of the scalar potential φ. The vanishing divergence $d\mathbf{E}$ expresses the conservation of electricity where none is created, away from the charges that produce $\mathbf{E}$ according to the Maxwell-Poisson equation $d\mathbf{E} = d{*}d\varphi = \rho$, ρ being the charge density. If the divergence $d\mathbf{E}$ vanished throughout the volume Ω, there would be no electricity produced and hence none radiated through the enclosing surface.[6] But a charge in Ω—say in a region[7] $\Lambda \subset \Omega$ isolated by $\Lambda' = \Omega - \Lambda$—would prevent electricity from being conserved throughout Ω.

We have the same formalism in Newton-Poisson gravity, where φ is the gravitational potential, $d\varphi$ and $\mathbf{E}$ both represent gravitational force, and ρ is the mass density. Gravity[8] would therefore be another topological$_1$ effect.

Again, the topology of the region where the 'potential' (A or $\mathbf{E}$ or whatever) is closed tells us relatively little: if the region were simply-connected, conservation would be *general* within the enclosing surface since there could be no holes containing sources; and if nothing were created inside the enclosing surface, the total radiation through it would vanish. But if the topology does *not* allow the presence of holes to be ruled out, the presence of sources in them cannot either; and sources would produce a flux through the enclosing surface.

A non-trivial topology cannot, on its own at any rate, rule out the *absence* of a source either. Nor does it provide the 'amount' or 'intensity' of the possible source (which would tell us the intensity of the effect—the flux through the enclosing surface). So the full Aharonov-Bohm effect, which can be considered 'differential,' is hardly topological$_1$, or at any rate no more so than electrostatics or Newton-Poisson gravity.

3 Topological$_2$ effects

3.1 Aharonov-Bohm

The Aharonov-Bohm effect is topological$_2$ in the sense that certain basic quantities (say the contour integral (1) and resulting interference pattern) are invariant under appropriate continuous deformations. It seems that (fundamental aspects of) electrostatics and Newton-Poisson gravity are just as topological$_2$. Magnetostatics may be even more topological$_2$.

Since the electromagnetic field $F = dA = dA'$ is easier to measure than the potential A, it is customary to view the freedom expressed by the substitution

(2) $$A \mapsto A' = A + d\xi$$

(ξ being a zero-form) as unobservable. Such transformations deform the level sets of A's local potential[9] γ. One can first imagine a purely 'angular' γ (with values running

[6] Over and above any divergence-free electrical background that may or may not be present.

[7] It will be convenient to view Λ and Ω as concentric spheres.

[8] Or rather the *total* gravitational attraction radiated by a mass.

[9] For wherever A is closed it can be written locally as the gradient $A = d\gamma$ of a zero-form γ—just as $\mathbf{E}$ can be written locally, wherever it is closed, as the curl $\mathbf{E} = d\zeta$ of a one-form ζ.

from say 0 to 2π),[10] whose level lines are straight rays radiating through the annulus λ' from the inner disk λ to the edge $\partial\omega$. A gauge transformation (2) would then deform the level rays, bending them without making them touch. The circle $\partial\omega$ can likewise be deformed into any loop going around the solenoid once. The Aharonov-Bohm effect is topological$_2$ in the sense that neither deformation affects the integral (1) (or the resulting interference pattern).

Whereas here the deformations are allowed by (and part of) theory, in the next cases they will be counterfactual.

3.2 Electrostatics (& Newton-Poisson gravity)

The basic law here, the Gauß-Maxwell equation

$$\iint_{\partial\Omega} \mathbf{E} = \iiint_{\Omega} d\mathbf{E} = \iiint_{\Omega} \rho \tag{3}$$

(or $d\mathbf{E} = \rho$), is topological$_2$ inasmuch as the boundary $\partial\Omega$ and electric field lines can be continuously deformed without affecting the integral (3)—which is the electrostatic analog of the (contour integral giving rise to the) Aharonov-Bohm effect. We can imagine a spherically symmetric charge distribution ρ contained in $\Lambda \subset \Omega$ (everything concentric): the electrical field lines radiated by the charge in Λ correspond to the level rays radiating from the solenoid in λ. Nothing in electrostatics prevents the deformation of $\partial\Omega$. Admittedly the electric rays cannot be bent without violating $\mathbf{E} = *d\varphi$; despite preserving the divergence $d\mathbf{E} = d\mathbf{E}'$, the transformation

$$\mathbf{E} \mapsto \mathbf{E}' = \mathbf{E} + d\alpha \tag{4}$$

(the three-dimensional version of (2), α being a one-form) is counterfactual—which does not prevent a conditional characterisation of the effect as topological$_2$: "the integral *would* remain the same *even if* the field lines *were* bent."

Most of this applies (*mutatis mutandis*) to Newton-Poisson gravity.

Why bother with obvious facts about integration? Because much is made in the literature of the homotopically deformable loop $\partial\omega$ (which corresponds to the homotopically deformable surface $\partial\Omega$) and gauge transformation (2) (which corresponds to the admittedly counterfactual transformation (4)).

3.3 Magnetostatics

The basic law here, Mawell's equation

$$\iint_{\partial\Omega} \mathbf{B} = \iiint_{\Omega} d\mathbf{B} = 0$$

(or $d\mathbf{B} = 0$), holds because a magnet has two poles, that act as source and sink of the same field lines, which form loops going from one pole to the other: all magnetism

[10] Such a γ cannot be continuous everywhere; we can imagine a single discontinuity, say on the ray $\gamma = 0 = 2\pi$.

produced is eventually recovered. If a magnetic loop crosses the boundary $\partial\Omega$ it will cross it again on the way back to the magnet, thus erasing whatever it contributed to the integral on the way out. The law is topological$_2$ in that the boundary $\partial\Omega$ and field lines can be deformed[11] continuously without affecting the integral. Nothing in magnetostatics prevents the deformation of $\partial\Omega$; the deformation of the field lines by $\mathbf{B} \mapsto \mathbf{B}' = \mathbf{B} + d\beta$ is again counterfactual, being prevented by the observability of the magnetic field $\mathbf{B} = d\mathbf{A}$.

But if I continue to dwell on these old three-dimensional theories I may give the impression I want to make a point about *them*, whereas my real point concerns the Aharonov-Bohm effect: that it is hardly topological, or at any rate no more topological than electrostatics *etc*.

My thanks to Nazim Bouatta, Dennis Dieks, Éric Gourgoulhon, Marc Lachièze-Rey and Jean-Philippe Nicolas for valuable clarifications and corrections.

References

Agricola, I. and T. Friedrich (2010) *Vektoranalysis: Differentialformen in Analysis, Geometrie und Physik*, Vieweg+Teubner, Berlin

Aharonov, Y. and D. Bohm (1959) “Significance of electromagnetic potentials in the quantum theory” *Physical Review* **115**, 485-91

Brading, K. and E. Castellani, editors (2003) *Symmetries in physics*, Cambridge University Press

Ehrenberg, W. and R. E. Siday (1949) “The refractive index in electron optics and the principles of dynamics” *Proceedings of the Physical Society B* **62**, 8-21

Martin, C. (2003) “On continuous symmetries and the foundations of modern physics” pp. 29-60 in Brading & Castellani (2003)

Nash, C. and S. Sen (1983) *Topology and geometry for physicists*, Academic Press, London

Nounou, A. (2003) “A fourth way to the Aharonov-Bohm effect” pp. 174-99 in Brading & Castellani (2003)

Ryder, L. (1996) *Quantum field theory*, Cambridge University Press

Wu, T. T. and C. N. Yang (1975) “Concept of nonintegrable phase factors and global formulation of gauge fields” *Physical Review D* **12**, 3845-57

[11] Deformations of $\partial\Omega$ can of course lead to the exclusion or inclusion of certain loops.