# Shortening the gauge argument

Alexander Afriat

August 25, 2012

**Abstract**

The "gauge argument" is often used to 'deduce' interactions from a symmetry requirement. A transition—whose justification can take some effort—from global to local transformations is typically made at the beginning of the argument. But one can spare the trouble by *starting* with local transformations, as global ones do not exist in general. The resulting economy seems noteworthy.

## 1 The gauge argument

One begins with a free field, of Dirac four-spinors $\psi \in \mathbb{C}^4$ for instance. The Dirac Lagrangian $\mathscr{L} = \bar{\psi}(i\partial\!\!\!/ - m)\psi$ is invariant under the *global* transformation

$$\psi \mapsto e^{i\kappa}\psi \tag{1}$$

(whose existence is assumed for the time being), where "global" means that $\kappa$ is a constant. It is then argued[1] that $\mathscr{L}$ should also be invariant under the *local* transformation

$$\psi \mapsto \psi_\zeta = e^{i\zeta}\psi, \tag{2}$$

where $\zeta : M \to \mathbb{R}$ is a smooth function on the base manifold $M$.

> Most immediately what are we to make of the initial, central demand of local gauge invariance? The demand is anything but self-evident and presumably, in the context of the gauge argument, must be argued for on some

[1] Göckeler & Schücker (1987) p. 48: "In physical terms we may interpret the requirement of local gauge invariance (independence of the fields at different spacetime points) as expressing the absence of (instantaneous) action at a distance." Ryder (1996) p. 93: "So when we perform a rotation in the internal space of $\varphi$ at one point, through an angle $\Lambda$, we must perform the same rotation *at all other points at the same time*. If we take this physical interpretation seriously, we see that it is impossible to fulfil, since it contradicts the letter and spirit of relativity, according to which there must be a minimum time delay equal to the time of light travel. To get round this problem we simply abandon the requirement that $\Lambda$ is a constant, and write it as an arbitrary function of space-time, $\Lambda(x^\mu)$. This is called a 'local' gauge transformation, since it clearly differs from point to point." Teller (2000) p. S469: "why should we expect invariance under a local phase transformation to begin with? The plausibility of such invariance probably arises with a misleading analogy with global phase transformations which can be imposed on individual state functions with no change of description." See also Sakurai (1967, p. 16), Aitchison & Hey (1982, p. 176), Mandl & Shaw (1984, p. 263), Ramond (1990, pp. 183-91), O'Raifeartaigh (1997, p. 118). One is reminded of Weyl's rejection (1929a, p. 331; 1929b, p. 286) of distant parallelism.

basis. Unlike the global invariance, the demand for the corresponding local invariance does not have an immediate physical counterpart. Is it to be taken as a direct implementation of some sort of unassailable first principle? If so, is the demand (or principle) something with which we are already familiar only in a different form?

A common justification for the demand of local gauge invariance in presenting the gauge argument is to present it as some sort of "locality" requirement. In outline, the "gauge locality argument" is that global gauge invariance is somehow at odds with the idea of a local field theory, and that to remedy this we must instead require local gauge invariance. This rather brief argument is just how Yang and Mills motivated the demand in their seminal 1954 paper,[2] very much setting the tone for subsequent treatments. Just what to make of this argument is not clear, however, there are many interrelated senses of locality that might be at issue. (Martin, 2002, p. S225)

One gathers at any rate that considerable and varied efforts have been devoted to the transition from (1) to (2).

As things stand the Lagrangian is not invariant, because of the derivative in the first term of $\mathscr{L}_\zeta = i\bar{\psi}_\zeta \partial\!\!\!/ \psi_\zeta - m\bar{\psi}\psi$. Writing

$$\bar{\psi}_\zeta \partial\!\!\!/ \psi_\zeta = \bar{\psi} e^{-i\zeta} \gamma^\mu \partial_\mu e^{i\zeta} \psi = \bar{\psi} \gamma^\mu (\partial_\mu + i\partial_\mu \zeta)\psi$$

we see that $\mathscr{L}_\zeta = \bar{\psi}[i\gamma^\mu(\partial_\mu + i\partial_\mu\zeta) - m]\psi$ has the derivative $\partial_\mu + i\partial_\mu\zeta$ instead of $\partial_\mu$. To offset (2) we therefore have to *subtract* the term $i\partial_\mu\zeta$ that alters $\mathscr{L}$, yielding the covariant differential $D = d - id\zeta$ with components $D_\mu = \partial_\mu - i\partial_\mu\zeta$. Writing $D\!\!\!\!/ = \gamma^\mu D_\mu$, the balanced Lagrangian $\mathscr{L}' = \bar{\psi}_\zeta(iD\!\!\!\!/ - m)\psi_\zeta$ will be equal to $\mathscr{L}$ for all $\zeta$. Another way of seeing that differentiation has to be balanced by $d\zeta$ to offset (2): The momentum operator $P$ becomes $-id$ in the position representation; applied to $\psi_\zeta$ it gives $-id\psi_\zeta = e^{i\zeta}(-id + d\zeta)\psi_\zeta$, in other words $UPU^\dagger U\psi = P_\zeta\psi_\zeta$, the position representation of the rotated momentum operator $P_\zeta$ being $-id + d\zeta$.[3]

It is then argued that an interaction $F = dA = d^2\zeta$ is thereby deduced,[4] whose potential $A$ is $d\zeta$. But since $d^2$ vanishes the interaction does too, as has often been

[2]Yang & Mills (1954) p. 192: "It seems that this [(1) but with $\mathbb{SU}(2)$ instead of $\mathbb{U}(1)$] is not consistent with the localized field concept that underlies the usual physical theories."

[3]My analysis owes much to Lyre (2001, 2002, 2004a,b). But

$$\langle\varphi|P|\varphi\rangle = \langle\varphi U|UPU^\dagger|U\varphi\rangle \neq \langle\varphi U|P|U\varphi\rangle$$

seems relevant to his claim (2004b, pp. 649-51) that local phase transformations are not observable. I would say they *are*—unless one compensates to restore invariance. P. 651 he writes that: "local phase transformations are already unmasked as *not* observable. From this insight, however, the whole logic of the received view breaks down. Since the introduction of an interaction field as intended by the received view seemingly changes physics (those fields are even directly observable themselves), it is necesary from this view to consider local gauge transformations as changing physics as well in order to tell the story about compensation. Since, however, local gauge transformations can be shown as not observable, the received view proves itself untenable." It is untenable because the added term $d\zeta$ is exact. But even if $d\zeta$ is *electromagnetically* unobservable, it is *quantum-mechanically* observable: $\langle\varphi|P|\varphi\rangle \neq \langle\varphi|P_\zeta|\varphi\rangle$.

[4]Ryder (1996) p. 95: "the electromagnetic field arises *naturally* by demanding invariance of the action [...] under *local* ($x$-dependent) rotations [...]."

pointed out.[5]

The gauge argument is fertile enough to produce another two Lagrangians,[6]

$$\mathscr{L}_A = j \wedge A = j^\mu A_\mu = \bar{\psi}\gamma^\mu A_\mu \psi \quad \text{and} \quad \mathscr{L}_F = F \wedge *F = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu},$$

where the current density three-form

$$j = \frac{1}{3!} \varepsilon_{\mu\nu\sigma\tau} j^\mu dx^\nu \wedge dx^\sigma \wedge dx^\tau$$

corresponds to the vector with components $j^\mu = \bar{\psi}\gamma^\mu \psi$. One can either leave $A = d\zeta$ in $\mathscr{L}'$ to offset (2), or balance $\mathscr{L}_\zeta$ with $\mathscr{L}_A$ in the sum $\mathscr{L}' = \mathscr{L}_\zeta + \mathscr{L}_A$. Again, a Lagrangian $\mathscr{L}_F$ derived from the gauge argument will vanish. But once the gauge argument has produced the exact potential $A = d\zeta$ and vanishing interaction $F = dA = d^2\zeta$ one can perhaps claim that $A$ is no longer exact.[7] The exact term $d\zeta$ would then be subtracted from one that isn't[8] in the gauge transformation

$$A \mapsto A' = A - d\zeta. \tag{3}$$

The total Lagrangian $\mathscr{L}' + \mathscr{L}_F$ is indifferent to (2) and (3).

## 2 Global and local gauge transformations

Let us return to the global transformation (1), which adds the same angle $\kappa$ everywhere on $M$. To do so one has to know where to start, there has to be an identity $\mathbb{1} = e^{i0} \in \mathbb{U}(1)$ *everywhere*, a *global* identity $\mathscr{I}$, of the gauge group $\mathscr{G}$. The structure group $G = \mathbb{U}(1)$ is there to act on the typical fiber, which here is $\mathbb{C}^4$; at a point $x \in M$, the identity $\mathbb{1}$ is the element that leaves any $\psi_x \in \mathbb{C}^4_x$ unaltered; the global identity $\mathscr{I} \in \mathscr{G}$ would leave a *global section* unaltered. But to leave a global section unaltered there has to be one in the first place. Since global sections do not exist in general,[9] the global gauge transformation (1) doesn't either, so the gauge argument can start with (2): "for the Lagrangian to remain invariant, the transformation (2) has to be balanced

[5] Auyang (1995, p. 58), Brown (1999, pp. 50-3), Teller (2000, pp. S468-9), Lyre (2001, 2002, 2004a,b), Healey (2001, p. 438), Martin (2002, p. S229), Martin (2003, p. 45), Catren (2008, pp. 512, 520). But the general *structure* of the covariant derivative is about right; Lyre (2002, p. 84): "Denn wenngleich das Eichprinzip [...] nicht zwingend auf nichtflache Konnektionen führt, so ist ja doch die in der kovarianten Ableitung vorgegebene Struktur des Wechselwirkungterms auch für den empirisch bedeutsamen Fall nicht-verschwindender Feldstärken korrekt beschrieben. Diese *Wechselwirkungsstruktur* is also tatsächlich aus der lokalen Eichsymmetrie-Forderung hergeleitet."

[6] *Cf.* Weyl (1929b, p. 283).

[7] Weyl (*ibid.*, p. 283) simply *provides* the inexact electromagnetic potential $\varphi$. He does not use an argument to produce an exact potential, which then becomes inexact. What he does on p. 291 (and on p. 348 of 1929a) is less clear; I would say that his argument only really justifies the exact term $d\lambda$, and that he adds the electromagnetic potential $\varphi$ ($f$ in 1929a) by hand.

[8] One can wonder what the gauge argument is for if the inexact potential $A$ was already there to begin with. The exact term subtracted in (3) has more to do with the invariance of $F = dA = dA'$ than with the gauge argument.

[9] See Göckeler & Schücker (1987, §9.7) for instance. The matter is of course topological—a simply-connected base manifold $M$ admits global sections.

by the exact term $d\zeta$ yielding the vanishing interaction $d^2\zeta$" and so on. The argument remains contrived and unconvincing,[10] but at least the exertions needed to reach (2) from (1) are spared. The local transformation (2) may (or may not) require justification in itself, regardless of context—but much less at any rate than if it is preceded by (1), which appears to represent a 'harmless' transformation from which effort is needed to reach and justify the 'troublesome' transformation (2). If (1) doesn't exist, (2) is neither troublesome nor harmless but all there is.

Of course (1) would—if available—be the natural place to start, as it corresponds to the only normal operator $N$ such that $\mathscr{L}(\psi) = \mathscr{L}(N\psi)$ without tinkering. The next most harmless normal operator is the unitary operator, whose action[11] is given by (2); the progression from (1) to (2) would therefore be entirely natural. But now that we're no longer starting with (1), (2) is no longer the natural successor of an equally natural initial transformation, so why start with (2)? Because the only alternative is the transformation $\psi \mapsto z\psi$ ($z \in \mathbb{C}$) corresponding to the most general normal operator—from which it is too hard to salvage the Lagrangian

$$i\bar{z}\bar{\psi}\gamma^\mu(z\partial_\mu + \partial_\mu z)\psi + m|z|^2\bar{\psi}\psi.$$

# References


Aitchison, I. J. R. and A. J. G. Hey (1982) *Gauge theories in particle physics*, Hilger, Bristol

Auyang, S. Y. (1995) *How is quantum field theory possible?*, Oxford University Press

Brown, H. (1999) "Aspects of objectivity in quantum mechanics" pp. 45-70 in J. Butterfield and C. Pagonis (editors) *From physics to philosophy*, Cambridge University Press

Catren, G. (2008) "Geometric foundations of classical Yang-Mills theory" *Studies in History and Philosophy of Modern Phyiscs* **39**, 511-31

Göckeler, M. and T. Schücker (1987) *Differential geometry, gauge theories, and gravity*, Cambridge University Press

Healey, R. (2001) "On the reality of gauge potentials" *Philosophy of Science* **68**, 432-55

Lyre, H. (2001) "The principles of gauging" *Philosophy of Science* **68**, S371-81

Lyre, H. (2002) "Zur Wissenschaftstheorie moderner Eichfeldtheorien" in A. Beckermann and C. Nimtz (editors) *Argument & Analyse – Sektionsvorträge*, Mentis, Paderborn


[10] In fact to get the gauge argument to work (and produce a nonvanishing interaction) it is enough to make the variation of the phase *path-dependent*.

[11] In its 'diagonal' eigenrepresentation, in which the kernel becomes the multiplication operator $e^{i\zeta}$, with just one variable.

Lyre, H. (2004a) *Lokale Symmetrien und Wirklichkeit: eine Naturphilosophische Studie über Eichtheorien und Strukturenrealismus*, Mentis, Paderborn

Lyre, H. (2004b) "Holism and structuralism in $U(1)$ gauge theory" *Studies in History and Philosophy of Modern Physics* **35**, 643-70

Mandl, F. and G. Shaw (1984) *Quantum field theory*, Wiley, Chichester

Martin, C. (2002) "Gauge principles, gauge arguments and the logic of nature" *Philosophy of Science* **69**, S221-34

Martin, C. (2003) "On continuous symmetries and the foundations of modern physics" pp. 29-60 in K. Brading and E. Castellani (editors) *Symmetries in physics: Philosophical reflections*, Cambridge University Press

O'Raifeartaigh, L. (1997) *The dawning of gauge theory*, Princeton University Press

Ramond, P. (1990) *Field theory: a modern primer*, Westview Press, Boulder

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

Sakurai, J. J. (1967) *Advanced quantum mechanics*, Addison-Wesley, Reading

Teller, P. (2000) "The gauge argument" *Philosophy of Science* **67**, S466-81

Weyl, H. (1929a) "Elektron und Gravitation" *Zeitschrift für Physik* **56**, 330-52

Weyl, H. (1929b) "Gravitation and the electron" *The Rice Institute Pamphlet* **16**, 280-95

Yang, C. N. and R. Mills (1954) "Conservation of isotopic spin and isotopic gauge invariance"