Weyl's gauge argument

Alexander Afriat

Weyl's gauge argument

Alexander Afriat

January 9, 2013

Abstract

The standard $\mathbb{U}(1)$ "gauge principle" or "gauge argument" produces an exact potential $A = d\lambda$ and a vanishing field $F = d^2\lambda = 0$. Weyl (1929b,c) has his own gauge argument, which is sketchy, archaic and hard to follow; but at least it produces an inexact potential A and a nonvanishing field $F = dA \neq 0$. I attempt a reconstruction.

1 Introduction

Weyl's first gauge theory (1918) was a generalisation of Einstein's general relativity; his second gauge theory,[1] which grew out of the first, remained a relativistic theory of curved spacetime, but with a matter field of two-spinors subject to Weyl's version of the Dirac equation. The proper orthochronous Lorentz group $\mathbb{SO}^+(1,3)$ changes neither the length, origin, spatial parity, nor temporal orientation of spacetime four-vectors, whose parallel propagation is accordingly governed in Weyl's second theory by a connection $\mathcal{A} = \mathcal{A}^a_\mu dx^\mu \otimes \mathbf{T}_a$ with values in the Lie algebra $\mathfrak{o}(1,3) = \mathrm{Lie}\,\mathbb{SO}^+(1,3)$. The parallel transport of Weyl's two-spinors, which are subject to a group we can call[2]

$$\mathbb{W}(2,\mathbb{C}) = \{g \in \mathbb{GL}(2,\mathbb{C}) : |\mathrm{det} g| = 1\},$$

is given by a connection $\mathfrak{A}$ with values in $\mathfrak{w}(2,\mathbb{C}) = \mathrm{Lie}\,\mathbb{W}(2,\mathbb{C})$; the homomorphism

$$h : \mathbb{W}(2,\mathbb{C}) \to \mathbb{SO}^+(1,3)$$

is therefore at the core of Weyl's theory. We'll see how he exploits the angular freedom $e^{i\lambda}$ left by h for "the critical part of the theory":[3] *the derivation of electromagnetism*.

The standard "gauge principle" or "gauge argument" is sometimes attributed to Weyl.[4] Not only is his argument quite different, but it avoids the exact connection $A = d\lambda$ and vanishing field $F = d^2\lambda$ that vitiate the standard argument.

[1] Weyl (1929a,b,c). See Straumann (1987), O'Raifeartaigh (1997), O'Raifeartaigh & Straumann (2000), Brading (2002) and Scholz (2005) for more recent accounts.

[2] It is sometimes known as $\mathrm{Spin}^{\mathbb{C}}$; see Friedrich (2000) §1.6 for instance.

[3] Weyl (1929b) p. 348: "§6. Elektrisches Feld. Wir kommen jetzt zu dem kritischen Teil der Theorie. Meiner Meinung nach liegt der Ursprung und die Notwendigkeit des elektromagnetischen Feldes in folgendem begründet."

[4] Brading (2002) pp. 3-4, Healey (2007) p. 160 for instance.

2 The standard gauge argument

One begins with a free field, of two-spinors $\psi \in \mathbb{C}^2$ for instance. The Lagrangian $\mathscr{L} = \bar{\psi}\not{\partial}\psi$ is invariant under the *global* transformation

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

where "global" means that κ is the same everywhere; $\not{\partial}$ stands for the sum $\sigma^\mu\partial_\mu$, in which σ^0 is the identity and σ^k the three Pauli operators. It is then argued[5] that $\mathscr{L}$ should also be invariant under the *local* transformation

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

where $\lambda : M \to \mathbb{R}$ 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 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,[6] 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)

[5] 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 φ at one point, through an angle Λ, 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 Λ 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.

[6] 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."

At any rate, as things stand the Lagrangian

$$\mathscr{L}_\lambda = \bar\psi_\lambda \partial\!\!\!/ \psi_\lambda = \bar\psi e^{-i\lambda}\sigma^\mu\partial_\mu(e^{i\lambda}\psi) = \bar\psi\sigma^\mu(\partial_\mu + i\partial_\mu\lambda)\psi$$

is not invariant since the derivative ∂_μ has become $\partial_\mu + i\partial_\mu\lambda$. To offset (2) we therefore have to subtract the term $i\partial_\mu\lambda$ that alters $\mathscr{L}$, yielding the covariant differential $D = d - id\lambda$ with components $D_\mu = \partial_\mu - i\partial_\mu\lambda$. Writing $D\!\!\!\!/ = \sigma^\mu D_\mu$, the balanced Lagrangian $\mathscr{L}' = \bar\psi_\lambda D\!\!\!\!/ \psi_\lambda$ will be equal to $\mathscr{L}$ for all λ. Another way of seeing that differentiation has to be balanced by $d\lambda$ to offset (2): The momentum operator P becomes $-id$ in the position representation; applied to ψ_λ it gives $-id\psi_\lambda = e^{i\lambda}(-id + d\lambda)\psi_\lambda$, in other words $UPU^\dagger U\psi = P_\lambda\psi_\lambda$, the position representation of the rotated momentum operator P_λ being $-id + d\lambda$.[7]

It is then argued that an interaction $F = dA = d^2\lambda$ is thereby deduced,[8] whose potential A is $d\lambda$. But since d^2 vanishes the interaction does too, as has often been pointed out.[9]

The gauge argument is fertile enough to produce another two Lagrangians,[10]

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

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 $j^\mu = \bar\psi\sigma^\mu\psi$. One can either leave $A = d\lambda$ in $\mathscr{L}'$ to offset (2), or balance $\mathscr{L}_\lambda$ with $\mathscr{L}_A$ in the sum $\mathscr{L}' = \mathscr{L}_\lambda + \mathscr{L}_A$. Again, a Lagrangian $\mathscr{L}_F$ derived from the gauge argument will vanish. But once the argument has produced the exact potential $A = d\lambda$ and vanishing interaction $F = dA = d^2\lambda$

---

[7]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 $d\lambda$ is exact. But even if $d\lambda$ is *electromagnetically* unobservable, it is *quantum-mechanically* observable: $\langle\varphi|P|\varphi\rangle \neq \langle\varphi|P_\lambda|\varphi\rangle$.

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

[9]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."

[10]*Cf.* Weyl (1929c) p. 283.

one can perhaps claim that A is no longer exact. The exact term $d\lambda$ would then be subtracted from one that isn't[11] in the gauge transformation

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

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

3 Weyl's argument

As pointed out in the Introduction, Weyl's two-spinors are subject to a group[12] $\mathbb{W}(2,\mathbb{C})$ slightly larger than $\mathbb{SL}(2,\mathbb{C}) = \{g \in \mathbb{GL}(2,\mathbb{C}) : \det g = 1\}$. To produce electromagnetism Weyl uses the $\mathbb{U}(1)$ freedom expressed by

$$h(e^{i\lambda} g) = h(g) \in \mathbb{SO}^+(1,3), \tag{4}$$

$g \in \mathbb{W}(2,\mathbb{C})$.[13]

$\mathbb{SO}^+(1,3) = G$ and $\mathbb{W}(2,\mathbb{C}) = G'$ are just 'structure' groups, acting at a generic spacetime point. What about the corresponding gauge groups $\mathscr{G}$, $\mathscr{G}'$ acting on all of spacetime M? In special relativity "there's just a single tetrad"; so there's just one $\mathbb{SO}^+(1,3) = G = \mathscr{G}$, one $\mathbb{W}(2,\mathbb{C}) = G' = \mathscr{G}'$, and above all one $e^{i\lambda}$.[14] But with spacetime curvature the tetrad varies,[15] and so does λ. This could mean the following:[16] Only a *flat* $\mathfrak{o}(1,3)$-valued connection $\mathcal{A}$ allows the assignment of the *same* tetrad to distant points—only with flatness can there be *global* constancy or 'sameness.' With curvature it becomes meaningless to say that tetrads at distant points are the same.

[11] 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.

[12] Weyl (1929b) p. 333: "man beschränke sich auf solche lineare Transformationen U von ψ_1, ψ_2, deren Determinante den absoluten Betrag 1 hat."

[13] Weyl (1929c) p. 291: "It is my firm conviction that we must seek the origin of the electromagnetic field in another direction. We have already mentioned that it is impossible to connect the transformations of the ψ in a unique manner with the rotations of the axis system; however we may attempt to accomplish this by means of invariants which can be used as constituents of an action quantity we always find that there remains an arbitrary "gauge factor" $e^{i\lambda}$. Hence the local axis-system does not determine the components of ψ uniquely, but only within such a factor of absolute magnitude 1." Weyl (1931) p. 195: "Aus der Natur, dem Transformationsgesetz der Größe ψ ergibt sich, daß die vier Komponenten ψ_ϱ relativ zum lokalen Achsenkreuz nur bis auf einen gemeinsamen Proportionalitätsfaktor $e^{i\lambda}$ durch den physikalischen Zustand bestimmt sind, dessen Exponent λ willkürlich vom Orte in Raum und Zeit abhängt, und daß infolgedessen zur eindeutigen Festlegung des kovarianten Differentials von ψ eine Linearform $\sum_\alpha f_\alpha dx_\alpha$ erforderlich ist, die so mit dem Eichfaktor in ψ gekoppelt ist, wie es das Prinzip der Eichinvarianz verlangt."

[14] Weyl (1929b) p. 348: "In der speziellen Relativitätstheorie muß man diesen Eichfaktor als eine Konstante ansehen, weil wir hier ein einziges, nicht an einen Punkt gebundes Achsenkreuz haben." Weyl (1929c) p. 291: "In the special theory of relativity, in which the axis system is not tied up to any particular point, this factor is a constant."

[15] The gauge groups become infinite-dimensional. Weyl (1929b) p. 348: "Anders in der allgemeinen Relativitätstheorie: jeder Punkt hat sein eigenes Achsenkreuz und darum auch seinen eigenen willkürlichen Eichfaktor; dadurch, daß man die starre Bindung der Achsenkreuze in verschiedenen Punkten aufhebt, wird der Eichfaktor notwendig zu einer willkürlichen Ortsfunktion." Weyl (1929c) p. 291: "But it is otherwise in the general theory of relativity when we remove the restriction binding the local axis-systems to each other; we cannot avoid allowing the gauge factor to depend arbitrarily on position."

[16] Here I am indebted to Johannes Huisman.

Where tetrads cannot remain constant, one has to suppose they *vary*. A flat real-valued phase connection A (see §4) alongside a curved $\mathcal{A}$ can of course be countenanced, but it is in the spirit of Weyl's argument for both to be flat or both curved. So if the tetrad varies, λ might as well too.

The group homomorphism h determines the Lie algebra homomorphism

$$\mathfrak{h} : \mathfrak{w}(2, \mathbb{C}) \to \mathfrak{o}(1, 3),$$

the Lie algebra $\mathfrak{w}(2, \mathbb{C})$ being the direct sum $\mathfrak{sl}(2, \mathbb{C}) \oplus i\mathbb{R}\mathbb{1}_2$, where $i\mathbb{R} = \mathrm{Lie}\,\mathbb{U}(1)$. Doing away with the additive freedom λ (or rather $i\lambda\mathbb{1}_2$) we're left with the isomorphism between $\mathfrak{w}(2, \mathbb{C})/i\mathbb{R}\mathbb{1}_2 = \mathfrak{sl}(2, \mathbb{C})$ and $\mathfrak{o}(1, 3)$. Instead of the phase $e^{i\lambda} \in \mathbb{U}(1)$ we have $i\lambda\mathbb{1}_2 \in i\mathbb{R}\mathbb{1}_2$; instead of $\mathbb{U}(1)$ we have the Lie algebra $i\mathbb{R}\mathbb{1}_2$; and instead of (4),

$$\mathfrak{h}(\gamma \oplus i\lambda\mathbb{1}_2) = \mathfrak{h}(\gamma) \in \mathfrak{o}(1, 3),$$

$\gamma \in \mathfrak{w}(2, \mathbb{C})$.[17]

4 Anholonomy

The additive freedom $i\lambda\mathbb{1}_2$ is in the Lie algebra $\mathfrak{w}(2, \mathbb{C})$ where the spin connection $\mathfrak{A}$ has its values; and connections are there to generate parallel transport—*in a direction*.[18] A direction $V \in T_xM$ will therefore characterise the propagation of λ, whose infinitesimal variation $\delta\lambda$ has to be linear in λ and in V. The object needed is a one-form; applied to the direction V it yields the infinitesimal generator $\langle A, V\rangle \in \mathbb{R}$, which then multiplies λ to produce the increment $\delta\lambda = \lambda\langle A, V\rangle$. So there's a connection for tetrads, another for spinors, *and a third one—A—for the residual* $\mathbb{U}(1)$ *freedom caught 'in between' tetrads and spinors*.

The whole point of allowing the propagation of λ to depend on direction is to admit anholonomies. So the curvature

$$F = \frac{1}{2}F_{\mu\nu}dx^\mu \wedge dx^\nu = dA = \frac{1}{2}(\partial_\mu A_\nu - \partial_\nu A_\mu)dx^\mu \wedge dx^\nu$$

[17] Weyl (1929b, p. 348): "Dann ist aber auch die infinitesimale lineare Transformation dE der ψ, welche der infinitesimalen Drehung $d\gamma$ entspricht, nicht vollständig festgelegt, sondern dE kann um ein beliebiges rein imaginäres Multiplum $i \cdot df$ der Einheitsmatrix vermehrt werden." Weyl (1929c, p. 291): "Then there remains in the infinitesimal linear transformation dE of ψ, which corresponds to the given infinitesimal rotation of the axis-system, an arbitrary additive term $+id\varphi \cdot 1$."

[18] Weyl (1929b, p. 348): "Zur eindeutigen Festlegung des kovarianten Differentials $\delta\psi$ von ψ hat man also außer der Metrik in der Umgebung des Punktes P auch ein solches df für jedes von P ausgehende Linienelement $\overrightarrow{PP'} = (dx)$ nötig. Damit $\delta\psi$ nach wie vor linear von dx abhängt, muß

$$df = f_p(dx)^p$$

eine Linearform in den Komponenten des Linienelements sein. Ersetzt man ψ durch $e^{i\lambda}$, so muß man sogleich, wie aus der Formel für das kovariante Differential hervorgeht, df ersetzen durch $df - d\lambda$." Weyl (1929c, p. 291): "The complete determination of the covariant differential $\delta\psi$ of ψ requires that such a $d\varphi$ be given. But it must depend linearly on the displacement PP': $d\varphi = \varphi_p(dx)^p$, if $\delta\psi$ shall depend linearly on the displacement. On altering ψ by multiplying it by the gauge factor $e^{i\lambda}$ we must at the same time replace $d\varphi$ by $d\varphi - d\lambda$ as is immediately seen from this formula of the covariant differential." Weyl's notation is confusing: whereas the one-form $d\lambda$ (which *is* a differential) is necessarily exact, *df* and $d\varphi$ (my A) aren't.

of A will not necessarily vanish; and since F is exact, it is also closed:

$$dF = d^2 A = \frac{1}{6}(\partial_\mu F_{\nu\sigma} + \partial_\nu F_{\sigma\mu} + \partial_\sigma F_{\mu\nu})\, dx^\mu \wedge dx^\nu \wedge dx^\sigma = 0.$$

In F, A and $dF = 0$ Weyl saw[19] the electromagnetic field, its potential and Maxwell's two homogeneous equations[20] (which are the same—up to Hodge duality—as the other two, away from sources).

5 Final remark

Whatever its idiosyncrasies, Weyl's gauge argument at least avoids the exact connection $A = d\lambda$ and vanishing curvature $F = d^2\lambda = 0$ produced by the standard argument.

I thank Ermenegildo Caccese, Johannes Huisman, Thierry Levasseur and Jean-Philippe Nicolas for many valuable conversations and clarifications.

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

Brading, K. (2002) "Which symmetry? Noether, Weyl, and the conservation of electric charge" *Studies in History and Philosophy of Modern Physics* **33**, 3-22

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

Friedrich, T. (2000) *Dirac operators in Riemannian geometry*, American Mathematical Society, Providence

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

Healey, R. (2007) *Gauging what's real: the conceptual foundations of contemporary gauge theories*, Oxford University Press, New York

[19] Weyl (1929b) p. 349, Weyl (1929c) pp. 291-2

[20] $\nabla \cdot \mathbf{B} = 0$ and $\nabla \times \mathbf{E} + \partial_t \mathbf{B} = 0$

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

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

O’Raifeartaigh, L. and N. Straumann (2000) “Gauge theory: historical origins and some modern developments”

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

Scholz, E. (2005) “Local spinor structures in V. Fock’s and H. Weyl’s work on the Dirac equation (1929)” pp. 284-301 in D. Flament *et al*. (editors) *Géométrie au vingtième siècle, 1930-2000*, Hermann, Paris

Straumann, N. (1987) “Zum Ursprung der Eichtheorien bei Hermann Weyl” *Physikalische Blätter* **43**, 414-21

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

Weyl, H. (1918) “Gravitation und Elektrizität” pp. 147-59 dans *Das Relativitätsprinzip*, Teubner, Stuttgart, 1990

Weyl, H. (1929a) “Gravitation and the electron” *Proceedings of the National academy of sciences, USA* **15**, 323-34

Weyl, H. (1929b) “Elektron und Gravitation” *Zeitschrift für Physik* **56**, 330-52

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

Weyl, H. (1931) *Gruppentheorie und Quantenmechanik*, Hirzel, Leipzig

Yang, C. N. and R. Mills (1954) “Conservation of isotopic spin and isotopic gauge invariance” *Physical review* **96**, 191-5