IASSNS-HEP-97/142

Riemann-Einstein Structure from Volume and Gauge Symmetry

Frank Wilczek
Institute for Advanced Study, School of Natural Science, Olden Lane, Princeton, NJ 08540

Abstract

It is shown how a metric structure can be induced in a simple way starting with a gauge structure and a preferred volume, by spontaneous symmetry breaking. A polynomial action, including coupling to matter, is constructed for the symmetric phase. It is argued that assuming a preferred volume, in the context of a metric theory, induces only a limited modification of the theory.

§1 Attempts to Use a Gauge Principle for Gravity

The fundamental laws of physics, as presently understood, largely follow from the two powerful symmetry principles: general covariance, and gauge invariance. There have been many attempts to unite these two principles. Indeed the very term, gauge invariance, originates from an early attempt [1] by Weyl to derive classical electromagnetism as a consequence of space-time symmetry, specifically symmetry under local changes of length scale. The modern understanding of gauge invariance, as a symmetry under transformations of quantum-mechanical wave functions, was reached by Weyl himself and also by London very shortly after the new quantum mechanics was first proposed. In this understanding of abelian gauge invariance, and in its nonabelian generalization [2], the space-time aspect is lost. The gauge transformations act only on internal variables. This formulation has had great practical success. Still, it is not entirely satisfactory to have two closely related, yet definitely distinct, fundamental principles, and several physicists have proposed ways to unite them.

One line of thought, beginning with Kaluza [3] and Klein [4], seeks to submerge gauge symmetry into general covariance. Its leading idea is that gauge symmetry arises as a reflection in the four familiar macroscopic space-time dimensions of general covariance in a larger number of dimensions, several of which are postulated to be small, presumably for dynamical reasons. Here we should take the opportunity to emphasize a point that is somewhat confused by the historically standard usages, but which it is vital to have clear for what follows. When physicists refer to general covariance, they usually mean the form-invariance of physical laws under coordinate transformations following the usual laws of tensor calculus, including the transformation of a given, preferred metric tensor. Without a metric tensor, one cannot form an action principle in the normal way, nor in particular formulate the accepted fundamental laws of physics, viz. general relativity and the Standard Model. From a purely mathematical point of view one might consider doing without the metric tensor; in that case general covariance becomes essentially the same concept as topological invariance. The existence of a metric tensor reduces the genuine symmetry to a much smaller one, in which space-times are required not merely to be topologically the same, but congruent (isometric), in order to be considered equivalent. In the Kaluza-Klein construction, for this reason, the gauge symmetries arise only from *isometries* of the compactified dimensions.

Another line of thought proceeds in the opposite direction, seeking to realize general covariance – in the metric sense – as a gauge symmetry. A formal parallelism is readily perceived [5]. Indeed, it is quite helpful, in bringing spinors to curved space, to introduce a local gauge $SO(3,1)$ symmetry [1]. The spin connection ω_μ^{ab} is the gauge potential of this symmetry. The lower index is a space-time index, and the upper indices are internal $SO(3,1)$ indices, in which ω is antisymmetric, as is appropriate for the gauge potential. In this construction one must, however, introduce an additional element that is not a feature of conventional gauge theories. That is, one postulates in addition the existence of a set of vierbein fields e_α^a with the properties

$$\eta_{ab} e_\alpha^a e_\beta^b = g_{\alpha\beta} \tag{1}$$

$$g^{\alpha\beta} e_\alpha^a e_\beta^b = \eta^{ab} \tag{2}$$

where η is the internal space metric tensor. The e_α^a are supposed to transform homogeneously under the local symmetry, and to behave as ordinary vectors under general coordinate transformations. The e_α^a link the internal and space-time indices, so that by contracting with them one can freely trade one kind of index for the other.

The formal nature of this local $SO(3,1)$ symmetry is evident from (1). The action of the symmetry simply interpolates between the different possible choices for the solution of this equation, all corresponding to the same space-time metric. Thus given the metric the vierbein adds no gauge-invariant information, at least locally. One could equally well have both a vierbein field and the gauge symmetry which effectively says it is arbitrary, or neither. For coupling to spinors, of course, it is very convenient to make the former choice. Another difference between the formal $SO(3,1)$ gauge invariance of general relativity and the usual gauge symmetries of the Standard Model is the postulated action. In general relativity the Einstein-Hilbert action density, in terms of the $SO(3,1)$ gauge field strength and the vierbeins, is given as

$$\mathcal{L} = \det e \; e_a^\alpha e_b^\beta F_{\alpha\beta}^{ab} \tag{3}$$

where the determinant involves e_α^a regarded as a square matrix, and the e_a^α are elements of the inverse matrix. The existence of an inverse to the introduced e_α^a is essential for expressing the action in this form. In this action the ω appear, after an integration by parts, only algebraically. They can therefore be eliminated in favor of the vierbeins, or ultimately the metric. All this is quite unlike what we have for conventional Yang-Mills theory.

A markedly different approach to casting gravity as a gauge theory was initiated by MacDowell and Mansouri [6], and independently Chamseddine and West [7]. It has proved useful as a way of introducing supergravity. Their construction, as improved by Stelle

[1]For definiteness I consider 3+1 dimensional space here.

and West [8], is along the following lines. One introduces an $SO(3,2)$ gauge symmetry (or $SO(1,4)$; $SO(3,2)$ is more appropriate for supergravity), with the associated gauge potentials A_μ^{ab}. One introduces as well a space-time scalar, internal-space vector field ϕ^a. Very important: no vierbein or metric is introduced. The action density is taken to be

$$\mathcal{L}_{\rm MM} \ = \ \epsilon_{abcde}\epsilon^{\alpha\beta\gamma\delta}F_{\alpha\beta}^{ab}F_{\gamma\delta}^{cd}\phi^e \tag{4}$$

and ϕ^e is taken to be frozen at $\phi^e = \kappa\delta_5^e$. Here of course the ϵs are numerical totally antisymmetric symbols, and κ is a constant. At this point one must distinguish two classes of gauge potentials, the $A_\alpha^{a5} \equiv h_\alpha^a$ and the A_α^{ab} with neither a nor b equal to 5. The field strengths F^{a5} do not appear in (4). The h_α^a appear in the remaining field strengths F^{ab} only in commutator terms. Let us separate the terms which are independent of, linear, or quadratic in these commutators. The term independent of the h_α^a is quadratic in the conventional $SO(3,1)$ potentials and is a total divergence. It does not contribute to the classical equations of motion. The term which is linear in the commutators – and therefore also linear in the conventional $SO(3,1)$ field strength – is most remarkable. Indeed, if we assume that the matrix h_α^a is non-singular, and make the identifications

$$\omega_\mu^{ab} \equiv A_\mu^{ab} \tag{5}$$

$$e_\alpha^a \equiv h_\alpha^a = A_\alpha^{a5} \tag{6}$$

we find that this term is proportional to the Einstein-Hilbert action density. Finally, in this interpretation, the term quadratic in the commutators yields a cosmological constant.

This construction certainly makes a more compelling connection of gravity to gauge theory than the preceding formal construction, and has the elegance of simplicity. Yet it is not entirely satisfactory, for several reasons. The assumed constancy of ϕ^e is quite *ad hoc*, and amounts to a way of sneaking in some assumptions about a higher symmetry, $SO(3,2)$ or $SO(1,4)$, that is not really present in the theory. Likewise, the expansion around non-singular h_α^a seems arbitrary, especially since these fields do not transform homogeneously under the higher symmetry (they do transform homogeneously under the residual $SO(3,1)$). Perhaps its most important limitation, however, is that it does not extend in any obvious way to include couplings to matter. In the remainder of this note I propose a way to remove these difficulties.

§2 Importance of the Volume Element for Index Contraction

Concretely, the difficulty with coupling to matter is as follows. To obtain a meaningful dynamics, one must include derivatives for the matter fields. These are all lower, or covariant, indices. In the absence of a metric, it is very difficult to soak up these indices and form an invariant action. The only contravariant object that is intrinsically defined is the numerical antisymmetric symbol $\epsilon^{\alpha\beta\gamma\delta}$, and it is a tensor density. Thus – assuming full general covariance – it can be used to the first power only, to soak up exactly four covariant indices. This, of course, is how it functioned in (4). In general, however, and in particular to couple to the fields of the Standard Model, one seems to require a more flexible way of forming invariants. Another aspect of the difficulty is that one cannot form fully covariant derivative-independent terms, such as are used to generate masses or effective potentials. With this in mind, it seems that if one wants to stop short of introducing a metric explicitly, a natural possibility is to forego full general covariance, and to require invariance only under transformations that leave the volume fixed. Then one can introduce potential terms, and have them introduce symmetry breaking dynamically, which opens up much wider possibilities for obtaining realistic theories.

§3 Covariance With a Preferred Volume Element

It might seem that relaxing the requirement of full general covariance would be a very drastic step, and that after taking it one would have great difficulties in recovering the main consequences of general relativity. This is not the case, however, as I shall now discuss. If one takes a conventional (metric) general covariant theory, and considers only variations of the metric that preserve the volume, one obtains in place of the full Einstein equations

$$R^\mu_\nu - \frac{1}{2} g^\mu_\nu R \ = \ T^\mu_\nu \tag{7}$$

only their traceless part

$$R^\mu_\nu - \frac{1}{4} g^\mu_\nu R \ = \ T^\mu_\nu - \frac{1}{4} g^\mu_\nu T \ . \tag{8}$$

It might appear that a great deal has been lost, or modified, in passing from (7) to (8), but that appearance is deceptive. Indeed, (7) has been written in a way that the contracted covariant derivative of each side vanishes separately. For the left-hand side, this follows from the Bianchi identity. On the right-hand side, it follows from *full* general covariance, with T^μ_ν defined according to the equation

$$\frac{\delta S_M}{\delta g_{\mu\nu}} = \sqrt{g} T^{\mu\nu} \tag{9}$$

for variation of the matter action. We define the energy-momentum tensor in (8) likewise, even though in the context of that equation we are not requiring that the action is stationary under full general covariance, after supplying appropriate powers of $\sqrt{g}$ to restore full covariance. Then it too has zero contracted covariant derivative (i.e., it is covariantly conserved). After subtracting $-\frac{1}{4} g^\mu_\nu R$ from both sides of (8) we can take the contracted covariant derivative and cancel off the vanishing terms, to obtain simply

$$\partial_\nu (R + T) \ = \ 0. \tag{10}$$

This means that $R + T$ is a world-constant – a number, not a function. This number is not determined by the gravitational equations of motion, but is an integration constant, or is determined by other equations of motion. In any case, we find that (8) differs from (7) only in that the cosmological term is no longer determined by the action and the equations of motion for the metric tensor, but is an integration constant (perhaps constrained by other equations of motion).

We have argued that solutions of (8) yield solutions of (7) for some value of the cosmological term, itself undetermined by (8). Conversely, if we suppose that (7) is valid, but that the energy-momentum tensor takes the form

$$T^\mu_\nu \;=\; T^\mu_\nu(0) + \lambda g^\mu_\nu. \tag{11}$$

with λ indeterminate, then we can eliminate λ by passing to (8).

It might seem surprising, at first sight, that the variation of an entire function determines only one number. However, of course, general covariance requires that the action is of a very special form, and indeed invariant under functional transformations, namely those of general covariance. Indeed, the situation above could, from another point of view, easily have been anticipated. One could, in the context of orthodox general relativity, take $\sqrt{g} =$ const. (or any fixed function) as a gauge choice. The equations of motion in this fixed gauge must then give the full equations of the theory, except insofar as fixing the gauge constrains gauge-invariant quantities. But the only gauge-invariant quantity associated with $\sqrt{g}$ is a single number, the world-volume. To demonstrate this, one must show the existence of a general coordinate transformation having an arbitrary specified Jacobian; but this is, apart from global questions, evident by counting parameters.

Other discussions of metric covariance with a fixed volume include [9].

§4 Metric Structure from Symmetry Breaking

After these preliminaries, the construction of a metric theory from spontaneous breakdown of gauge symmetry is not difficult. The necessary field content is the same as before: an $SO(3,2)$ gauge field A^{ab}_α, and a spacetime scalar, internal space vector ϕ^a. The desired pattern of symmetry breaking is dynamically favored by action terms of two types. The first term we need is of the type

$$\mathcal{L}_1 \;=\; -\,\kappa_1(\eta_{ab}\phi^a\phi^b - v^2)^2. \tag{12}$$

Clearly if $\kappa_1 > 0$ this will be stationarized, and $-\mathcal{L}_1$ minimized, when $\phi^a(x) = v\delta^a_5$. Other solutions can be put in this form by a gauge transformation. As $\kappa_1 \to \infty$ this value is effectively frozen in.

The second term we need is of the type

$$\mathcal{L}_2 \;=\; \kappa_2(J - w)^2 \tag{13}$$

where

$$J \;\equiv\; \epsilon^{\alpha\beta\gamma\delta}\epsilon_{abcde}\phi^e\nabla_\alpha\phi^a\nabla_\beta\phi^b\nabla_\gamma\phi^c\nabla_\delta\phi^d \;, \tag{14}$$

and ∇ denotes the $SO(3,2)$ gauge covariant derivative. This term is evidently stationarized at $J = w$. If we suppose that ϕ takes the form discussed in the preceding paragraph, then we find

$$J \;=\; v^5\epsilon^{\alpha\beta\gamma\delta}\epsilon_{abcd}A^{a5}_\alpha A^{b5}_\beta A^{c5}_\gamma A^{d5}_\delta \tag{15}$$

where now the internal space variables run from 1 to 4. By stationarizing these two terms, we have reproduced two crucial elements of the structure described in §1, that is, an appropriate symmetry-breaking "director" δ^a_5 and a non-singular "vierbein" $e^a_\alpha \propto A^{a5}_\alpha$, as in (6), but now on a dynamical basis, by spontaneously breaking a legitimate symmetry. The terms we have used are among the simplest ones consistent with the assumed symmetries, that is, $SO(3,2)$ or $SO(1,4)$ and volume-preserving reparametrizations.

§5 Curvature Term

To complete the identification we should discuss the curvature term. In the new framework we can use a simpler construction than (4). The relevant term is

$$\mathcal{L}_3 \;=\; \kappa_3 \epsilon^{\alpha\beta\gamma\delta}\epsilon_{abcde}F^{ab}_{\alpha\beta}\nabla_\gamma\phi^c\nabla_\delta\phi^d\phi^e \;. \tag{16}$$

When this is expanded, subject to the freezing of ϕ as before and of the determinant as in (15), and with the now familiar identifications, it yields the Einstein-Hilbert action. In this form, generalization to any number of dimensions is immediate.

§6 Matter Couplings

The structure here proposed allows coupling to matter fields with no substantial difficulty, but one subtlety. Consider, for example, the problem of constructing a kinetic energy term for a space-time scalar, gauge scalar field ψ. The standard construction uses both the determinant of the metric tensor and the inverse metric tensor. Since we do not require full general covariance, but only covariance under volume-preserving reparametrizations, we do not need the determinant. However we cannot take the inverse metric for granted, for our effective metric only emerges after $\mathcal{L}_1$ is stationarized, and only becomes non-singular when $\mathcal{L}_2$ is stationarized. The appropriate construction is most easily expressed in terms of the auxiliary field

$$w^\alpha_a \;=\; \epsilon^{\alpha\beta\gamma\delta}\epsilon_{abcde}\nabla_\beta\phi^b\nabla_\gamma\phi^c\nabla_\delta\phi^d\phi^e \;. \tag{17}$$

This field becomes proportional to the inverse vierbein after our two steps of symmetry breaking, but is a perfectly well-defined tensor with the same index structure even in the unbroken phase. In terms of it, we can construct the desired kinetic term as

$$\mathcal{L}_{\text{kin.}} \;=\; w^\alpha_a w^\beta_b \eta^{ab}\partial_\alpha\psi\partial_\beta\psi \;. \tag{18}$$

One should note that these terms involve high powers of the covariant derivatives of ϕ.

In a very similar fashion, one can construct kinetic terms for space-time spinor fields. In this case only one power of w is required. Yang-Mills fields require the form

$$\mathcal{L}_{\text{Y-M}} \;=\; G^I_{\alpha\beta}G^I_{\gamma\delta}\eta^{ac}w^\alpha_a w^\gamma_c \eta^{bd}w^\beta_b w^\delta_d \tag{19}$$

or more economically

$$\mathcal{L}'_{\text{Y-M}} \;=\; G^I_{\alpha\beta}G^I_{\gamma\delta}\epsilon^{\alpha\beta\mu\nu}\epsilon^{\gamma\delta\rho\tau}\nabla_\mu\phi^a\nabla_\nu\phi^b\nabla_\rho\phi^c\nabla_\tau\phi^d\eta_{ac}\eta_{bd}. \tag{20}$$

It is very important, from the point of view of general dynamics, that the interactions we need are of no higher than second order in time derivatives. Thus they have a well posed initial value problem. This requirement greatly constrains the form of possible additional terms [10].

§7 On the Cosmological Term

A most puzzling feature of current physical theory is that the vacuum is supposed to contain several symmetry-breaking condensates (e.g., electroweak $SU(2) \times U(1)$ breaking and QCD chiral $SU(2) \times SU(2)$ breaking), yet for purposes of gravity displays very small or vanishing energy density. This is the famous problem of the cosmological term.

In the framework discussed here, this problem takes a very different appearance. We have already touched on this in a general way in §3. Now we can illustrate concretely how extra (non-metric) field equations can augment that discussion.

For present purposes it is useful to introduce symbols σ and A for the vacuum expectation values of ϕ^5 and of $(\det A_\alpha^{5a})^{\frac{1}{4}}$ – the "determinantal" part of the vierbein – respectively. Note that one can swing the expectation value ϕ to lie in a fixed direction, say 5, by appropriate gauge transformations.

Let us assume that σ and A are constants, and examine the dependence of various possible terms in the Lagrangian on these constants. We have for the 'gravitational' terms

$$\begin{array}{lll}
\lambda_1(\phi^a\phi^b\eta_{ab})^2 & \sim \lambda_1\sigma^4 & ;[\lambda_1] = \ \ 4 \\
\lambda_2(\phi^a\phi^b\eta_{ab}) & \sim \lambda_2\sigma^2 & ;[\lambda_2] = \ \ 4 \\
\lambda_3 J^2 & \sim \lambda_3\sigma^{10}A^8 & ;[\lambda_3] = -4 \\
\lambda_4 J & \sim \lambda_4\sigma^5 A^4 & ;[\lambda_4] = \ \ 0 \\
\lambda_5\epsilon^{\alpha\beta\gamma\delta}\epsilon_{\alpha\beta\gamma\delta}F^{ab}_{\alpha\beta}\nabla_\lambda\phi^c\nabla_\delta\phi^d\phi^e & \sim \lambda_5\sigma^3A^2\partial^2 \text{ or } \lambda_5\sigma^3A^4 & ;[\lambda_5] = \ \ 0
\end{array}$$

Here σ is taken to have mass dimension 0, A to have mass dimension 1, and the mass dimensions of the couplings are then as indicated.

In the last line the two terms arising from the spin connection A_α^{ab} and vierbein A_α^{5a} pieces of F are indicated separately. Also indicated is the occurrence of derivatives, when one solves for the A_α^{ab}. The first of these terms corresponds to the usual Einstein action, and leads us to identify the Planck mass according to

$$M^2_{\text{Pl.}} \sim \lambda_5\sigma^3A^2. \tag{21}$$

The second of these terms is an effective cosmological term. Its coefficient is given as

$$\lambda_5\sigma^3A^4 \sim M^4_{\text{Pl.}}/(\lambda_5\sigma^3). \tag{22}$$

The terms in J depend on the combination

$$\sigma^5A^4 \sim M^4_{\text{Pl.}}/(\lambda_5^2\sigma). \tag{23}$$

There is no A-dependence from the remaining terms, nor (we recall) from conventional condensation energy.

Thus if the pure numbers λ_5, σ are large, the effective cosmological term, appearing in the equations determining A, is suppressed. The largeness of λ_5 and σ translates, according to (21), into the smallness of the mass scale A. This in turn points to a very large length scale associated to the metric, such as might be induced by inflation. In any case, it seems significant that the effective cosmological term can here be suppressed by a multiplicative, as opposed to a subtractive, mechanism.

§8 Remarks

1. If these considerations correspond to reality, the structures assumed in general relativity, including the conventional notion of distance and distinction of time, are not primary. Indeed, in the unbroken symmetry phase there is no distinction among the variables describing space and time, and no notion of distance (only volume). It will be very interesting to consider the possibility that a more symmetric phase was realized in the early Universe, or near singularities of general relativity.

2. Our considerations have been entirely classical. The actions postulated here, although polynomial in the fields, are of a very unfamiliar type for quantum field theory. Considerations of a different order will be required to see if it is possible to embed them into a consistent quantum theory. This could well require supersymmetry to soften the ultraviolet behavior, and extra fields or extra dimensions to cancel anomalies. Since the symmetry-broken phase is close to conventional general relativity, one might suspect that the structures used in successful quantizations of that theory could also be brought to bear here.

3. There is no apparent reason why the internal group could not be substantially larger. It could break down to a product of a non-trivial internal group and local Lorentz invariance, rather than just the latter. The occurrence, in attractive unification schemes based on $SO(10)$, of representations that are spinors both for the internal and the space-time symmetry is suggestive in this regard.

Acknowledgments I wish to thank Mark Alford, John March-Russell and Lisa Randall for valuable discussions.

REFERENCES

[1] H. Weyl, *Sitz. p. Preuss. Alcad. d. Wiss*, 1918, reprinted in A. Einstein, et al., “The Principle of Relativity,” (Dover, N.Y. 1952).

[2] C.N. Yang and R. Mills, *Phys. Rev.* **96**, 191 (1954).

[3] T.L. Kaluza, *Sitz. Bal. Akad.*, 966 (1921).

[4] O. Klein, *Z. Phys.* **37**, 895 (1926).

[5] R. Utiyama, *Phys. Rev.* **101**, 1597 (1956).

[6] S. MacDowell and F. Mansouri, *Phys. Rev. Lett.* **38** 739 (1977).

[7] A. Chamseddine and P. West *Nucl. Phys.* **B 129**, 39 (1977); also quite relevant is A. Chamseddine, *Ann. Phys.* **113**, 219 (1978).

[8] K. Stelle and P. West, *Nucl. Phys.* **B 129**, 39 (1979).

[9] A. Zee, in Proceedings Orbis Scientae **20**, (Plenum Press, N.Y., 1983) and references therein; Y. Ng and H. van Dam, *Phys. Rev. Lett.* **65**, 1972 (1990).

[10] I am very grateful to R. Sundrum for raising this issue.