

General Relativity - its beauty, its curves, its rough edges . . . and its lessons for gauge fields *

Tom Lawrence

Stafford, UK

tom.lawrence@ronininstitute.org

May 2, 2022

Abstract

The core principle of General Relativity (GR) is that curved spacetime is manifested as gravity. Kaluza-Klein theories use this principle to interpret non-gravitational interactions. Most require spacetime to be extended in a *complex* way. If, instead, spacetime is extended in the most natural way, orthogonal gauge symmetries result. These represent the action of the ‘internal’ symmetries on vectorial matter. The unitary groups act on their spinor representations. Complex spaces are needed to describe these fields.

This essay pinpoints the beauty and explanatory power of GR in the notion of spacetime curvature, rather than its field equation. A similar argument from first principles must underlie any unified fundamental theory of physics - including its field equations. We conclude by discussing some features that might be expected of these.

*Essay written for the Gravity Research Foundation 2022 Awards for Essays on Gravitation

Introduction

The core principle of GR, that gravity is how we experience the curvature of spacetime, is perhaps the most elegant in physics, as this essay will argue. Its beauty lies in how it is logically argued from first principles, yet the consequences represent a revolutionary change in understanding – and explain many things that the existing Newtonian theory did not.

This step-by-step inevitability does not extend to the field equation. Consequently, variants and modifications of GR have proliferated.

To develop a unified theory of the fundamental interactions, additional degrees of freedom must be incorporated. Einstein looked at several possible ways of doing this[1]. This essay describes how the logical approach of GR plays out for this task, when supplemented by our modern understanding of symmetries and gauge fields.

The first part of this essay looks at pure GR. It describes how notions from Riemannian geometry arise when incorporating non-inertial frames into a spacetime framework. Gravity as curvature follows quickly and logically from there. The precise form of the field equation does not seem to arise from such a clear physical reasoning. But the basic notions of GR – of space-time curvature, the equivalence principle, and geodesics which carry local inertial frames and diverge/converge in the vicinity of matter – seem widely accepted, even in alternatives to GR.

The second part considers how this conceptual structure may be extended to incorporate gauge fields. The entry point for this is considering the symmetry transformations already present in GR and how they are related to the components of the Jacobian matrix for coordinate changes. The degrees of freedom carried by the Jacobian matrix include those of an $O(1,3)$ group. Increasing the number of spatial dimensions results in new orthogonal group symmetries. These are related to the unitary groups of the gauge interactions. The gauge symmetries act directly on spinors and work is still ongoing to determine how to incorporate spinors into this theory. But the induced action of these gauge symmetries on tensor fields constructed from the spinors can be calculated, at least for the electromagnetic

interaction and the full electroweak interaction.

The final part considers the complex spaces that may underlie the definition of spinors - the mathematical objects that might characterise these, and what a geometric field equation might look like for spinors.

General relativity

Curvature = gravity

Newton's law of gravity was itself revolutionary, particularly when considered as part of the framework of mechanics that Newton had developed. It is essentially a phenomenological law - its form is constructed to fit careful observations of the motion of objects in our solar system. There is no explanation of why the law should take this form. Neither is there any explanation of how one object can act on another hundreds of millions of miles away across seemingly empty space - with a change to the first object instantly affecting the second. A universal measurement of time is assumed and all calculations within this framework must be carried out in an inertial frame, or otherwise modified by introducing inertial forces.

The differential form of the law cast a new light on the problem of remote action. It involves a gravitational potential - a field spread across space, which emanates from the gravitating body and falls off with distance. But the mathematical form and the nature of this field remained unexplained.

Maxwell's theory of electromagnetism was also based on fields, but disturbances in these fields travelled at a fixed velocity in vacuum.

This was a key motivation for the development of Special Relativity (SR). In SR, every observer has their own spatial and time coordinates - the co-ordinate system in a sense encodes the experience of the observer. For observers who are in relative motion with constant velocity, the coordinates of their frames of reference are linearly related. For example, for a

boost in the x^1 -direction,

$$x'^1 = \beta x^1 - \frac{v}{c} \beta x^0, \quad x'^0 = \beta x^0 - \frac{v}{c} \beta x^1, \quad (1)$$

where

$$\beta = \left(1 - \frac{v^2}{c^2}\right)^{-1/2} \quad (2)$$

By analogy with distance between two neighbouring points, we have an infinitesimal interval:

$$ds^2 = dx^0 dx_0 + dx^1 dx_1 + dx^2 dx_2 + dx^3 dx_3 = \eta_{\mu\nu} dx^\mu dx^\nu \quad (3)$$

This preserves its form under a boost, that is, the metric is invariant.

This spacetime formulation of SR, proposed by Minkowski, underpins not just GR but quantum field theories as well.

For observers who are in relative motion with varying velocity, such as linear acceleration or one orbiting around the other, the coordinate transformations do not preserve the metric and in general are non-linear. This leads to the experience of inertial forces. To describe such situations covariantly, we need to apply concepts from Riemannian geometry.

By way of illustration, take a system of Minkowski coordinates x^μ . If we transform to a second coordinate system x'^μ which is given by rotating the x - y plane of x^μ at a constant rate of k radians per second,

$$x' = x \cos(kt) + y \sin(kt), \quad y' = -x \sin(kt) + y \cos(kt), \quad z' = z, \quad t' = t \quad (4)$$

This coordinate system has a metric which differs from the Minkowski one (indeed it has off-diagonal components)[2].

The Jacobian matrix components $\partial x'/\partial x$, $\partial x'/\partial t$, $\partial y'/\partial x$ and $\partial y'/\partial t$ are then functions of x , y and t . This spoils the covariant transformation of partial derivatives. For example, if

a vector field V^μ is defined along a curve with parameter λ which is a function of x , y and t ,

$$\frac{dV^\mu}{d\lambda} = \frac{dx^\mu}{dx'^\nu} \frac{dV'^\nu}{d\lambda} + \frac{d}{d\lambda} \left(\frac{dx^\mu}{dx'^\nu} \right) V'^\nu \quad (5)$$

Had we transformed to another set of Minkowski coordinates, we would just have the first term on the right, so the derivative would transform as a vector.

In the case where V^μ is the four-momentum of an object and λ is its proper time, the additional term represents an inertial force.

This is occurring because the transformation from the x^μ coordinate basis to that for x'^μ is varying from event to event. To allow for this, we can introduce a notion of parallel transport between events, enabling us to construct derivatives which transform as tensors.

Note that all the analysis so far has been in the absence of gravitational fields. We have introduced concepts from Riemannian geometry purely to allow SR to operate smoothly in the curvilinear coordinate systems required by mutually accelerating reference frames.

In the absence of gravitational fields, it is always possible to transform to an inertial frame across the entire spacetime, in which free particles travel along straight lines. Their four-momentum is conserved and the connection reduces to zero in this frame. (In any non-inertial frame, the connection is non-zero and the four-momentum is not conserved absolutely, but only covariantly.)

In Riemannian geometry, the existence of a co-ordinate system in which the connection is everywhere zero defines a flat space.

On a curved space, there is more than one sensible definition of parallel transport and each definition has a corresponding connection, as described below. For now, we adopt the Levi-Civita connection. There are paths passing through every point on which this connection can be reduced to zero by a change of coordinates – these are geodesics. Immediate consequences are that the tangent vector to the path at any point is parallel to the tangent vector at each other point, and so the path obeys the geodesic equation. This equation can also be found

using the principle of least action, showing that this is an extremal path[3].

We then recall the equivalence of gravitational and inertial mass. This implies that every object at the same point in a gravitational field accelerates at the rate. Consequently, the effect of a gravitational field can be eliminated locally by transforming to a freely falling frame. An observer falling freely will not be able to tell whether the object they have just let go of is subject to a gravitational field - it will hover in front of them. Its four-momentum is conserved, so the observer is in a local inertial frame.

It is not a global inertial frame. If an observer is falling directly towards a gravitating body, for example, they will observe a distant object also falling towards that body as moving towards them, with varying four-momentum. This acceleration cannot be eliminated by a choice of frame – indicating that spacetime is curved. The straight world lines in a local inertial frame allows one to identify freefall trajectories with geodesics and free fall frames with geodesic coordinates.

This seemed a radical idea when Einstein proposed it. But the geometry it relies on can be easily adapted from that which describes familiar two-dimensional curved spaces.

The curvature is characterised by the Riemann tensor, which determines the rate of convergence of geodesics. The resulting theory, unlike Newtonian theory, gave the correct value for the precession of the perihelion of Mercury. It also overcomes the problems with Newton's law mentioned above. In the absence of gravitational fields, it reduces to SR, so is automatically consistent with it. There is no problem with remote action. And in appropriate limits, it reduces to Newton's law, thus explaining its inverse square form[4].

Expressed in this way, the passage from the covariant formulation of SR to the description of gravity as the curvature of space-time seems a natural one, following almost inevitably from the principle of equivalence. If spacetime is curved, the key features of gravity must result – the equivalence principle, tidal forces, the convergence of freefall paths, one body orbiting relative to another and so forth. It would seem absurd to suggest that this is coincidence, or an induced or derived effect – that Einstein did not uncover a fundamental truth about

the universe, reasoned out from first mathematical principles. (As will be described below, curvature of spacetime can also result in key features of other forces.)

Small wonder, then, that Einstein, when asked what he would have done if Eddington's observations had disapproved GR, replied "then I would have felt sorry for the dear Lord. The theory is correct".

However, of course, the narrative above is not the full story. A complete theory required an equation relating curvature to the properties of the gravitating matter.

The GR field equation

By considering the Newtonian correspondence of the equation of geodesic deviation, it is possible to deduce that the Ricci tensor vanishes in vacuum[3]. This suggests that the Ricci tensor may be related to the matter distribution. Dimensional arguments can be used to deduce that the analogue of the matter density in the Poisson equation for gravity should be a symmetric rank-two tensor[4]. This is expected to be conserved by way of it being a Noether current. This means it is divergenceless. The Ricci tensor, in contrast, has a divergence which is proportional to its trace. Einstein therefore proposed that the divergenceless part of the Ricci tensor was proportional to the energy-momentum density tensor.

The field equation is thus well motivated, but the motivations rely on a combination of Newtonian theory and technical constraints, for which the physical meaning is not entirely clear. The action of curvature on test particles has been reasoned out from first principles – it cannot be anything else. This is not true of the relation between matter and the space-time curvature it causes and we may reasonably ask why it should take this form.

Now the field equation may be derived by varying the Einstein-Hilbert action with respect to the metric – and in the case of the Palatini approach, the connection – using the principle of stationary action. But this leaves the question of why the action should be the Einstein-Hilbert one. We may add to this the questions of why, under such variations, the action should be stationary and the surface term vanish.

A range of theorems and results have been published on the uniqueness of field equation, but these all impose technical constraints on the equation or the action[5, 6, 7, 8, 9].

This lack of uniqueness of the field equation has provided space for alternative theories to be proposed – Brans-Dicke theory, $f(R)$ theories, Gauss-Bonnet gravity and so on. However, it is conspicuous that all of these assume gravity to be a manifestation of dynamic geometry.

Applying the conceptual structure to gauge interactions

Tangent space transformations and teleparallelism

The key to understanding how the theory may be extended to incorporate gauge fields is to look at the symmetry transformations already present in the theory. The fundamental transformations of GR are changes of coordinate system. Such transformations are taken to be analytic and invertible; these form a group with an infinite number of parameters.

However, they induce a group of transformation on a tangent space which is far simpler – it only has 16 parameters. It is the simplicity of tensor transformations under this group which enables us to carry out calculations in GR[2].

Einstein himself looked into these 16 quantities in the decade after general relativity was published[10]. He noted that the Riemannian metric is determined by just 10 of them. He wondered whether the other six could represent the electromagnetic field[1]. This led him to look at the Weitzenböck connection, which has torsion but zero field strength – a subject on which he corresponded with Cartan.

At this point, a misconception crept in, which has persisted in much of the literature, and which is rectified below. When Cartan first wrote to Einstein in 1929, he explained that there were spaces containing curvature and torsion; “in spaces where parallelism is defined in the Levi-Civita way, the torsion is zero; in spaces where parallelism is absolute the curvature is zero”. Einstein started his reply with the sentence “I see, indeed, that the manifolds used by me are a special case of those studied by you”[11].

However, Cartan had focused on manifolds over which a parallelism could be defined *globally*. There do exist such manifolds. However, those with a Riemannian or pseudo-Riemannian metric are a subclass of Riemannian/pseudo-Riemannian manifolds – not the other way round. This is because for *any* Riemannian or pseudo-Riemannian manifold, one can choose a coordinate neighbourhood around any point across which such ‘absolute’ parallelisms can be defined. The Weitzenböck connection is then well-defined across the neighbourhood and exists on the *same* manifold as the Levi-Civita one[2]. (Indeed, it is now known that any non-symmetric, metric-compatible connection, including the Weitzenböck connection, can be written as the sum of its contorsion and the Levi-Civita connection[12].)

All manifolds which admit a global parallelism by definition admit one over a coordinate neighbourhood, but the converse is certainly not true – global parallelism is a far tighter restriction. (For example, it is well known that even-dimensional spheres do not admit a parallel field globally, but they are Riemannian manifolds and admit them locally.)

A proper analysis of the 16 components of the Jacobian matrix on a given tangent space requires a knowledge of subgroup-coset space structures. The importance of these structures to field theories was only understood when the theory of non-linear realisations was developed in the late 1960s [13, 14, 15, 16]. These realisations came to be understood as effective theories resulting from situations of symmetry breaking[17].

I have used this knowledge to identify the parameters as follows[2]:

- The 16 parameters are those of the group $GL(4, R)$;
- An orthonormal frame basis is mapped to another orthonormal frame basis by an $O(1, 3)$ subgroup of this, with six parameters. These relate to a freedom of choice of which parallelism to base the Weitzenböck connection upon;
- The 10 independent components of the metric are formed from the parameters of the corresponding coset space.

Actually, an analysis using the Weitzenböck connection in the way Einstein attempted

results in the Teleparallel Equivalent of GR (TEGR)[18, 19, 12, 20].

Gauge symmetries and additional dimensions

In order to incorporate extra gauge degrees of freedom, the tangent space must be expanded, by increasing the dimensionality of the underlying spacetime. Kaluza and Klein took this approach, adding one spatial dimension - which piqued Einstein's interest[1].

However, they did not have our modern understanding of gauge fields. The archetypal gauge theory was published by Yang and Mills in 1954, the year before Einstein's death[21]. Two years later, the general theory – covering both electromagnetism and Yang-Mills theory – was provided by Utiyama, who tried to represent general relativity in the same way[22].

Consequently, we now know that electromagnetism is the gauge field of a $U(1)$ symmetry – a local change of complex phase. Similarly, the electroweak interaction is mediated by an $SU(2) \otimes U(1)$ gauge field and the strong interaction by an $SU(3)$ gauge field. Unitary transformation matrices have complex components and map an orthonormal basis on a complex vector space to another orthonormal basis. It does not make sense to apply such transformations to real vector spaces.

Instead, they should be applied to spinors. However, it is always possible to construct the outer product of a spinor and its adjoint spinor, and use a Clifford algebra to decompose this into tensors. Amongst these, there is always a real vector. The action of a unitary transformation on a vector representation formed in this way is actually an orthogonal one. The vector representation of electromagnetic $U(1)$ is two-dimensional, so we need an extra two real dimensions to incorporate these gauge degrees of freedom. Naturally, the same also goes for a hypercharge symmetry. Similarly, the vector representation of $SU(2)$ isospin is three-dimensional, so we need a further three dimensions to incorporate these degrees of freedom, giving a total of five extra dimensions for the electroweak symmetry.

I have shown that when space-time forms a product of our familiar four dimensions and a two-dimensional space, and a generalisation of Kaluza's cylinder condition is satisfied,

then there are extra connection components that can be identified as $U(1)$ gauge fields. Similarly, a product of a four-dimensional space-time and a three-dimensional space, with the generalised cylinder condition applying, gives rise to $SU(2)$ gauge fields[23].

The field strength of these gauge fields are components of the Riemann tensor, which do not contribute to the Ricci tensor of either factor space.

The strong interaction cannot be handled in exactly the same way, as the dimensionality of the spinor representations of orthogonal groups is always a power of two. Colour $SU(3)$ must therefore be embedded in an $SU(4)$ group.

The geometry of spinors

The previous section summarises how both gravity and the electroweak gauge fields can be manifestations of the curvature of real spacetime and couple to/act on tensorial matter. This is fully covariant in all dimensions.

We can then go further and ask whether we can describe fundamental spinors as manifestations of geometry. This would realise Clifford's idea that "...this variation of the curvature of space is what really happens in that phenomenon we call the *motion of matter*... That in the physical world nothing else takes place but this variation"[24]. Or, as Davies puts it, "the forces and fields . . . themselves being explained in terms of geometry"[25].

This aspect of the theory is considerably less developed. However, the above considerations give us clues as to how to proceed.

Unitary transformations operate on spinors which inhabit a complex vector space. These may be used to construct vectors which inhabit a real vector space and are operated on by orthogonal transformations. The real vector space is tangent to the underlying spacetime manifold. It therefore seems reasonable to postulate that the complex vector space inhabited by the spinors is tangent to a curved complex manifold.

We can then ask what this manifold looks like - how may we characterise it?

Again, we can turn to GR for inspiration. In GR, the geometry of spacetime is determined by the field equation and characterised by geometric tensors such as the metric, Ricci tensor and Einstein tensor. However, the values of the components of these tensors are different in different co-ordinate systems, but this is irrelevant to the shape of the space-time. We can factor out the effect of coordinate changes by looking at their action on the operator form of these tensors. The induced action of a change of coordinates on the operator form of a tensor is a similarity transformation:

$$j : X_I^J \rightarrow X_I'^J = j_I^K X_K^L (j^{-1})_L^J \quad (6)$$

where j is the Jacobian matrix[23]. The set of all possible values of the tensor is partitioned into orbits under this action, which are distinguished by their eigenvalues.

This form of the tensor is known as the operator form because it maps one vector to another vector, under the operation of index contraction.

Indeed, the GR field equation itself may be contracted with a vector, resulting in a matrix or differential operator equation for the vector:

$$G_\sigma^\rho V^\sigma - \Lambda V^\rho = (D_\sigma D^\rho - D^\rho D_\sigma) V^\sigma - \left(\frac{R}{2} - \Lambda\right) V^\rho = \frac{8\pi G}{c^4} T_\sigma^\rho V^\sigma \quad (7)$$

We then recall that spinors were first suggested by Dirac as solutions of a relativistic wave equation involving first-order differential operators. We might therefore expect a curved complex space to be characterised by complex differential or matrix operators with two spinor indices, which can map one spinor into another. Its eigenvalue equation might then be a curved space generalisation of the Dirac equation.

References

- [1] T. Sauer, Einstein’s unified field theory program, in *The Cambridge Companion to Einstein*, eds. M. Janssen and C. Lehner (Cambridge University Press, 2014). pp.281–305
- [2] T. Lawrence, Tangent space symmetries in general relativity and teleparallelism, *Int. J. Geom. Meth. Mod. Phys.* **18** (2021) supp01, 2140008
- [3] R. D’Inverno, *Introducing Einstein’s Relativity* (Oxford University Press, Oxford, 1995)
- [4] I. Kenyon, *General Relativity* (Oxford University Press, Oxford, 1995)
- [5] D. Lovelock, The Einstein tensor and its generalizations, *J. Math. Phys.* **12** (1971) 498–501; D. Lovelock, The four-dimensionality of space and the Einstein tensor, *J. Math. Phys.* **13** (1972) 874–876
- [6] A. Navarro and J. Navarro, Lovelock’s theorem revisited, *J. Geom. Phys.* **61** (2011) 1950–1956, arXiv:1005.2386 [math-ph]
- [7] E. Curiel, A simple proof of the uniqueness of the Einstein field equation in all dimensions, arXiv:1601.03032 [gr-qc]
- [8] J. Kijowski, Universality of the Einstein theory of gravitation, *Int. J. Geom. Meth. Mod. Phys.* **13** (2016) 1640008, arXiv:1604.05052 [gr-qc]
- [9] V. Afonso, G. Olmo and D. Rubiera-Garcia, Mapping Ricci-based theories of gravity into general relativity, *Phys. Rev D* **97** (2018) 021503, arXiv:1801.10406 [gr-qc]
- [10] A. Einstein, Riemann-Geometrie mit Aufrechterhaltung des Begriffes des Fernparallelismus, *Sitzber. Preuss. Akad. Wiss.* **17** (1928) 217–222; translated into English in A. Unzicker and T. Case, Translation of Einstein’s Attempt of a Unified Field Theory with Teleparallelism, arXiv:physics.hist-ph/0503046

- [11] A. Einstein and E. Cartan, *Letters on Absolute Parallelism 1929-1932*, translated into English by J. Leroy and J. Ritter, ed. R. Debever (Princeton University Press, 1979), p.11
- [12] J. Pereira, Lorentz connections and gravitation, *AIP Conf.Proc.* **1483** (2012), 239–259, arXiv:1210.0379 [gr-qc]
- [13] S. Coleman, J. Wess and B. Zumino, Structure of phenomenological Lagrangians. I, *Phys. Rev.* **177** (1969), 2239–2247
- [14] C. Callan, S. Coleman, J. Wess and B. Zumino, Structure of phenomenological Lagrangians. II, *Phys. Rev.* **177** (1969), 2247–2250
- [15] C. Isham, Metric structures and chiral symmetries, *Nuov. Cim. A* **61** (1969) 188–202
- [16] C. Isham, The embedding of non-linear meson transformations in a Euclidean space *Nuov. Cim. A* **61** 1969 729–36
- [17] A. Salam and J. Strathdee, Nonlinear realizations. I. The role of Goldstone bosons, *Phys. Rev.* **184** (1969), 1750–59.
- [18] V. de Andrade, H. Arcos and J. Pereira, Torsion as alternative to curvature in the description of gravitation, arXiv:gr-qc/0412034
- [19] R. Aldrovandi, T. Gribl Lucas and J. Pereira, Inertia and gravitation in teleparallel gravity, arXiv:0812.0034 [gr-qc]
- [20] J. Maluf, The teleparallel equivalent of general relativity, *Ann. Phys. (Berlin)* **525** (2013), 339–57
- [21] C. Yang and R. Mills, Conservation of isotopic spin and isotopic gauge invariance, *Phys. Rev.* **96** (1954), 191–195.

- [22] R. Utiyama, Invariant theoretical interpretation of interaction, *Phys. Rev.* **101** (1955), 1597–1607
- [23] T. Lawrence, Product manifolds as realisations of general linear symmetries, arXiv:2203.09473 [gr-qc] - accepted for publication in IJGMMP, doi:10.1142/S0219887822400060
- [24] W. Clifford, On the space-theory of matter (1870), *Proc. Cambridge Philos. Soc.* **2** (1876), 157–158
- [25] P. Davies, *Superforce - The Search for a Grand Unified Theory of Nature* (Unwin Paperbacks, London, 1989)