MacDowell–Mansouri Gravity and Cartan Geometry

Derek K. Wise
Department of Mathematics
University of California
Riverside, CA 92521, USA

email: derek@math.ucr.edu

November 29, 2006

Abstract

The geometric content of the MacDowell–Mansouri formulation of general relativity is best understood in terms of Cartan geometry. In particular, Cartan geometry gives clear geometric meaning to the MacDowell–Mansouri trick of combining the Levi–Civita connection and coframe field, or soldering form, into a single physical field. The Cartan perspective allows us to view physical spacetime as tangentially approximated by an arbitrary homogeneous 'model spacetime', including not only the flat Minkowski model, as is implicitly used in standard general relativity, but also de Sitter, anti de Sitter, or other models. A 'Cartan connection' gives a prescription for parallel transport from one 'tangent model spacetime' to another, along any path, giving a natural interpretation of the MacDowell–Mansouri connection as 'rolling' the model spacetime along physical spacetime. I explain Cartan geometry, and 'Cartan gauge theory', in which the gauge field is replaced by a Cartan connection. In particular, I discuss MacDowell–Mansouri gravity, as well as its recent reformulation in terms of BF theory, in the context of Cartan geometry.

Contents

1 Introduction

The geometry of ordinary general relativity is by now well understood—spacetime geometry is described by the Levi–Civita connection on the tangent bundle of a Lorentzian manifold. In the late 1970s, MacDowell and Mansouri introduced a new approach, based on broken symmetry in a type of gauge theory [20]. This approach has been influential in such a wide array of gravitational theory that it would be a difficult task to compile a representative bibliography of such work. The original MacDowell–Mansouri paper continues to be cited in work ranging from supergravity [15, 16, 22, 27] to background-free quantum gravity [13, 14, 25].

However, despite their title "Unified geometric theory of gravity and supergravity", the geometric meaning of the MacDowell–Mansouri approach is relatively obscure. In the original paper, and in much of the work based on it, the technique seems like an unmotivated "trick" that just happens to give the equations of general relativity. One point of the present paper is to show that MacDowell–Mansouri theory is no trick after all, but rather a theory with a rich geometric structure, which may offer insights into the geometry of gravity itself.

In fact, the secret to understanding the geometry behind their work had been around in some form for over 50 years by the time MacDowell and Mansouri introduced their theory. The geometric foundations had been laid in the 1920s by Élie Cartan, but were for a long time largely forgotten. The relevant geometry is a generalization of Felix Klein's celebrated *Erlanger Programm* to include inhomogeneous spaces, called 'Cartan geometries', or in Cartan's own terms, *espaces généralisés* [9, 10]. The MacDowell–Mansouri gauge field is a special case of a 'Cartan connection', which encodes geometric information relating the geometry of spacetime to the geometry of a homogeneous 'model spacetime' such as de Sitter space. Cartan connections have been largely replaced in the literature by what is now the usual notion of 'connection on a principal bundle' [11], introduced by Cartan's student Charles Ehresmann [12].

The MacDowell–Mansouri formalism has recently seen renewed interest among researchers in gravitational physics, especially over the past 5 years. Over a slightly longer period, there has been a resurgence in the mathematical literature of work related to Cartan geometry, no doubt due in part to the availability of the first modern introduction to the subject [24]. Yet it is not clear that there has been much communication between researchers on the two sides—physical and mathematical—of what is essentially the same topic.

In this paper I review Klein geometry, leading up to Cartan geometry. I show how Cartan geometry is suited to describing the classical constraint problem of rolling a homogeneous manifold on another manifold, and use this idea to understand the geometry of MacDowell–Mansouri gravity.

MacDowell–Mansouri gravity

MacDowell–Mansouri gravity is based on a gauge theory with gauge group $G \supset \mathrm{SO}(3,1)$ depending on the sign of the cosmological constant[1]:

$$G = \begin{cases} \mathrm{SO}(4,1) & \Lambda > 0 \\ \mathrm{SO}(3,2) & \Lambda < 0 \end{cases}$$

[1]For simplicity, I restrict attention to MacDowell–Mansouri theory for gravity, as opposed to supergravity.

To be definite, let us focus on the case of $\Lambda > 0$, where $G = \mathrm{SO}(4,1)$. The Lie algebra has a splitting:

$$\mathfrak{so}(4,1) \cong \mathfrak{so}(3,1) \oplus \mathbb{R}^{3,1}, \tag{1}$$

not as Lie algebras but as vector spaces with metric.

If F is the curvature of the $\mathrm{SO}(4,1)$ gauge field A, the Lagrangian is:

$$S_{\mathrm{MM}} = \frac{-3}{2G\Lambda} \int \mathrm{tr}\,(F \wedge \star \hat{F}) \tag{2}$$

Here $\hat{F}$ denotes the projection of F into the subalgebra $\mathfrak{so}(3,1)$, and $\star$ is an internal Hodge star operator. This projection breaks the $\mathrm{SO}(4,1)$ symmetry, and the resulting equations of motion are, quite surprisingly, the Einstein equation for ω with cosmological constant Λ, and the vanishing of the torsion.

The orthogonal splitting (1) provides the key to the MacDowell–Mansouri approach. Extending from the Lorentz Lie algebra $\mathfrak{so}(3,1)$ to $\mathfrak{so}(4,1)$ lets us view the connection ω and coframe field e of Palatini-style general relativity as two aspects of the connection A. The reason this is possible *locally* is quite simple. In local coordinates, these fields are both 1-forms, valued respectively in the Lorentz Lie algebra $\mathfrak{so}(3,1)$ and Minkowski vector space $\mathbb{R}^{3,1}$. Using the splitting, we can combine these local fields in an $\mathrm{SO}(4,1)$ connection 1-form A, which has components $A^I_{\mu J}$ given by[2]

$$A^i_{\mu j} = \omega^i_{\mu j} \qquad A^i_{\mu 4} = \frac{1}{\ell} e^i.$$

where ℓ is a scaling factor with dimensions of length.

This connection form A has a number of nice properties, as MacDowell and Mansouri realized. The curvature $F[A]$ also breaks up into $\mathfrak{so}(3,1)$ and $\mathbb{R}^{3,1}$ parts. The $\mathfrak{so}(3,1)$ part is the curvature $R[\omega]$ plus a cosmological constant term, while the $\mathbb{R}^{3,1}$ part is the torsion $d_\omega e$:

$$F_{\mu\nu}{}^i{}_j = R_{\mu\nu}{}^i{}_j - \frac{\Lambda}{3}(e \wedge e)_{\mu\nu}{}^i{}_j \qquad F_{\mu\nu}{}^i{}_4 = (d_\omega e)_{\mu\nu}{}^i$$

where we choose $\ell^2 = 3/\Lambda$. This shows that when the curvature $F[A]$ vanishes, so that $R - \frac{\Lambda}{3}\, e \wedge e = 0$ and $d_\omega e = 0$, we get a torsion free connection for a universe with positive cosmological constant.

The BF reformulation

Recently, the basic MacDowell–Mansouri technique has been used with a different action [13, 21, 25], based on 'BF theory' [4]. This work has in turn been applied already in a variety of ways, from cosmology [2] to the particle physics [19]. The setup for this theory is much like

[2]Here, I use the Latin alphabet for internal indices, with capital indices running from 0 to 4, and lower case indices running from 0 to 3:

$$I, J, K, \ldots \in \{0,1,2,3,4\}$$
$$i, j, k, \ldots \in \{0,1,2,3\}.$$

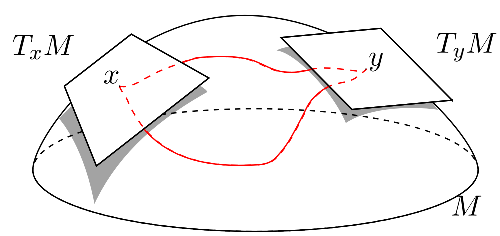
that of the original MacDowell–Mansouri theory, but in addition to the G-connection, there is a 2-form B with values in the Lie algebra $\mathfrak{g}$ of the gauge group. The action proposed by Freidel and Starodubtsev has the appearance of a perturbed BF theory[3]:

$$S = \int \mathrm{tr}\left(B \wedge F - \frac{G\Lambda}{6} B \wedge \star \hat{B}\right). \tag{3}$$

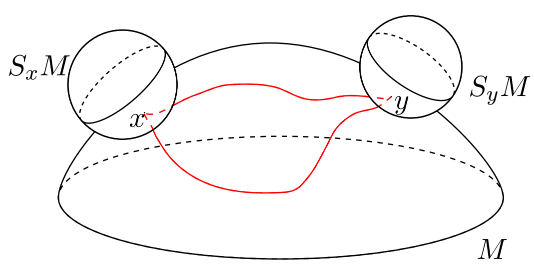
I give a treatment of both the original MacDowell–Mansouri action and the Freidel–Starodubtsev reformulation in Section 5.3, after developing the appropriate geometric setting for such theories, which lies in Cartan geometry.

The idea of a Cartan geometry

What is the geometric meaning of the splitting of an SO(4, 1) connection into an SO(3, 1) connection and coframe field? For this it is easiest to first consider a lower-dimensional example, involving SO(3) and SO(2). An oriented 2d Riemannian manifold is often thought of in terms of an SO(2) connection since, in the tangent bundle, parallel transport along two different paths from x to y gives results which differ by a rotation of the tangent vector space at y:



In this context, we can ask the geometric meaning of extending the gauge group from SO(2) to SO(3). The group SO(3) acts naturally not on the bundle TM of tangent *vector spaces*, but on some bundle SM of 'tangent *spheres*'. We can construct such a bundle, for example, by compactifying each fiber of TM. Since SO(3) acts to rotate the sphere, an SO(3) connection on a Riemannian 2-manifold may be viewed as a rule for 'parallel transport' of tangent spheres, which need not fix the point of contact with the surface:



An obvious way to get such an SO(3) connection is simply to *roll a ball on the surface*, without twisting or slipping. Rolling a ball along two paths from x to y will in general give different

[3]For simplicity, I ignore a term in the Freidel–Starodubtsev action proportional to $\mathrm{tr}\,(B \wedge B)$ that vanishes if we choose the Immirzi parameter $\gamma = 0$.

results, but the results differ by an element of SO(3). Such group elements encode geometric information about the surface itself.

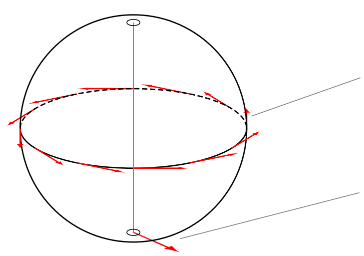
In our example, just as in the extension from the Lorentz group to the de Sitter group, we have an orthogonal splitting of the Lie algebra

$$\mathfrak{so}(3) \cong \mathfrak{so}(2) \oplus \mathbb{R}^2$$

given in terms of matrix components by

$$\begin{bmatrix} 0 & u & a \\ -u & 0 & b \\ -a & -b & 0 \end{bmatrix} = \begin{bmatrix} 0 & u & \\ -u & 0 & \\ & & 0 \end{bmatrix} + \begin{bmatrix} & & a \\ & & b \\ -a & -b & \end{bmatrix}.$$

As in the MacDowell–Mansouri case, this allows an SO(3) connection A on an oriented 2d manifold to be split up into an SO(2) connection ω and a coframe field e. But here it is easy to see the geometric interpretation of these components: an infinitesimal rotation of the tangent sphere, as it begins to move along some path, breaks up into a part which rotates the sphere about its point of tangency and a part which moves the point of tangency:

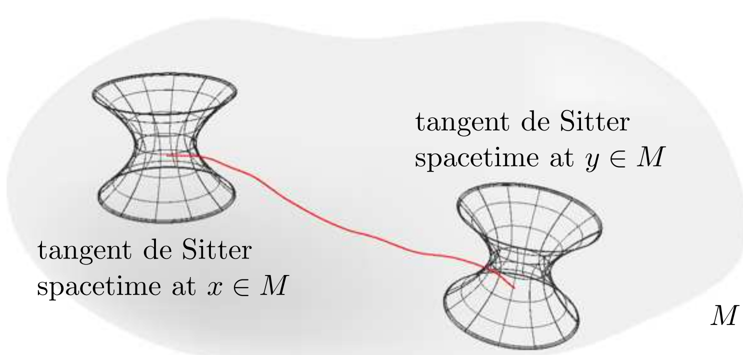


The $\mathfrak{so}(2)$ part gives an infinitesimal rotation around the axis through the point of tangency.

The $\mathbb{R}^2$ part gives an infinitesimal translation of the point of tangency.

The connection thus *defines* a method of rolling a tangent sphere along a surface.

Extrapolating from this example to the extension $\mathrm{SO}(3,1) \subset \mathrm{SO}(4,1)$, we surmise a geometric interpretation for MacDowell–Mansouri gravity: the SO(4,1) connection $A = (\omega, e)$ encodes the geometry of spacetime M by "rolling de Sitter spacetime along M":



This idea is appealing since, for spacetimes of positive cosmological constant, we expect de Sitter spacetime to be a better infinitesimal approximation than flat Minkowski vector space. The geometric beauty of MacDowell–Mansouri gravity, and related approaches, is that they study spacetime using 'tangent spaces' that are truer to the mean geometric properties of the spacetime itself. Exploring the geometry of a surface M by rolling a ball on it may not seem like a terribly useful thing to do if M is a plane; if M is some slight deformation of a sphere, however, then exploring its geometry in this way is very sensible! Likewise, approximating a spacetime by de Sitter space is most interesting when the spacetime has the same cosmological constant.

More generally, this idea of studying the geometry of a manifold by "rolling" another manifold—the 'model geometry'—on it provides an intuitive picture of 'Cartan geometry'. Cartan geometry, roughly speaking, is a generalization of Riemannian geometry obtained by replacing linear tangent spaces with more general homogeneous spaces. As Sharpe explains in the preface to his textbook on the subject [24], Cartan geometry is a common generalization of Riemannian and Klein geometries. The following diagram is an adaptation of one of Sharpe's:

$$
\begin{tikzcd}
\begin{array}{c}\text{Euclidean}\\ \text{Geometry}\end{array} \arrow[r, "\begin{array}{c}\text{allow}\\ \text{curvature}\end{array}"] \arrow[d, "\begin{array}{c}\text{generalize}\\ \text{symmetry group}\end{array}"'] & \begin{array}{c}\text{Riemannian}\\ \text{Geometry}\end{array} \arrow[d, "\begin{array}{c}\text{generalize tangent}\\ \text{space geometry}\end{array}"] \\
\begin{array}{c}\text{Klein}\\ \text{Geometry}\end{array} \arrow[r, "\begin{array}{c}\text{allow}\\ \text{curvature}\end{array}"'] & \begin{array}{c}\text{Cartan}\\ \text{Geometry}\end{array}
\end{tikzcd}
$$

Like Euclidean geometry, a Klein geometry is homogeneous, meaning that there is a symmetry of the geometry taking any point to any other point. Cartan geometry provides 'curved' versions of arbitrary Klein geometries, in just the same way that Riemannian geometry is a curved version of Euclidean geometry.

But besides providing a beautiful geometric interpretation, and a global setting for the MacDowell–Mansouri way of doing gravity, Cartan geomety also helps in understanding the sense in which MacDowell–Mansouri theory is a deformation of a topological field theory.

Plan of the present paper

In Section 2, I briefly review Klein's viewpoint on homogeneous geometry using symmetry groups, focussing on the six cases most relevant to gravity: de Sitter, Minkowski, and anti de Sitter, together with their Wick rotated versions, the spherical, Euclidean, and hyperbolic Riemannian spacetimes. These six Klein geometries provide the homogeneous 'model spacetimes' which are used to describe inhomogeneous spacetimes via Cartan geometry.

Section 3 provides an introduction to Cartan geometry, and explains why this is just the right sort of geometry to describe rolling a homogeneous space on a manifold. In Section 4, I investigate further issues relevant to Cartan geometry, and particularly to the idea of doing gauge theory using a Cartan connection in place of the usual gauge connection.

Section 5 focusses on viewing general relativity through the lens of Cartan geometry. I begin with a review of the Palatini formalism and show how it can be viewed in terms of Cartan geometry. I then describe the construction of the MacDowell–Mansouri action from the Palatini action, and discuss the BF reformulation.

2 Homogeneous spacetimes and Klein geometry

Klein revolutionized modern geometry with the realization that almost everything about a homogeneous geometry—with a very broad interpretation of what constitutes a 'geometry'—is encoded in its groups of symmetries. From the Kleinian perspective, the objects of study in geometry are 'homogeneous spaces'. While many readers will be familiar with homogeneous geometry, the idea is essential to understanding Cartan geometry, so I review it here in some detail.

2.1 Klein geometry

A **homogeneous space** (G, X) is an abstract space[4] X together with a group G of transformations of X, such that G acts transitively: given any $x, y \in X$ there is some $g \in G$ such that $gx = y$.

The main tools for exploring a homogeneous space (G, X) are subgroups $H \subset G$ which preserve, or 'stabilze', interesting 'features' of the geometry. What constitutes an interesting feature of course depends on the geometry. For example, Euclidean geometry, $(\mathbb{R}^n, \mathrm{ISO}(n))$, has points, lines, planes, polyhedra, and so on, and one can study subgroups of the Euclidean group $\mathrm{ISO}(n)$ which preserve any of these. 'Features' in other homogeneous spaces may be thought of as generalizations of these notions. We can also work backwards, *defining* a feature of a geometry abstractly as that which is preserved by a given subgroup. If H is the subgroup preserving a given feature, then the space of all such features of X may be identified with the coset space G/H:

$$G/H = \{gH : g \in G\} = \text{ the space of “features of type } H\text{”}.$$

Let us illustrate why this is true using the most basic of features, the feature of 'points'. Given a point $x \in X$, the subgroup of all symmetries $g \in G$ which fix x is called the **stabilizer**, or **isotropy group** of x, and will be denoted H_x. Fixing x, the transitivity of the G-action implies we can identify each $y \in X$ with the set of all $g \in G$ such that $gx = y$. If we have two such symmetries:

$$gx = y \qquad g'x = y$$

then clearly $g^{-1}g'$ stabilizes x, so $g^{-1}g' \in H_x$. Conversely, if $g^{-1}g' \in H_x$ and g sends x to y, then $g'x = gg^{-1}g'x = gx = y$. Thus, the two symmetries move x to the same point if and only if $gH_x = g'H_x$. The points of X are thus in one-to-one correspond with cosets of H_x in G. Better yet, the map $f \colon X \to G/H_x$ induced by this correspondence is G-equivariant:

$$f(gy) = gf(y) \qquad \forall g \in G, y \in X$$

so X and G/H_x are isomorphic as H-spaces.

All this depends on the choice of x, but if x' is another point, the stabilizers are conjugate subgroups:

$$H_x = gH'_x g^{-1}$$

[4] I am deliberately vague here about what sort of 'space' a Klein geometry is. In general, X might be a discrete set, a topological space, a Riemannian manifold, etc. For our immediate purposes, the most important cases are when X has at least the structure of a smooth manifold.

where $g \in G$ is any element such that $gx' = x$. Since these conjugate subgroups of G are all isomorphic, it is common to simply speak of "the" point stabilizer H, even though fixing a particular one of these conjugate subgroups gives implicit significance to the points of X fixed by H. By the same looseness of vocabulary, the term 'homogeneous space' often refers to the coset space G/H itself.

To see the power of the Kleinian point of view, an example familiar from special relativity is $(n+1)$-dimensional Minkowski spacetime. While this is most obviously thought of as the 'space of events', there are other interesting 'features' to Minkowski spacetime, and the corresponding homogeneous spaces each tell us something about the geometry of special relativity. The group of symmetries preserving orientation and time orientation is the connected Poincaré group $\mathrm{ISO}_0(n,1)$. The stabilizer of an event is the connected Lorentz group $\mathrm{SO}_0(n,1)$ consisting of boosts and rotations. The stabilizer of an event *and* a velocity is the group of spatial rotations around the event, $\mathrm{SO}(n)$. The stabilizer of a spacelike hyperplane is the group of Euclidean transformations of space, $\mathrm{ISO}(n)$. This gives us a piece of the lattice of subgroups of the Poincaré group, with corresponding homogeneous spaces:

$$\begin{tikzcd}[column sep=large, row sep=large]
 & \mathrm{ISO}_0(n,1) \arrow[dl, dash, "\substack{\mathrm{ISO}_0(n,1)/\mathrm{ISO}(n) \\ \text{`space of spacelike hyperplanes'}}"'] \arrow[dr, dash, "\substack{\mathrm{ISO}_0(n,1)/\mathrm{SO}_0(n,1) \\ \text{`event space' (Minkowski)}}"] & \\
\mathrm{ISO}(n) \arrow[dr, dash, "\substack{\mathrm{ISO}(n)/\mathrm{SO}(n) \\ \text{`position space' (Euclidean)}}"'] & & \mathrm{SO}_0(n,1) \arrow[dl, dash, "\substack{\mathrm{SO}_0(n,1)/\mathrm{SO}(n) \\ \text{`velocity space' (hyperbolic)}}"] \\
 & \mathrm{SO}(n) &
\end{tikzcd}$$

'Klein geometries', for the purposes of this paper, will be certain types of homogeneous spaces. The geometries we are interested in are all 'smooth' geometries, so we require that the symmetry group G be a Lie group. We also require the subgroup H to be a closed subgroup of G. This is obviously necessary if we want the quotient G/H to have a topology where 1-point subsets are closed sets. In fact, the condition that H be closed in G suffices to guarantee H is a Lie subgroup and G/H is a smooth homogeneous manifold.

We also want Klein geometries to be connected. Leaving this requirement out is sometimes useful, particularly in describing discrete geometries. However, our purpose is not Klein geometry *per se*, but Cartan geometry, where the key idea is comparing a manifold to a 'tangent Klein geometry'. Connected components not containing the 'point of tangency' have no bearing on the Cartan geometry, so it is best to simply exclude disconnected homogeneous spaces from our definition.

Definition 1 *A (smooth, connected)* **Klein geometry** (G,H) *consists of a Lie group G with closed subgroup H, such that the coset space G/H is connected.*

As Sharpe emphasizes [24], for the purposes of understanding Cartan geometry it is useful to view a Klein geometry (G, H) as the principal right H bundle

$$\begin{tikzcd} G \arrow[d] \\ G/H \end{tikzcd}$$

This is a principal bundle since the fibers are simply the left cosets of H by elements of G, and these cosets are isomorphic to H as right H-sets.

Strictly speaking, a 'homogeneous space' clearly should not have a preferred basepoint, whereas the identity coset $H \in G/H$ is special. Mathematically speaking, it would thus be better to define a Klein geometry to be a principal H bundle $P \to X$ which is merely *isomorphic to* the principal bundle $G \to G/H$:

$$\begin{tikzcd} P \arrow[r, "\sim"] \arrow[d, "\pi"'] & G \arrow[d] \\ X \arrow[r, "\sim"] & G/H \end{tikzcd}$$

but not canonically so. For our purposes, however, it will actually be good to have an obvious basepoint in the Klein geometry. Since we are interested in approximating the local geometry of a manifold by placing a Klein geometry *tangent* to it, the preferred basepoint $H \in G/H$ will serve naturally as the 'point of tangency'.

2.2 Metric Klein geometry

For studying the essentially distinct types of Klein geometry, it is enough to consider the coset spaces G/H. However, for many applications, including MacDowell–Mansouri, one is interested not just in the symmetry properties of the homogeneous space, but also in its metrical properties. If we wish to distinguish between spheres of different sizes, or de Sitter spacetimes of different cosmological constants, for example, then we need more information than the symmetry groups. For such considerations, we make use of the fact that there is a canonical isomorphism of vector bundles [24]

$$\begin{tikzcd} T(G/H) \arrow[rr, "\sim"] \arrow[dr, "p"'] & & G \times_H \mathfrak{g}/\mathfrak{h} \arrow[dl, "\pi"] \\ & G/H & \end{tikzcd}$$

where the bundle on the right is the bundle associated to the principal bundle $G \to G/H$ via the adjoint representation on $\mathfrak{g}/\mathfrak{h}$. The tangent space at any point in the Klein geometry G/H is thus $\mathfrak{g}/\mathfrak{h}$, up to the adjoint representation, so an $\mathrm{Ad}(H)$-invariant metric on $\mathfrak{g}/\mathfrak{h}$ gives a metric on $T(G/H)$. In physically interesting examples, this metric will generally be nondegenerate of Riemannian or Lorentzian signature. One way to obtain such a metric is to use the Killing form on $\mathfrak{g}$, which is invariant under $\mathrm{Ad}(G)$, hence under $\mathrm{Ad}(H)$, and passes to a metric on $\mathfrak{g}/\mathfrak{h}$. When $\mathfrak{g}$ is semisimple the Killing form is nondegenerate. But even when $\mathfrak{g}$ is not semisimple, it is often possible to find a nondegenerate H-invariant metric on $\mathfrak{g}/\mathfrak{h}$, hence on $T(G/H)$. This leads us to define:

Definition 2 *A* **metric Klein geometry** (G, H, η) *is a Klein geometry* (G, H) *equipped with a (possibly degenerate)* $\mathrm{Ad}(H)$*-invariant metric on* $\mathfrak{g}/\mathfrak{h}$*.*

Notice that any Klein geometry can be made into a metric Klein geometry in a trivial way by setting $\eta = 0$. In cases of physical interest, it is generally possible to choose η to be nondegenerate.

2.3 Homogeneous model spacetimes

For studying 4d gravity, there are 4 essential homogeneous spacetimes we are interested in, corresponding to Lorentzian or Riemannian gravity with cosmological constant either positive or negative. These are the de Sitter, anti de Sitter, spherical, and hyperbolic models. We can also consider the $\Lambda \to 0$ limits of these, obtaining two more: Minkowski and Euclidean. This gives us six homogeneous 'model spacetimes', each of which can be described as a Klein geometry G/H:

	$\Lambda < 0$	$\Lambda = 0$	$\Lambda > 0$
Lorentzian	anti de Sitter $\mathrm{SO}(3,2)/\mathrm{SO}(3,1)$	Minkowski $\mathrm{ISO}(3,1)/\mathrm{SO}(3,1)$	de Sitter $\mathrm{SO}(4,1)/\mathrm{SO}(3,1)$
Riemannian	hyperbolic $\mathrm{SO}(4,1)/\mathrm{SO}(4)$	Euclidean $\mathrm{ISO}(4)/\mathrm{SO}(4)$	spherical $\mathrm{SO}(5)/\mathrm{SO}(4)$

We can reduce the number of independent cases by noting that, in their fundamental representations, the Lie algebras $\mathfrak{so}(4,1)$, $\mathfrak{iso}(3,1)$, and $\mathfrak{so}(3,2)$ consist of matrices of the form[5]

$$\begin{bmatrix} 0 & u & v & w & a \\ u & 0 & x & y & b \\ v & -x & 0 & z & c \\ w & -y & -z & 0 & d \\ \epsilon a & -\epsilon b & -\epsilon c & -\epsilon d & 0 \end{bmatrix}$$

where the value of ϵ depends on the algebra:

$$\epsilon = \begin{cases} 1 & \mathfrak{g} = \mathfrak{so}(4,1) \\ 0 & \mathfrak{g} = \mathfrak{iso}(3,1) \\ -1 & \mathfrak{g} = \mathfrak{so}(3,2) \end{cases} \tag{4}$$

and similarly for the Lie algebras $\mathfrak{so}(5)$, $\mathfrak{iso}(4)$ and $\mathfrak{so}(4,1)$ of the corresponding Riemannian models.

These are all nondegenerate *metric* Klein geometries. For the cases with $\epsilon \neq 0$, the Lie algebra has a natural metric given by[6]

$$\langle \xi, \zeta \rangle = -\frac{1}{2}\mathrm{tr}\,(\xi\zeta).$$

[5]In each case, the Lie algebra is $\mathfrak{so}(V)$ where V is a vector space with metric $(-1, +1, +1, +1, \epsilon)$.

[6]This metric is nondegenerate and invariant under the adjoint action of $\mathrm{SO}(4,1)$, hence is proportional to the Killing form, since $\mathrm{SO}(5)$, $\mathrm{SO}(4,1)$, and $\mathrm{SO}(3,2)$ are semisimple.

which is invariant under $\mathrm{Ad}(G)$, hence under $\mathrm{Ad}(H)$. With respect to this metric, we have the orthogonal direct sum decomposition of $\mathfrak{g}$

$$\begin{bmatrix} 0 & u & v & w & \\ u & 0 & x & y & \\ v & -x & 0 & z & \\ w & -y & -z & 0 & \\ & & & & 0 \end{bmatrix} + \begin{bmatrix} & & & & a \\ & & & & b \\ & & & & c \\ & & & & d \\ \epsilon a & -\epsilon b & -\epsilon c & -\epsilon d & \end{bmatrix}$$

into $\mathfrak{so}(3,1)$ and a complement $\mathfrak{p} \cong \mathfrak{g}/\mathfrak{h} \cong \mathbb{R}^{3,1}$, where on the latter subspace the metric $\langle \cdot, \cdot \rangle$ restricts to the Minkowski metric of signature $(-{+}{+}{+})$. To discuss spacetimes of various cosmological constant, we scale this metric by choosing a fundamental length ℓ and replacing the components (a, b, c, d) in the above matrices by x^i/ℓ where the x^i are dimensionful. Then, on the $\mathbb{R}^{3,1}$ subspace, the metric $\langle \cdot, \cdot \rangle$ becomes

$$\frac{\epsilon}{\ell^2} x^i y_i$$

The choice of ℓ (and ϵ) selects the value of the cosmological constant to be

$$\lambda = \frac{3\epsilon}{\ell^2} \tag{5}$$

This can be seen by comparing, for example, to de Sitter spacetime, which is the 4-dimensional submanifold of 5d Minkowski vector space given by

$$M_{\mathrm{dS}} = \left\{ (t, w, x, y, z) \in \mathbb{R}^{4,1} \;\middle|\; -t^2 + w^2 + x^2 + y^2 + z^2 = \frac{3}{\lambda} \right\}$$

where $\lambda > 0$ is the cosmological constant. But the the relationship between ℓ and λ will become clearer in Section 3.4, in the context of Cartan geometry.

So far we have discussed the metric only in the cases where the cosmological constant is nonzero. For the Minkowski case, $\mathrm{ISO}(3,1)/\mathrm{SO}(3,1)$ (or for the Euclidean case, $\mathrm{ISO}(4)/\mathrm{SO}(4)$) the trace no longer provides a nondegenerate metric. In fact, if we try using the obvious decomposition of $\mathfrak{iso}(3,1)$ analogous to the one used above, the metric induced by the trace *vanishes* on the subspace corresponding to $\mathbb{R}^{3,1}$. However, we require a metric on this subspace to be invariant only under $\mathrm{SO}(3,1)$, not under the full Poincaré group. Such a metric is easily obtained, noting the semidirect product structure:

$$\mathfrak{iso}(3,1) = \mathfrak{so}(3,1) \ltimes \mathbb{R}^{3,1}$$

of the Poincaré Lie algebra. Using the trace on $\mathfrak{so}(3,1)$ together with the usual Minkowski metric on $\mathbb{R}^{3,1}$ gives an an nondegenerate $\mathrm{SO}(3,1)$-invariant metric on the entire Poincaré Lie algebra. In particular, the metric on the $\mathbb{R}^{3,1}$ part makes $\mathrm{ISO}(3,1)/\mathrm{SO}(2)$ into a nondegenerate metric Klein geometry.

Notice that in the Minkowski case, as in de Sitter or anti de Sitter, we are still free to choose a fundamental length ℓ by which to scale vectors in $\mathbb{R}^{3,1}$, but now this choice is not constrained by the value of the cosmological constant. This points out a key difference beween the $\lambda = 0$ and $\lambda \neq 0$ cases: Minkowski spacetime has an extra 'rescaling' symmetry that is broken as the cosmological constant becomes nonzero.

3 Cartan geometry

While the beauty of Klein's perspective on geometry is widely recognized, the spacetime we live in is clearly not homogeneous. This does not mean, however, that Kleinian geometry offers no insight into actual spacetime geometry! Cartan discovered a beautiful generalization of Klein geometry—a way of modeling inhomogeneous spaces as 'infinitesimally Kleinian'. The goal of this section is explain this idea as it relates to spacetime geometry.

While this section and the next are intended to provide a fairly self-contained introduction to basic Cartan geometry, I refer the reader to the references for further details on this very rich subject. In particular, the book by Sharpe [24] and the article by Alekseevsky and Michor [1] are helpful resourses, and serve as the major references for my explanation here.

I begin with a review the idea of an 'Ehresmann connection'. Such a connection is just the type that shows up in ordinary gauge theories, such as Yang–Mills. My purpose in reviewing this definition is merely to easily contrast it with the definition of a 'Cartan connection', to be given in Section 3.2.

3.1 Ehresmann connections

Before giving the definition of Cartan connection I review the more familiar notion of an Ehresmann connection on a principal bundle. I fact, both Ehresmann and Cartan connections are related to the Maurer–Cartan form, the canonical 1-form any Lie group G has, with values in its Lie algebra $\mathfrak{g}$:

$$\omega_G \in \Omega^1(G, \mathfrak{g}).$$

This 1-form is simply the derivative of left multiplication in G:

$$\begin{aligned} &\omega_G \colon TG \to \mathfrak{g} \\ &\omega_G(x) := (L_{g^{-1}})_*(x) \quad \forall x \in T_g G. \end{aligned}$$

Since the fibers of a principal G bundle look just like G, they inherit a Maurer–Cartan form in a natural way. Explicitly, the action of G on a principal right G bundle P is such that, if P_x is any fiber and $y \in P_x$, the map

$$\begin{aligned} G &\to P_x \\ g &\mapsto yg \end{aligned}$$

is invertible. The inverse map lets us pull the Maurer–Cartan form back to P_x in a unique way:

$$T_{xg}P \to T_g G \to \mathfrak{g}$$

Because of this canonical construction, the 1-form thus obtained on P is also called a **Maurer–Cartan form**, and denoted ω_G.

Ehresmann connections can be defined in a number of equivalent ways [11]. The definition we shall use is the following one.

Definition 3 *An* **Ehresmann connection** *on a principal H bundle*

$$\begin{array}{c} P \\ \big\downarrow{\scriptstyle \pi} \\ M \end{array}$$

is an $\mathfrak{h}$-valued 1-form ω on P

$$\omega\colon TP \to \mathfrak{h}$$

satisfying the following two properties:

1. $R_h^*\omega = \mathrm{Ad}(h^{-1})\omega$ *for all $h \in H$;*
2. *ω restricts to the Maurer–Cartan form $\omega_H\colon TP_x \to \mathfrak{h}$ on fibers of P.*

Here $R_h^*\omega$ is the pullback of ω by the right action

$$\begin{aligned} R_h\colon P &\to P \\ p &\mapsto ph \end{aligned}$$

of $h \in H$ on P, and

$$V_pP := \ker\left[d\pi_p\colon T_pP \to T_{\pi(p)}M\right]$$

is the vertical component of the tangent space T_pP.

The **curvature** of an Ehresmann connection ω is given by the familiar formula

$$\Omega[\omega] = d\omega + \frac{1}{2}[\omega, \omega]$$

where the bracket of $\mathfrak{h}$-valued forms is defined using the Lie bracket on Lie algebra parts and the wedge product on form parts.

3.2 Definition of Cartan geometry

We are ready to state the formal definition of Cartan geometry, essentially as given by Sharpe [24].

Definition 4 *A* **Cartan geometry** $(\pi\colon P \to M, A)$ *modeled on the Klein Geometry* (G, H) *is a principal right H bundle*

$$\begin{array}{c} P \\ \big\downarrow{\scriptstyle \pi} \\ M \end{array}$$

equipped with a $\mathfrak{g}$-valued 1-form A on P

$$A\colon TP \to \mathfrak{g}$$

called the **Cartan connection**, *satisfying three properties:*

0. *For each* $p \in P$, $A_p \colon T_pP \to \mathfrak{g}$ *is a linear isomorphism;*
1. $(R_h)^*A = \mathrm{Ad}(h^{-1})A \quad \forall h \in H$;
2. *A takes values in the subalgebra* $\mathfrak{h} \subseteq \mathfrak{g}$ *on vertical vectors, and in fact restricts to the Maurer–Cartan form* $\omega_H \colon TP_x \to \mathfrak{h}$ *on fibers of* P.

Compare this definition to the definition of Ehresmann connection. The most obvious difference is that the Cartan connection on P takes values not in the Lie algebra $\mathfrak{h}$ of the gauge group of the bundle, but in the larger algebra $\mathfrak{g}$. The addition of the 0th requirement in the above definition has important consequences. Most obviously, G must be chosen to have the same dimension as T_pP. In other words, the Klein geometry G/H must have the same dimension as M. In this way Cartan connections have a more "concrete" relationship to the base manifold than Ehresmann connections, which have no such dimensional restrictions. Also, the isomorphisms $A \colon T_pP \to \mathfrak{g}$ may be inverted at each point to give an injection

$$X_A \colon \mathfrak{g} \to \mathrm{Vect}(P)$$

so any element of $\mathfrak{g}$ gives a vector field on P. The restriction of X_A to the subalgebra $\mathfrak{h}$ gives vertical vector fields on P, while the restriction of X_A to a complement of $\mathfrak{h}$ gives vector fields on the base manifold M itself [1].

When the model Klein geometry G/H is a metric Klein geometry, i.e. when it is equipped with an H-invariant metric on $\mathfrak{g}/\mathfrak{h}$, M inherits this metric via the isomorphism $T_xM \cong \mathfrak{g}/\mathfrak{h}$, which comes from the isomorphism $T_pP \cong \mathfrak{g}$.

The **curvature** of a Cartan connection is given by the same formula as in the Ehresmann case:

$$F[A] = dA + \frac{1}{2}[A, A].$$

This curvature is a 2-form valued in the Lie algebra $\mathfrak{g}$. It can be composed with the canonical projection onto $\mathfrak{g}/\mathfrak{h}$:

$$\begin{tikzcd} \Lambda^2(TP) \arrow[r, "F"] \arrow[rr, bend right, "T"'] & \mathfrak{g} \arrow[r] & \mathfrak{g}/\mathfrak{h} \end{tikzcd}$$

and the composite T is called the **torsion** for reasons that will become particularly clear in Section 3.4.

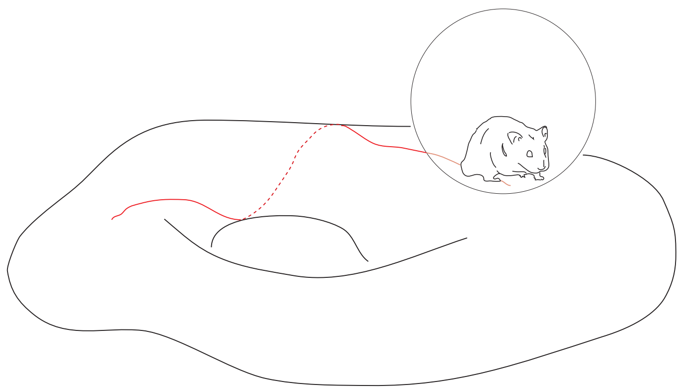
3.3 Geometric interpretation: rolling Klein geometries

In Section 1, I claimed that Cartan geometry is about "rolling the model Klein geometry on the manifold." Let us now see why a Cartan geometry on M modeled on G/H contains just the right data to describe the idea of rolling G/H on M. To understand this, we return to the example of the sphere rolling on a surface M embedded in $\mathbb{R}^3$. For this example we have

$$G = SO(3)$$
$$H = SO(2)$$

and the model space is $S^2 = \mathrm{SO}(3)/\mathrm{SO}(2)$. The Cartan geometry consists of a principal SO(2) bundle P over M together with a 1-form $\omega \in \Omega^1(M, \mathfrak{so}(3))$ satisfying the three properties above.

To understand the geometry, it is helpful to consider the situation from the point of view of an 'observer' situated at the point of tangency between the "real" space and the homogeneous model. In fact, in the rolling ball example, such an observer is easily imagined. Imagine the model sphere as a "hamster ball"—a type of transparent plastic ball designed to put a hamster or other pet rodent in to let it run around the house. But here, the hamster gets to run around on some more interesting, more lumpy surface than your living room floor, such as a Riemann surface:



It may sound silly, but in fact this is the easiest way to begin to visualize Cartan connections! In this context, what is the geometric meaning of the SO(2) bundle P in the definition of Cartan geometry? Essentially, it is best to think of P as the bundle of "hamster configurations", where a hamster configuration is specified by the hamster's position on the surface M, together with the direction the hamster is facing.

One key point, which is rather surprising on first sight, is that an element of P tells us nothing about the configuration of the rolling sphere itself. It tells us only where the hamster is, and which direction he is pointing. Naively, we might try describing the rolling of a ball on a surface using the space of all configurations of the ball itself, which would be a principal SO(3) bundle over the surface. But in fact, our principal SO(2) bundle is sufficient to describe rolling *without slipping or twisting.* This becomes obvious when we consider that the motion of the hamster completely determines the motion of the ball.

Now a Cartan connection:

$$A\colon TP \to \mathfrak{so}(3)$$

takes 'infinitesimal changes in hamster configuration' and gives infinitesimal rotations of the sphere he is sitting inside of. An 'infinitesimal change in hamster configuration' consists of a tiny rotation together with a 'transvection'— a pure translation of the point of tangency. The resulting element of $\mathfrak{so}(3)$ is the tiny rotation of the sphere, as seen by the hamster.

I now describe in detail the geometric interpretation of conditions 0, 1, and 2 in the definition of a Cartan connection, in the context of this example.

0. $A_p\colon T_pP \to \mathfrak{so}(3)$ *is a linear isomorphism.* The hamster can move in such a way as to produce any tiny rotation of the sphere desired, and he can do this in just one way. In the case of a tiny rotation that lives in the stabilizer subalgebra $\mathfrak{so}(2)$, note that the sphere's rotation is always viewed relative to the hamster: his corresponding movement is just a

tiny rotation of his body, while fixing the point of tangency to the surface. In particular, the isomorphism is just the right thing to impose a 'no twisting' constraint. Similarly, since the hamster can produce any transvection in a unique way, the isomorphism perfectly captures the idea of a 'no slipping' constraint.

1. $(R_h)^*A = \mathrm{Ad}(h^{-1})A$ *for all* $h \in \mathrm{SO}(2)$. This condition is 'SO(2)-equivariance', and may be interpreted as saying there is no absolute significance to the specific direction the hamster is pointing in. A hamster rotated by $h \in \mathrm{SO}(2)$ will get different elements of $\mathfrak{so}(3)$ for the same infinitesimal motion, but they will differ from the elements obtained by the unrotated hamster by the adjoint action of h^{-1} on $\mathfrak{so}(3)$.

2. *A restricts to the* SO(2) *Maurer-Cartan form on vertical vectors.* A vertical vector amounts to a slight rotation of the hamster inside the hamster ball, without moving the point of tangency. Using the orientation, there is a canonical way to think of a slight rotation of the hamster as an element of $\mathfrak{so}(2)$, and A assigns to such a motion precisely this element of $\mathfrak{so}(2)$.

Using this geometric interpretation, it is easy to see that the model Klein geometries themselves serve as the prototypical examples of flat Cartan geometries. Rolling a Klein geometry on *itself* amounts to simply moving the point of tangency around. Thus, just as $\mathbb{R}^n$ has a canonical way of identifying all of its linear tangent spaces, S^n has a canonical way of identifying all of its tangent spheres, H^n has a canonical way of identifying all of its tangent hyperbolic spaces, and so on.

It is perhaps worth mentioning another example—an example that is sort of 'dual' to the hamster ball rolling on a flat plane—which I find equally instructive. Rather than a hamster in a sphere, exploring the geometry of a plane, consider a person (a 15th century European, say) standing on a plane tangent to a spherical Earth. The plane rolls as she steps, the point of tangency staying directly beneath her feet. This rolling gives an ISO(2)/SO(2) Cartan geometry on the Earth's surface. She can even use the rolling motion to try drawing a local map of the Earth on the plane. As long as she doesn't continue too far, this map will even be fairly accurate.

In MacDowell–Mansouri gravity, we are in a related geometric situation. The principal $\mathrm{SO}(3,1)$ bundle describes possible event/velocity pairs for an "observer". This observer may try drawing a map of spacetime M by rolling Minkowski spacetime along M, giving an $\mathrm{ISO}(3,1)/\mathrm{SO}(3,1)$ Cartan connection. A smarter observer, if M has $\Lambda > 0$, might prefer getting an $\mathrm{SO}(4,1)/\mathrm{SO}(3,1)$ Cartan connection by rolling de Sitter spacetime along M.

3.4 Reductive Cartan geometry

The most important special case of Cartan geometry for our purposes is the 'reductive' case. Since $\mathfrak{h}$ is a vector subspace of $\mathfrak{g}$, we can always write

$$\mathfrak{g} \cong \mathfrak{h} \oplus \mathfrak{g}/\mathfrak{h}$$

as vector spaces. A Cartan geometry is said to be **reductive** if this direct sum is $\mathrm{Ad}(H)$-invariant. A reductive Cartan connection A may thus be written as

$$A = \omega + e \qquad \begin{array}{l} \omega \in \Omega^1(P, \mathfrak{h}) \\ e \in \Omega^1(P, \mathfrak{g}/\mathfrak{h}) \end{array}$$

Diagrammatically:

$$\begin{tikzcd} & & \mathfrak{h} \\ TP \arrow[r, "A"] \arrow[urr, bend left, "\omega"] \arrow[drr, bend right, "e"'] & \mathfrak{g} \arrow[ur] \arrow[dr] & \\ & & \mathfrak{g}/\mathfrak{h} \end{tikzcd}$$

It is easy to see that the $\mathfrak{h}$-valued form ω is simply an Ehresmann connection on P, and we interpret the $\mathfrak{g}/\mathfrak{h}$-valued form e as a generalized **coframe field**.

The concept of a reductive Cartan connection provides a geometric foundation for the MacDowell–Mansouri action. In particular, it gives global meaning to the trick of combining the local connection and coframe field 1-forms of general relativity into a connection valued in a larger Lie algebra. Physically, for theories like MacDowell–Mansouri, the reductive case is most important because gauge transformations of the principal H bundle act on $\mathfrak{g}$-valued forms via the adjoint action. The $\mathrm{Ad}(H)$-invariance of the decomposition says gauge transformations do not mix up the 'connection' parts with the 'coframe' parts of a reductive Cartan connection.

One can of course use the $\mathrm{Ad}(H)$-invariant decomposition of $\mathfrak{g}$ to split any other $\mathfrak{g}$-valued differential form into $\mathfrak{h}$ and $\mathfrak{g}/\mathfrak{h}$ parts. Most importantly, we can split the curvature F of the Cartan connection A:

$$\begin{tikzcd} & & \mathfrak{h} \\ \Lambda^2(TP) \arrow[r, "F"] \arrow[urr, bend left, "\hat{F}"] \arrow[drr, bend right, "T"'] & \mathfrak{g} \arrow[ur] \arrow[dr] & \\ & & \mathfrak{g}/\mathfrak{h} \end{tikzcd}$$

The $\mathfrak{g}/\mathfrak{h}$ part T is the **torsion**. The $\mathfrak{h}$ part $\hat{F}$ is related to the curvature of the Ehresmann connection ω, but there is an important difference: *The 'curvature' $\hat{F}$ is the Ehresmann curvature modified in such a way that the model Klein geometry becomes the standard for 'flatness'.*

To see that this is true, consider a reductive Klein geometry G/H. The canonical G/H Cartan connection on the principal bundle $G \to G/H$ is the Maurer–Cartan form ω_G, and this splits in the reductive case into two parts $\omega_G = \omega_H + e$:

$$\begin{tikzcd} & & \mathfrak{h} \\ TG \arrow[r, "\omega_G"] \arrow[urr, bend left, "\omega_H"] \arrow[drr, bend right, "e"'] & \mathfrak{g} \arrow[ur] \arrow[dr] & \\ & & \mathfrak{g}/\mathfrak{h} \end{tikzcd}$$

The well-known 'structural equation' for the Maurer–Cartan form,

$$d\omega_G = -\frac{1}{2}[\omega_G, \omega_G], \tag{6}$$

is interpreted in this context as the statement of vanishing Cartan curvature. In particular, this means both parts of the curvature vanish.

Let us work out the curvature in the cases most relevant to gravity. The six Klein model spacetimes listed in Section 2—de Sitter, Minkowski, anti de Sitter, and their Riemannian analogs—are all reductive.

For any of these models, the Cartan connection is an $\mathfrak{g}$-valued 1-form A on a principal H bundle, which we take to be the frame bundle FM on spacetime:

$$A \in \Omega^1(FM, \mathfrak{g}).$$

We identify $\mathfrak{g}/\mathfrak{h}$ with Minkowski vector space $\mathbb{R}^{3,1}$ (or Euclidean $\mathbb{R}^4$ in the Riemannian cases) by picking a unit of length ℓ.

In index notation, we write the two parts of the connection as

$$A^i{}_j = \omega^i{}_j \quad \text{and} \quad A^i{}_4 = \frac{1}{\ell} e^i.$$

This gives

$$A^4{}_j = \frac{-\epsilon}{\ell}\, e_j,$$

where ϵ is chosen according to the choice of $\mathfrak{g}$, by (4). We use these components to calculate the two parts of the curvature

$$F^I{}_J = dA^I{}_J + A^I{}_K \wedge A^K{}_J$$

as follows. For the $\mathfrak{so}(3,1)$ part:

$$\begin{aligned} F^i{}_j &= dA^i{}_j + A^i{}_k \wedge A^k{}_j + A^i{}_4 \wedge A^4{}_j \\ &= d\omega^i{}_j + \omega^i{}_k \wedge \omega^k{}_j - \frac{\epsilon}{\ell^2}\, e^i \wedge e_j \\ &= R^i{}_j - \frac{\epsilon}{\ell^2}\, e^i \wedge e_j \end{aligned}$$

where R is the curvature of the $\mathrm{SO}(3,1)$ Ehresmann connection ω; for the $\mathbb{R}^{3,1}$ part:

$$\begin{aligned} F^i{}_4 &= dA^i{}_4 + A^i{}_k \wedge A^k{}_4 \\ &= \frac{1}{\ell}\left(de^i + \omega^i{}_k \wedge e^k\right) \\ &= \frac{1}{\ell}\, d_\omega e^i. \end{aligned}$$

The same calculations hold formally in the Riemannian analogs as well, the only difference being that indices are lowered with δ_{ij} rather than η_{ij}.

This has a remarkable interpretation. The above calculations give the condition for a Cartan connection $A = \omega + e$ based on any of our six models to be flat:

$$F = 0 \qquad \Longleftrightarrow \qquad R - \frac{\epsilon}{\ell^2} e \wedge e = 0 \ \text{ and } \ d_\omega e = 0$$

The second condition says ω is torsion-free. The first says says not that ω is flat, but that it is homogeneous with cosmological constant

$$\Lambda = \frac{3\epsilon}{\ell^2}$$

In other words, A is flat when ω is the Levi-civita connection for a universe with only cosmological curvature, and the cosmological constant matches the **internal cosmological constant**—the cosmological constant (5) of the model homogeneous spacetime. Indeed, the Maurer–Cartan form ω_G is a Cartan connection for the model spacetime, and the structural equation (6) implies $\lambda = 3\epsilon/\ell^2$ is the cosmological constant of the model.

The point here is that one *could* try describing spacetime with cosmological constant Λ using a model spacetime with $\lambda \neq \Lambda$, but this is not the most natural thing to do. But in fact, this is what is done all the time when we use semi-Riemannian geometry ($\lambda = 0$) to describe spacetimes with nonzero cosmological constant.

If we agree to use a model spacetime with cosmological constant Λ, the parts of a reductive connection and its curvature can be summarized diagrammatically in the three Lorentzian cases as follows:

$$\begin{tikzcd}
 & & \mathfrak{so}(3,1) \\
T(FM) \arrow[r, "A"] \arrow[urr, bend left, "\omega"] \arrow[drr, bend right, "\frac{1}{\ell}e"'] & \mathfrak{g} \arrow[ur] \arrow[dr] & \\
 & & \mathbb{R}^{3,1}
\end{tikzcd}
\qquad
\begin{tikzcd}
 & & \mathfrak{so}(3,1) \\
\Lambda^2(T(FM)) \arrow[r, "F"] \arrow[urr, bend left, "R-\frac{\Lambda}{3}e\wedge e"] \arrow[drr, bend right, "\frac{1}{\ell}d_\omega e"'] & \mathfrak{g} \arrow[ur] \arrow[dr] & \\
 & & \mathbb{R}^{3,1}
\end{tikzcd}$$

where ℓ and Λ are related by the equation

$$\ell^2\Lambda = 3\epsilon.$$

As observed earlier, for $\Lambda = \epsilon = 0$ the value of ℓ^2 is not constrained by the cosmological constant, so there is an additional scaling symmetry in Cartan geometry modelled on Minkowski or Euclidean spacetime.

As a final note on reductive Cartan geometries, in terms of the constituent fields ω and e, the **Bianchi identity**

$$d_A F = 0$$

for a reductive Cartan connection A breaks up into two parts. One can show that these two parts are the Bianchi identity for ω and another familiar identity:

$$d_\omega R = 0 \qquad d_\omega^2 e = R \wedge e.$$

4 Cartan-type gauge theory

Part of the case I wish to make is that gravity—particularly in MacDowell–Mansouri-like formulations—should be seen as based on a type of gauge theory where the connection is not an Ehresmann connection but a Cartan connection. Unlike gauge fields in 'Ehresmann-type' gauge theories, like Yang-Mills theory, the gravitational field does not encode purely 'internal' degrees of freedom. Cartan connections give a concrete correspondence between spacetime and a Kleinian model, in a way that is ideally suited to a geometric theory like gravity.

In this section, I discuss issues—such as holonomy and parallel transport—relevant to doing gauge theory with a Cartan connection as the gauge field. As it turns out, some of these issues are clarified by considering associated bundles of the Cartan geometry.

4.1 A sequence of bundles

Just as Klein geometry involves a sequence of H-spaces:

$$H \to G \to G/H,$$

Cartan geometry can be seen as involving the induced sequence of bundles:

$$\begin{tikzcd}
P \arrow[r, "\iota"] \arrow[dr, "\substack{\text{principal}\\ H\text{ bundle}}"'] & P \times_H G \arrow[r] \arrow[d, "\substack{\text{principal}\\ G\text{ bundle}}"] & P \times_H G/H \arrow[dl, "\substack{\text{bundle of tangent}\\ \text{Klein geometries}}"] \\
& M &
\end{tikzcd}$$

The bundle $Q = P \times_H G \to M$ is associated to the principal H bundle P via the action of H by left multiplication on G. This Q is a principal right G bundle, and the map

$$\begin{aligned}
\iota\colon P &\to P \times_H G \\
p &\mapsto [p, 1_G]
\end{aligned}$$

is a canonical inclusion of H bundles. I call the associated bundle $\kappa\colon P \times_H G/H \to M$, the **bundle of tangent Klein geometries**. This is an appropriate name, since it describes a bundle over M whose fibers are copies of the Klein geometry G/H, each with a natural 'point tangency'. Explicitly, for $x \in M$, the **Klein geometry tangent to M at x** is the fiber $\kappa^{-1}x$, and the **point of tangency** in this tangent geometry is the equivalence class $[p, H]$ where p is any point in P_x and H is the coset of the identity. This is well defined since any other 'point of tangency' is of the form $[ph, H] = [p, hH] = [p, H]$, where $h \in H$.

There is an interesting correspondence between Cartan connections on P and Ehresmann connections on $Q = P \times_H G$. To understand this correspondence, we introduce the notion of a **generalized Cartan connection** [1], in which we replace the 0th requirement in Definition 4, that $A_p\colon T_pP \to \mathfrak{g}$ be an isomorphism, by the weaker requirement that T_p and $\mathfrak{g}$ have the same dimension. It is not hard to show that if

$$\tilde{A}\colon TQ \to \mathfrak{g}$$

is an Ehresmann connection on Q then

$$A := \iota^*\tilde{A}\colon TP \to \mathfrak{g}$$

is a generalized Cartan connection on P. In fact, given a generalized Cartan connection on P, there is a unique Ehresmann connection $\tilde{A}$ on Q such that $A = \iota^*\tilde{A}$, so the generalized Cartan connections on P are in one-to-one correspondence with Ehresmann connections on Q [1]. Moreover, the generalized Cartan connection A associated to an Ehresmann connection $\tilde{A}$ on Q is a Cartan connection if and only if $\ker\tilde{A} \cap \iota_*(TP) = 0$ [24].

4.2 Parallel transport in Cartan geometry

How does an observer in a spacetime of positive cosmological constant decide how much her universe deviates from de Sitter spacetime? From the Cartan perspective, one way is to do parallel transport in the bundle of tangent de Sitter spacetimes.

There are actually two things we might mean by 'parallel transport' in Cartan geometry. First, if the geometry is reductive, then the $\mathfrak{h}$ part of the G/H-Cartan connection is an Ehresmann connection ω. We can use this Ehresmann connection to do parallel transport in the bundle of tangent Klein geometries in the usual way. Namely, if

$$\gamma\colon [t_0, t_1] \to M$$

is a path in the base manifold, and $[p, gH]$ is a point in the tangent Klein geometry at $\gamma(t_0)$, then the translation of $[p, gH]$ along γ is

$$[\tilde{\gamma}(t), gH]$$

where $\tilde{\gamma}$ is the horizontal lift of γ starting at $p \in P$. However, this method, aside from being particular to the reductive case, is also not the sort of parallel transport that is obtained by rolling the model geometry, as in our intuitive picture of Cartan geometry. In particular, the translation of the point of tangency $[p, H]$ of the tangent Klein geometry at $x = \gamma(t_0) \in M$ is always just the point of tangency at in the tangent Klein geometry at $\gamma(t)$. This is expected, since the gauge group H only acts in ways that stabilize the basepoint. We would like to describe a sort of parallel transport that does not necessarily fix the point of tangency.

The more natural notion of parallel transport in Cartan geometry, does not require the geometry to be reductive. A Cartan connection cannot be used in the same way as an Ehresmann connection to do parallel transport, because Cartan connections do not give 'horizontal lifts'. Horizontal subspaces are given by the kernel of an Ehresmann connection; Cartan connections have no kernel. To describe the general notion of parallel transport in a Cartan geometry, we make use of the associated Ehresmann connection.

To understand the general parallel transport in Cartan geometry is to observe that we have a canonical isomorphism of fiber bundles

$$\begin{tikzcd} P \times_H G/H \arrow[rr, "\cong"] \arrow[dr] & & Q \times_G G/H \arrow[dl] \\ & M & \end{tikzcd}$$

where $Q = P \times_H G$ is the principal G bundle associated to P, as in the previous section. To see this, note first that the H bundle inclusion map $\iota\colon P \to Q$ induces an inclusion of the associated bundles by

$$\begin{aligned} \iota'\colon P \times_H G/H &\to Q \times_G G/H \\ [p, gH] &\mapsto [\iota(p), gH]. \end{aligned}$$

This bundle map has an inverse which we construct as follows. An element of $Q \times_G G/H = P \times_H G \times_G G/H$ is a an equivalence class $[p, g', gH]$, with $p \in P$, $g' \in G$, and $gH \in G/H$. Any such element can be written as $[p, 1, g'gH]$, so we can define a map that simply drops this "1" in the middle:

$$\begin{aligned}\beta\colon Q \times_G G/H &\to P \times_H G/H \\ [p, g', gH] &\mapsto [p, g'gH].\end{aligned}$$

It is easy to check that this is a well-defined bundle map, and

$$\begin{aligned}\beta\iota'[p, gH] &= \beta[\iota(p), gH] = \beta[p, 1, gH] = [p, gH] \\ \iota'\beta[p, g', gH] &= \iota'[p, g'gH] = [p, 1, g'gH] = [p, g', gH]\end{aligned}$$

so $\iota' = \beta^{-1}$ is a bundle isomorphism. While these are isomporphic as fiber bundles, the isomorphism is not an isomorphism of associated bundles (in the sense described by Isham [17]), since it does not come from an isomorphism of the underlying principal bundles. In fact, while $P \times_H G/H$ and $Q \times_G G/H$ are isomorphic as fiber bundles, there is a subtle difference between them two descriptions: the latter bundle does not naively have a natural 'point of tangency' in each fiber, except via the isomorphism ι'.

Given the above isomorphism of fiber bundles, and given the associated Ehresmann connection defined in the previous section, we have a clear prescription for parallel transport. Namely, given any $[p, gH]$ in the tangent Klein geometry at $x = \gamma(t_0) \in M$, we think of this point as a point in $Q \times_G G/H$, via the isomorphism ι', use the Ehresmann connection on Q to translate along γ, then turn the result back into a point in the bundle of tangent Klein geometries, $P \times_H G/H$, using β. That is, the parallel transport is

$$\beta([\hat{\gamma}(t), gH])$$

where

$$\hat{\gamma}\colon [t_0, t_1] \to Q$$

is the horizontal lift of $\gamma\colon [t_0, t_1] \to M$ starting at $\iota(p) \in Q$, with respect to the Ehresmann connection associated with the Cartan connection on P. Note that this sort of parallel transport need not fix the point of tangency.

Holonomy and development

Just as a Cartan geometry has two notions of parallel translation, it also has two notions of holonomy, taking values either G or H. Whenever the geometry is reductive, we can take the holonomy along a loop using the Ehresmann connection part of the Cartan connection. This gives a holonomy for each loop with values in H. In fact, without the assumption of reductiveness, there is a general notion of this H holonomy, which is I shall not describe. In general there is a topological obstruction to defining this type of holonomy of a Cartan connection: it is not defined for all loops in the base manifold, but only those loops that are the images of loops in the principal H bundle [24].

The other notion of holonomy, with values in G, can of course can be calculated by relying on the associated Ehresmann connection on $Q = P \times_H G$.

Besides holonomies around loops, a Cartan connection gives a notion of 'development on the model Klein geometry'. Suppose we have a Cartan connection A on $P \to M$ and a piecewise-smooth path in P,

$$\gamma\colon [t_0, t_1] \to P,$$

lifting a chosen path in M. Given any element $g \in G$, the **development of γ on G** starting at g is the unique path

$$\gamma_G\colon [t_0, t_1] \to G$$

such that $\gamma(t_0) = g$ and $\gamma^* A = \gamma_G^* \omega_G \in \Omega^1([t_0, t_1], \mathfrak{g})$. To actually calculate the development, one can use the path-ordered exponential

$$\gamma_G(t) = P e^{-\int_{t_0}^{t} \omega(\tilde{\gamma}'(s)) ds} \in G.$$

Composing γ_G with the quotient map $G \to G/H$ gives a path on the model Klein geometry:

$$\gamma_{G/H}\colon [t_0, t_1] \to G/H.$$

called the **development of γ on G/H** starting at gH. This path is independent of the lifting γ, depending only on the path in the base manifold M. [24]

In the SO(3)/SO(2) example of Section 3.3, the development is the path traced out on the ball itself by the point of tangency on the surface, as the ball rolls.

4.3 BF Theory and flat Cartan connections

As an example of a gauge theory with Cartan connection, let us consider using a Cartan connection in the topological gauge theory known as 'BF theory' [4]. Special cases of such a theory have already been considered by Freidel and Starodubtsev, in connection with MacDowell–Mansouri gravity [13], but without the explicit Cartan-geometric framework.

In ordinary BF theory with gauge group H, on n-dimensional spacetime, the fields are an Ehresmann connection A on a principal H bundle P, and an $\mathrm{Ad}(P)$-valued $(n-2)$-form B, where

$$\mathrm{Ad}(P) = P \times_H \mathfrak{h}$$

is the vector bundle associated to P via the adjoint representation of H on its Lie algebra. Denoting the curvature of A by F, the BF theory action

$$S_{BF} = \int \mathrm{tr}\,(B \wedge F)$$

leads to the equations of motion:

$$F = 0$$
$$d_A B = 0.$$

That is, the connection A is flat, and the field B is covariantly constant.

We wish to copy this picture as much as possible using a Cartan connection of type G/H in place of the Ehresmann H-connection. Doing so requires, first of all, picking a Klein model

G/H of the same dimension as M. For the B field, the obvious analog is an $(n-2)$-form with values in the bundle

$$\mathrm{Ad}_{\mathfrak{g}}(P) := P \times_H \mathfrak{g}$$

where H acts on $\mathfrak{g}$ via the restriction of the adjoint representation of G. Formally, we obtain the same equations of motion

$$\begin{aligned} F &= 0 \\ d_A B &= 0. \end{aligned}$$

but these must now be interpreted in the Cartan-geometric context.

In particular, the equation $F = 0$ says the Cartan connection is flat. In other words, 'rolling' the tangent Klein geometry on spacetime is trivial, giving an isometric identification between any contractible neighborhood in spacetime and a neighborhood of the model geometry G/H. Of course, the rolling can still give nontrivial holonomy around noncontractible loops. This indicates that solutions of Cartan-type BF theory are related to 'geometric structures' [26], which have been used to study a particular low-dimensional case of BF theory, namely, 3d quantum gravity [7].

Let us work out a more explicit example: Cartan-type BF theory based on one of the Lorentzian reductive models discussed in Sections 2.3 and 3.4. Since the geometry is reductive, we can decompose our $\mathfrak{g}$-valued fields into A, F, and B into $\mathfrak{so}(3,1)$ and $\mathbb{R}^{3,1}$ parts. We already know how to do this for A and F. For B, let $b = \hat{B}$ denote the $\mathfrak{so}(3,1)$ part, and x the $\mathbb{R}^{3,1}$ part, so

$$B^i{}_j = b^i{}_j \qquad B^i{}_4 = \frac{1}{\ell} x^i.$$

Note that this gives

$$B^4{}_j = -\frac{\epsilon}{\ell} x_j.$$

with ϵ chosen by the sign of the cosmological constant according to (4). We need to know how to write $d_A B$ in terms of these component fields. We know that

$$\begin{aligned} d_A B^{IJ} :&= dB^{IJ} + [A, B]^{IJ} \\ &= dB^{IJ} + A^I{}_K \wedge B^{KJ} - B^I{}_K \wedge A^{KJ}, \end{aligned}$$

so for both indices between 0 and 3 we have

$$\begin{aligned} d_A B^{ij} :&= dB^{ij} + A^i{}_k \wedge B^{kj} - B^i{}_k \wedge A^{kj} + A^i{}_4 \wedge B^{4j} - B^i{}_4 \wedge A^{4j} \\ &= d_\omega b^{ij} - \frac{\epsilon}{\ell} e^i \wedge x^j + \frac{\epsilon}{\ell} x^i \wedge e^j \end{aligned}$$

and for an index 4,

$$\begin{aligned} d_A B^{i4} = d_A b^{i4} &= dB^{i4} + A^i{}_k \wedge B^{k4} - B^i{}_k \wedge A^{k4} \\ &= d_\omega x^i - \frac{1}{\ell} b^i{}_k \wedge e^k. \end{aligned}$$

The equations for BF theory with Cartan connection based on de Sitter, Minkowski, or anti de Sitter model geometry are thus

$$\begin{aligned} R - \frac{\epsilon}{\ell^2} e \wedge e &= 0 \\ d_\omega e &= 0 \\ d_\omega b + \frac{\epsilon}{\ell^2}(x \wedge e - e \wedge x) &= 0 \\ d_\omega x - \frac{1}{\ell} b \wedge e &= 0 \end{aligned}$$

In terms of the constituent fields of the reductive geometry, classical Cartan-type BF theory is thus described by the Levi–Civita connection on a spacetime of purely cosmological curvature, with constant $\Lambda = 3\epsilon/\ell^2$, together with an pair of auxiliary fields b and x, satisfying two equations. We shall encounter equations very similar to these in the BF reformulation of MacDowell–Mansouri gravity.

5 From Palatini to MacDowell–Mansouri

In this section, I show how thinking of the standard Palatini formulation of general relativity in terms of Cartan geometry leads in a natural way to the MacDowell–Mansouri formulation.

5.1 The Palatini formalism

Before describing the MacDowell–Mansouri approach, I briefly recall in this section the better-known Palatini formalism. The main purpose in doing this is to firmly establish the global differential geometric setting of the Palatini approach, in order to compare to that of MacDowell–Mansouri. Experts may safely skip ahead after skimming to fix notation.

The Palatini formalism de-emphasizes the metric g on spacetime, making it play a subordinate role to the **coframe field** e, a vector bundle morphism:

$$\begin{tikzcd} TM \arrow[rr, "e"] \arrow[dr, "p"'] & & \mathcal{T} \arrow[dl, "\pi"] \\ & M & \end{tikzcd}$$

Here $\mathcal{T}$ is the **fake tangent bundle** or **internal space**—a bundle over spacetime M which is isomorphic to the tangent bundle TM, but also equipped with a fixed metric η. The name coframe field comes from the case where TM is trivializable, and $e\colon TM \to \mathcal{T} = M \times \mathbb{R}^{3,1}$ is a choice of trivialization. In this case e restricts to a coframe $e_x\colon T_xM \to \mathbb{R}^{3,1}$ on each tangent space. In any case, since $\mathcal{T}$ is *locally* trivializable, we can treat e locally as an $\mathbb{R}^{3,1}$-valued 1-form.

The tangent bundle acquires a metric by pulling back the metric on $\mathcal{T}$:

$$g(v, w) := \eta(ev, ew)$$

for any two vectors in the same tangent space T_xM. In index notation, this becomes $g_{\alpha\beta} = e^a_\alpha e^b_\beta \eta_{ab}$. In the case where the metric g corresponds to a classical solution of general relativity,

$e\colon TM \to \mathcal{T}$ is an *isomorphism*, so that g is nondegenerate. However, the formalism makes sense when e is any bundle morphism, and attempting path-integral quantization of the theory makes it unnatural to impose a nondegeneracy constraint.

When e is an isomorphism, we can also pull a connection ω on the vector bundle $\mathcal{T}$ back to a connection on TM as follows. Working in coordinates, the covariant derivative of a local section s of $\mathcal{T}$ is

$$(D_\mu s)^a = \partial_\mu s^a + \omega^a_{\mu b} s^b$$

When e is an isomorphism, we can use D to differentiate a section w of TM in the obvious way: use e to turn w into a section of $\mathcal{T}$, differentiate this section, and use e^{-1} to turn the result back into a section of TM. This defines a connection on TM by:

$$\nabla_v w = e^{-1} D_v e w$$

for any vector field v. In particular, if $v = \partial_\mu$, $\nabla_\mu := \nabla_{\partial_\mu}$, we get, in index notation:

$$(\nabla_\mu w)^\alpha = \partial_\mu w^\alpha + \Gamma^\alpha_{\mu\beta} w^\beta$$

where

$$\Gamma^\alpha_{\mu\beta} := e^\alpha_a (\delta^a_b \partial_\mu + A^a_{\mu b}) e^b_\beta$$

The Palatini action is

$$S_{\mathrm{Pal}}(\omega, e) = \frac{1}{2G} \int_M \mathrm{tr} \left(e \wedge e \wedge R + \frac{\Lambda}{6}\, e \wedge e \wedge e \wedge e \right). \tag{7}$$

where R is the curvature of ω and the wedge product $\wedge$ denotes antisymmetrization on both spacetime indices and internal Lorentz indices. Compatibility with the metric η forces the curvature R to take values in $\Lambda^2\mathcal{T}$. Hence, the expression in parentheses is a $\Lambda^4\mathcal{T}$-valued 1-form on M, and the 'trace' is really a map that turns such a form into an ordinary 1-form using the volume form on the internal space $\mathcal{T}$:

$$\mathrm{tr} : \Omega(M, \Lambda^4\mathcal{T}) \to \Omega(M, \mathbb{R})$$

For computations, the action is often written leaving internal indices in as:

$$S_{\mathrm{Pal}}(\omega, e) = \frac{1}{2G} \int_M \left(e^i \wedge e^j \wedge R^{k\ell} + \frac{\Lambda}{6}\, e^i \wedge e^j \wedge e^k \wedge e^\ell \right) \epsilon_{ijk\ell}.$$

The variations of ω and e give us the respective equations of motion

$$d_\omega(e \wedge e) = 0 \tag{8}$$

$$e \wedge R + \frac{\Lambda}{3}\, e \wedge e \wedge e = 0. \tag{9}$$

In the classical case where e is an isomorphism, the first of these equations is equivalent to

$$d_\omega e = 0$$

which says precisely that the induced connection on TM is torsion free, hence that $\Gamma^\alpha_{\mu\beta}$ is the Christoffel symbol for the Levi–Civita connection. The other equation of motion, rewritten in terms of the metric and Levi–Civita connection, is Einstein's equation.

5.2 The coframe field

Let us study the precise sense in which the field

$$e\colon TP \to \mathfrak{g}/\mathfrak{h}$$

in a reductive Cartan geometry is a generalization of the coframe field

$$e\colon TM \to \mathcal{T}$$

used in the Palatini formulation of general relativity. The latter is a $\mathcal{T}$-valued 1-form on spacetime. The former seems superficially rather different: a 1-form not on spacetime M, but on some principal bundle P over M, with values not in a vector bundle, but in a mere vector space $\mathfrak{g}/\mathfrak{h}$.

To understand the relationship between these, we first note that from the Cartan perspective, there is a natural choice of fake tangent bundle $\mathcal{T}$. To be concrete, Consider the case of Cartan geometry modeled on de Sitter spacetime, so $G = \mathrm{SO}(4,1)$, $H = \mathrm{SO}(3,1)$. The frame bundle $FM \to M$ is a principal H bundle, and the Lie algebra $\mathfrak{g} = \mathfrak{so}(4,1)$ has an $\mathrm{Ad}(H)$-invariant splitting $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{g}/\mathfrak{h}$, so the geometry is reductive. As we have seen, we can pick an invariant metric on $\mathfrak{g}/\mathfrak{h}$ of signature $(-{+}{+}{+})$ which is invariant under the adjoint representation of H. This representation gives us an associated bundle of FM, which we take as the **fake tangent bundle**:

$$\mathcal{T} := FM \times_H \mathfrak{g}/\mathfrak{h}.$$

This is isomorphic, as a vector bundle, to the tangent bundle TM, but is equipped with a metric induced by the metric on $\mathfrak{g}/\mathfrak{h}$. As I explain below, with this choice of $\mathcal{T}$, the two versions of the 'coframe field' are in fact equivalent ways of describing the same field, given an Ehresmann connection on the frame bundle. Since the geometry is reductive, the $\mathfrak{so}(3,1)$ part of the Cartan connection is such an Ehresmann connection.

In the more general case, where the principal H bundle P is not necessarily the frame bundle, we may consider the generalized coframe field either as an H-equivariant 1-form on P, or as a 1-form on M with values in the associated bundle:

$$e\colon TP \to \mathfrak{g}/\mathfrak{h} \qquad \text{or} \qquad e\colon TM \to P \times_H \mathfrak{g}/\mathfrak{h}.$$

provided we have an Ehresmann connection on P, such as the $\mathfrak{h}$ part of a reductive Cartan connection.

To prove the equivalence of these two perspectives on the coframe field, suppose we have an Ehresmann connection ω on a principal H bundle $p\colon P \to M$, and a Lie algebra $\mathfrak{g} \supset \mathfrak{h}$. Given a 1-form $e\colon TM \to P \times_H \mathfrak{g}/\mathfrak{h}$ valued in the associated bundle, we wish to construct an H-equivariant 1-form $\tilde{e}\colon TP \to \mathfrak{g}/\mathfrak{h}$. For any $v \in T_yP$, taking $e(d\pi(v))$ gives an element $[y', X] \in P \times_H \mathfrak{g}/\mathfrak{h}$. This element is by definition an equivalence class such that $[y', X] = [y'h, \mathrm{Ad}(h^{-1})X]$ for all $h \in H$. We thus define $\tilde{e}(v)$ for $v \in T_yP$ to be the unique element of $\mathfrak{g}/\mathfrak{h}$ such that $e(d\pi(v)) = [y, \tilde{e}(v)]$. This construction makes $\tilde{e}$ equivariant with respect to the actions of H, since on one hand

$$e(d\pi(v)) = [y, \tilde{e}(v)] = [yh, \mathrm{Ad}(h^{-1})\tilde{e}(v)],$$

while on the other

$$e(d\pi(v)) = e(d\pi(R_{h*}v)) = [yh, \tilde{e}(R_{h*}v)] = [yh, R_h^*\tilde{e}(v)],$$

so that

$$R_h^*\tilde{e}(v) = \mathrm{Ad}(h^{-1})\tilde{e}(v).$$

Conversely, given the equivariant 1-form $\tilde{e}\colon TP \to \mathfrak{g}/\mathfrak{h}$, define $e\colon TM \to P \times_H \mathfrak{g}/\mathfrak{h}$ as follows. If $v \in T_xM$, pick any $y \in p^{-1}(x)$ and let $\tilde{v}_y \in T_yP$ be the unique horizontal lift of v relative to the connection ω. Then let $e(v) = [y, \tilde{e}(\tilde{v}_y)] \in FM \times_H \mathfrak{g}/\mathfrak{h}$. This is well-defined, since for any other $y' \in p^{-1}(x)$, we have $y' = yh$ for some $h \in H$, and hence

$$[y', \tilde{e}(\tilde{v}_{y'})] = [yh, R_h^*\tilde{e}(v_y)] = [yh, \mathrm{Ad}(h^{-1})\tilde{e}(v_y)] = [y, \tilde{e}(v_y)]$$

where the second equality is equivariance and the third follows from the definition of the associated bundle $FM \times_H \mathfrak{g}/\mathfrak{h}$. It is straightforward to show that the construction of $\tilde{e}$ from e and vice-versa are inverse processes, so we are free to regard the coframe field e in either of these two ways.

As mentioned in the previous section, for applications to quantum gravity it is often best to allow degenerate coframe fields, which don't correspond to classical solutions of general relativity. Here I merely point out that the remarks of this section still hold for possibly degenerate coframe fields, provided we replace the Cartan connection with a generalized Cartan connection, as defined in Section 4.1.

5.3 MacDowell–Mansouri gravity

Using results of the previous section, the Palatini action for general relativity can be viewed in terms of Cartan geometry, simply by thinking of the coframe field and connection as parts of a Cartan connection $A = \omega + e$. However, in its usual form:

$$S_{\mathrm{Pal}} = \frac{1}{2G}\int (e^i \wedge e^j \wedge R^{k\ell} - \frac{\Lambda}{6} e^i \wedge e^j \wedge e^k \wedge e^\ell)\epsilon_{ijk\ell}$$

the action is not written directly in terms of the Cartan connection. The MacDowell–Mansouri action can be seen as a rewriting of the Palatini action that makes the underlying Cartan-geometric structure more apparent.

To obtain the MacDowell–Mansouri action, let us begin by rewriting the Palatini action (7) using the internal Hodge star operator as

$$S_{\mathrm{Pal}} = \frac{-1}{G}\int \mathrm{tr}\left((e \wedge e \wedge \star R - \frac{\Lambda}{6} e \wedge e \wedge \star(e \wedge e)\right). \tag{10}$$

Here we are using the isomorphism $\Lambda^2\mathbb{R}^4 \cong \mathfrak{so}(3,1)$ to think of both R and $e \wedge e$ as $so(3,1)$-valued 2-forms, and using Hodge duality in $\mathfrak{so}(3,1)$, as described in the Appendix. In each of the models discussed in Section 2.3, the $\mathfrak{so}(3,1)$ or $\mathfrak{so}(4)$ part of the curvature of the reductive Cartan connection, with appropriate internal cosmological constant, is given by

$$\hat{F} = R - \frac{\Lambda}{3} e \wedge e,$$

When $\Lambda \neq 0$, this gives us an expression for $e \wedge e$ which can be substituted into the Palatini action to obtain

$$\begin{aligned} S &= \frac{-1}{G} \int \mathrm{tr}\, \left(\frac{3}{\Lambda}(R - \hat{F}) \wedge \star R - \frac{3}{2\Lambda}(R - \hat{F}) \wedge \star(R - \hat{F}) \right) \\ &= \frac{-3}{2G\Lambda} \int \mathrm{tr}\, \left(\hat{F} \wedge \star\hat{F} + R \wedge \star R \right). \end{aligned}$$

The $R \wedge R$ term here is a topological invariant, having vanishing variation due to the Bianchi identity. For physical purposes, we may thus discard this term. If we also recognize that $\hat{F} \wedge \star\hat{F} = F \wedge \star\hat{F}$, we obtain the **MacDowell–Mansouri action** (2) as I presented it in Section 1:

$$S_{\mathrm{MM}} = \frac{-3}{2G\Lambda} \int \mathrm{tr}\, (F \wedge \star\hat{F})$$

The BF reformulation of MacDowell–Mansouri introduced by Freidel and Starodubtsev is given by the action

$$S = \int \mathrm{tr}\, \left(B \wedge F - \frac{\alpha}{2} B \wedge \star\hat{B} \right).$$

where

$$\alpha = \frac{G\Lambda}{3}$$

Calculating the variation, we get:

$$\begin{aligned} \delta S &= \int \mathrm{tr}\, (\delta B \wedge (F - \alpha\star\hat{B}) + B \wedge \delta F) \\ &= \int \mathrm{tr}\, (\delta B \wedge (F - \alpha\star\hat{B}) + d_A B \wedge \delta A) \end{aligned}$$

where in the second step we use the identity $\delta F = d_A \delta A$ and integration by parts. The equations of motion resulting from the variations of B and A are thus, respectively,

$$F = \alpha\star\hat{B} \tag{11}$$

$$d_A B = 0 \tag{12}$$

Why are these the equations of general relativity? Freidel and Starodubtsev approach this question indirectly, substituting the first equation of motion back into the Lagrangian to eliminate the B field. If we do this, and note that $\star^2 = -1$, we obtain

$$\begin{aligned} S &= \int \mathrm{tr}\, (-\frac{1}{\alpha}\star F \wedge \hat{F} - \frac{1}{2\alpha} \star \hat{F} \wedge F) \\ &= \frac{-3}{2G\Lambda} \int \mathrm{tr}\, (F \wedge \star F) \end{aligned}$$

which is precisely the MacDowell–Mansouri action. However, it is instructive to see Einstein's equations coming directly from the equations of motion (11) and (12).

Decomposing the F and B fields into reductive components, we can rewrite the equations of motion as:

$$\begin{aligned} R - \frac{\epsilon}{\ell^2} e \wedge e &= G\Lambda \star b \\ d_\omega e &= 0 \\ d_\omega b + \frac{\epsilon}{\ell^2}(x \wedge e - e \wedge x) &= 0 \\ d_\omega x - \frac{1}{\ell} b \wedge e &= 0 \end{aligned}$$

These are strikingly similar to the equations for Cartan-type BF theory obtained in Section 4.3. Indeed, if we take $G = 0$, they are identical. This is good, because it says turning Newton's gravitational constant off turns 4d gravity into 4d Cartan-type BF theory!

But we still have a bit of work to show that these equations of motion are in fact the equations of general relativity. They can be simplified as follows. Taking the covariant differential of the first equation shows, by the Bianchi identity $d_\omega R = 0$ and the second equation of motion—the vanishing of the torsion $d_\omega e$—that

$$d_\omega \star b = 0$$

But this covariant differential passes through the Hodge star operator, as shown in the Appendix, and hence

$$d_\omega b = 0.$$

This reduces the third equation of motion to

$$e^i \wedge x^j = e^j \wedge x^i.$$

The matrix part of the form $e \wedge x$ is thus a *symmetric* matrix which lives in $\Lambda^2\mathbb{R}^4$, hence is zero. When the coframe field e is invertible, we therefore get

$$x = 0$$

and hence by the fourth equation of motion,

$$b \wedge e = 0.$$

Taking the Hodge dual of the first equation of motion and wedging with e we therefore get precisely the second equation of motion arising from the Palatini action (10):

$$\star(R - \frac{1}{\ell^2} e \wedge e) \wedge e = 0,$$

namely, Einstein's equation.

6 Conclusions and Outlook

One major point of this work has been to make a case for Cartan geometry as a means to a deeper understanding of the geometry of general relativity, particularly in the MacDowell–Mansouri formulation. Cartan geometry not only gives geometric meaning to the MacDowell–Mansouri formalism, but also deepens the connection between 4d BF theory and gravity. There are many unanswered questions raised by this work, which I must leave to future research. I list just a few issues.

From my perspective, one major issue is that the $\Lambda \to 0$ limit of MacDowell–Mansouri gravity still seems poorly understood. Cartan geometry based on the flat Minkowski model $\mathrm{ISO}(3,1)/\mathrm{SO}(3,1)$ is as viable as de Sitter or anti de Sitter models, and Cartan-type BF theory based on this model makes perfect sense. As $\Lambda \to 0$, the de Sitter group undergoes a Wigner contraction to the Poincaré group, so from the Cartan-geometric perspective, the $\Lambda \to 0$ limit of MacDowell–Mansouri gravity should be "MacDowell–Mansouri with gauge group $\mathrm{ISO}(3,1)$". However, this choice of gauge group in the action (2) does not naively give general relativity with $\Lambda = 0$, but a mere topological theory. Is there a sense in which the limit of MacDowell–Mansouri gravity as $\Lambda \to 0$ is full-fledged general relativity with $\Lambda = 0$?

A related issue is that, from the Cartan perspective, 'doubly special relativity' may show up quite naturally as a limit of MacDowell–Mansouri gravity, given the relationship between DSR and de Sitter spacetime [18].

Another issue deals with describing matter in quantum gravity, by first studying matter in BF theory and then using the BF reformulation of MacDowell–Mansouri. Some work has already been done in this direction [5, 6, 14], but the switch to *Cartan-BF* theory may have important consequences for this effort, which should be investigated.

Appendix: Hodge duality in $\mathfrak{so}(3,1)$

The 6-dimensional Lie algebras $\mathfrak{so}(4)$, $\mathfrak{so}(3,1)$, and $\mathfrak{so}(2,2)$ inherit a notion of Hodge duality by the fact that they are isomorphic as vector spaces to $\Lambda^2\mathbb{R}^4$. In the $\mathfrak{so}(3,1)$ case:

$$\mathfrak{so}(3,1) \xrightarrow[\text{lower an index}]{} \Lambda^2\mathbb{R}^4 \xrightarrow[\text{Hodge duality}]{\star} \Lambda^2\mathbb{R}^4 \xrightarrow[\text{raise an index}]{} \mathfrak{so}(3,1)$$

Explicitly, this Hodge dual permutes $\mathfrak{so}(3,1)$ matrix entries as follows:

$$\star \begin{bmatrix} 0 & a & b & c \\ a & 0 & d & e \\ b & -d & 0 & f \\ c & -e & -f & 0 \end{bmatrix} = \begin{bmatrix} 0 & -f & e & -d \\ -f & 0 & c & -b \\ e & -c & 0 & a \\ -d & b & -a & 0 \end{bmatrix}$$

It is straightforward to verify the following properties of $\star$:

- $\star\star X = -X$
- $\star[X, X'] = [X, \star X']$

For MacDowell–Mansouri gravity and its BF reformulation, the essential application of the second property is that if ω is an $\mathrm{SO}(3,1)$ connection, the covariant differential d_ω commutes with the internal Hodge star operator:

$$d_\omega(\star X) = d(\star X) + [\omega, \star X] = \star(dX + [\omega, X]) = \star d_\omega X.$$

Acknowledgments

I thank John Baez, Aristide Baratin, Jim Dolan, Laurent Freidel, Bill Goldman, Jeff Morton, Artem Starodubtsev, Danny Stevenson, and Josh Willis for helpful discussions.

References

[1] D. V. Alekseevsky and P. W. Michor, Differential geometry of Cartan connections, *Publ. Math. Debrecen* **47** (1995), 349–375.

[2] S. Alexander, A quantum gravitational relaxation of the cosmological constant, *Phys. Lett.* B **629** (2005) 53-59. Also available as `hep-th/0503146`.

[3] J. C. Baez, Spin foam models, *Class. Quantum Grav.* **15** (1998), 1827–1858. Also available as `gr-qc/9709052`.

[4] J. C. Baez, An introduction to spin foam models of BF theory and quantum gravity, in *Geometry and Quantum Physics*, eds. H. Gausterer and H. Grosse, Lecture Notes in Physics **543**, Springer, Berlin, 2000, pp. 25-93. Also available as `gr-qc/9905087`.

[5] J. C. Baez, A. S. Crans, and D. K. Wise, Exotic statistics for strings in 4d BF theory, available as `gr-qc/0603085`.

[6] J. C. Baez and A. Perez, Quantization of strings and branes coupled to BF theory, available as `gr-qc/0605087`.

[7] S. Carlip, Geometric structures and loop variables in (2+1)-dimensional gravity, in *Knots and Quantum Gravity*, ed. J. Baez, Oxford University Press, Oxford, 1994. Also available as `gr-qc/9309020`.

[8] E. Cartan, Les espaces à connexion conforme, *Ann. Soc. Pol. Math.* **2**, 171–221, 1923.

[9] E. Cartan, Les groupes d'holonomie des espaces généralisés, *Acta Math.* **48**, 142, 1926.

[10] E. Cartan, *La Méthode du Repère Mobile, la Théorie des Groupes Continus et les Espaces Généralisés*, Exposés de Géométrie No. 5, Hermann, Paris, 1935.

[11] Y. Choquet-Bruhat, C. DeWitt-Morette, with M. Dillard-Bleick, *Analysis, Manifolds, and Physics*, rev. ed., North-Holland, Amsterdam, 1982.

[12] C. Ehresmann, Les connexions infinitésimales dans un espace fibré différentiable, *Colloque de Toplogie*, Bruxelles (1950) 29-55.

[13] L. Freidel and A. Starodubtsev, Quantum gravity in terms of topological observables, available as `hep-th/0501191`.

[14] L. Freidel, J. Kowalski–Glikman, and A. Starodubtsev, Particles as Wilson lines of gravitational field, to be published in *Phys. Rev.* D. Available as `gr-qc/0607014`.

[15] H. Garcia–Compean, A. Nieto, O. Obregon, and C. Ramirez, Dual Description of Supergravity MacDowell-Mansouri Theory, *Phys. Rev.* D **59** (1999) 124003. Also available as `hep-th/9812175`

[16] H. Garcia–Compean, O. Obregon, and C. Ramirez, Gravitational Duality in MacDowell-Mansouri Gauge Theory, *Phys. Rev.* D **58** (1998) 104012. Also available as `hep-th/9802063`

[17] C. J. Isham, *Modern Differential Geometry for Physicists*, World Scientific, River Edge, NJ, 1999.

[18] J. Kowalski–Glikman and S. Nowak, Doubly special relativity and de Sitter space, *Class. Quantum Grav.* **20** (2003) 4799-4816.

[19] G. Lisi, Clifford bundle formulation of BF gravity generalized to the standard model, available as `gr-qc/0511120`

[20] S. W. MacDowell and F. Mansouri, Unified geometric theory of gravity and supergravity, *Phys. Rev. Lett.* **38** (1977), 739–742. Erratum, *ibid.* **38** (1977), 1376.

[21] L. Smolin and A. Starodubtsev, General relativity with a topological phase: an action principle, available as `hep-th/0311163`.

[22] J. A. Nieto, Superfield Description of a Self-Dual Supergravity a la MacDowell-Mansouri, *Class. Quantum Grav.* **23** (2006) 4387-4398. Also available as `hep-th/0509169`.

[23] E. Ruh, Cartan connections, *Proc. Sympos. Pure Math.* **54**, part 3, A.M.S., Providence, RI, 1993.

[24] R. W. Sharpe, *Differential Geometry: Cartan's Generalization of Klein's Erlangen Program*, Springer, New York, 1997.

[25] A. Starodubtsev, Topological methods in quantum gravity, Ph.D. Thesis, University of Waterloo, 2005.

[26] W. P. Thurston *Three-Dimensional Geometry and Topology*, Vol. 1, Princeton University Press, Princeton, 1997.

[27] R. S. Tung, Gravitation as a super $\mathrm{SL}(2,\mathbb{C})$ gauge theory, Proceedings of the 9th Marcel Grossmann Meeting, Rome, 2-8 July, 2000. Also available as `gr-qc/0101034`